7450 Ethernet Service Switch
7750 Service Router
7950 Extensible Routing System

Releases up to 24.3.R2

Layer 2 Services and EVPN Advanced Configuration Guide for Classic CLI

3HE 20792 AAAA TQZZA
Edition: 01
July 2024

Nokia is committed to diversity and inclusion. We are continuously reviewing our customer documentation and consulting with standards bodies to ensure that terminology is inclusive and aligned with the industry. Our future customer documentation will be updated accordingly.

Table of contents

List of tables

List of figures

Preface

About This Guide

Each Advanced Configuration Guide is organized alphabetically and provides feature and configuration explanations, CLI descriptions, and overall solutions. The Advanced Configuration Guide chapters are written for and based on several Releases, up to 24.7.R2. The Applicability section in each chapter specifies on which release the configuration is based.

The Advanced Configuration Guides supplement the user configuration guides listed in the 7450 ESS, 7750 SR, and 7950 XRS Guide to Documentation.

Audience

This manual is intended for network administrators who are responsible for configuring the routers. It is assumed that the network administrators have a detailed understanding of networking principles and configurations.

AC-Influenced DF Election on an ES

This chapter provides information about Attachment Circuit (AC) influenced Designated Forwarder (DF) election on an Ethernet Segment (ES).

Topics in this chapter include:

Applicability

The information and configuration in this chapter are based on SR OS Release 22.5.R1. Attachment Circuit (AC) influenced Designated Forwarder (DF) election on an Ethernet Segment (ES) is always enabled in SR OS releases earlier than 21.5.R1. The AC-DF election capability can be disabled in SR OS Release 21.5.R1 and later.

Overview

RFC 8584, section “The AC-Influenced DF Election Capability”, describes the AC-DF capability that modifies the EVPN DF election process in RFC 7432. RFC 8584 states that when PEs build their candidate DF election list, they do not include PEs when no Auto-Discovery (AD) per-ES or per-EVI routes for those PEs are present. In SR OS, this behavior is default for all ESs, configured as **ac-df-capability include**.

The **ac-df-capability** command is configurable in the **config>service>system>bgp-evpn>eth-seg** context:

```
*A:PE-2>config>service>system>bgp-evpn>eth-seg# ac-df-capability ?
  - ac-df-capability {include|exclude}
```

The command **ac-df-capability exclude** disables AC-DF on the ES, so the presence of an AD per-ES or per-EVI does not influence the candidate DF election list. When **ac-df-capability exclude** is configured:

- The candidate DF election list is not influenced by the presence or absence of AD per-ES/EVI routes (type 1) from the ES peers.
- PEs are only removed from the candidate DF election list when their ES route (type 4) is not present.
- The local ES route is active if there are active SAPs on the ES.
- When the local AC is operationally down, due to admin shutdown or reason other than Multi Homing (MH) standby, this does not trigger a DF switchover.

The **ac-df-capability exclude** option:

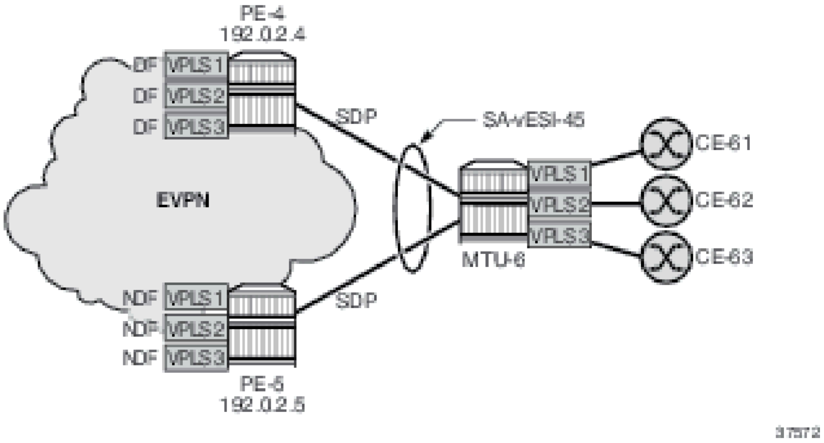
- is supported with any type of service-carving (DF Election)
- is recommended in ESs that use an operational group monitored by the access LAG to signal standby LACP or power-off
- must be configured consistently on all PEs attached to the same ES

AC-DF enabled – default

The following example illustrates the default behavior, where a PE builds the list of DF candidates with nodes that have sent EVPN AD per-ES/EVI routes. This behavior is compatible with the behavior in SR OS releases earlier than 21.5.R1.

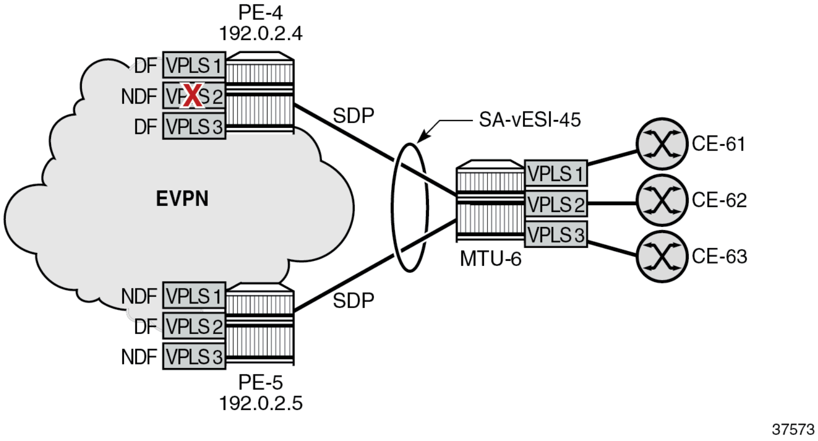
Figure 1: PE-4 as the DF on a single-active ES for three VPLSs shows a topology with MTU-6 connected via SDPs to the single-active ES "SA-vESI-45". PE-4 is the DF for three services: VPLS 1, VPLS 2, and VPLS 3. Traffic for these services passes via PE-4, while PE-5 is standby.

Figure 1: PE-4 as the DF on a single-active ES for three VPLSs



When a failure occurs on the spoke-SDP in VPLS 2 on PE-4, PE-4 sends an EVPN-AD per-EVI withdrawal and PE-4 becomes the Non-Designated Forwarder (NDF) for VPLS 2, while remaining the DF for VPLS 1 and VPLS 3, as shown in Figure 2: AC failure in VPLS 2 on PE-4 causes PE-5 to become the DF for VPLS 2.

Figure 2: AC failure in VPLS 2 on PE-4 causes PE-5 to become the DF for VPLS 2



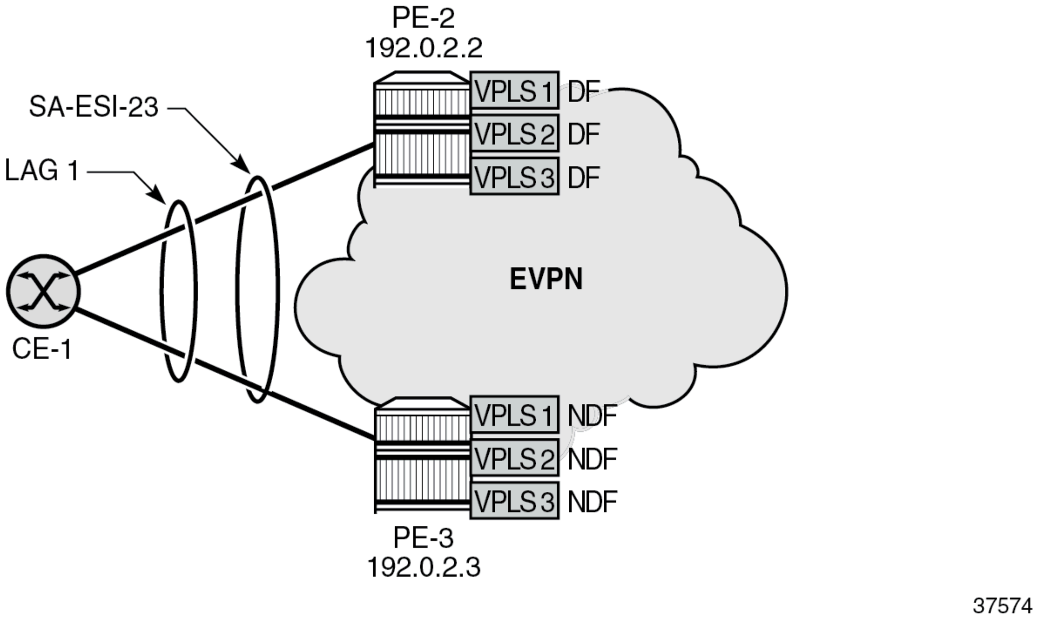
VPLS 2 traffic to and from MTU-6 passes via DF PE-5, while VPLS 1 and VPLS 3 traffic will pass via DF PE-4. No traffic is dropped. The AC failure in VPLS 2 does not have an impact on the other services.

Problem with AC-DF on ES with the operational group monitored by LAG

In this example, a failure in an access circuit of a particular service also impacts other services when the AC-DF capability is enabled.

Figure 3: PE-2 is DF on single-active ES for three VPLSs shows a single-active ES with LAG 1 associated with it. An operational group is assigned to the ES and monitored by the LAG to signal standby LACP (default) or power off. Three VPLSs are configured on PE-2 and PE-3. PE-2 is the DF for each of these VPLSs.

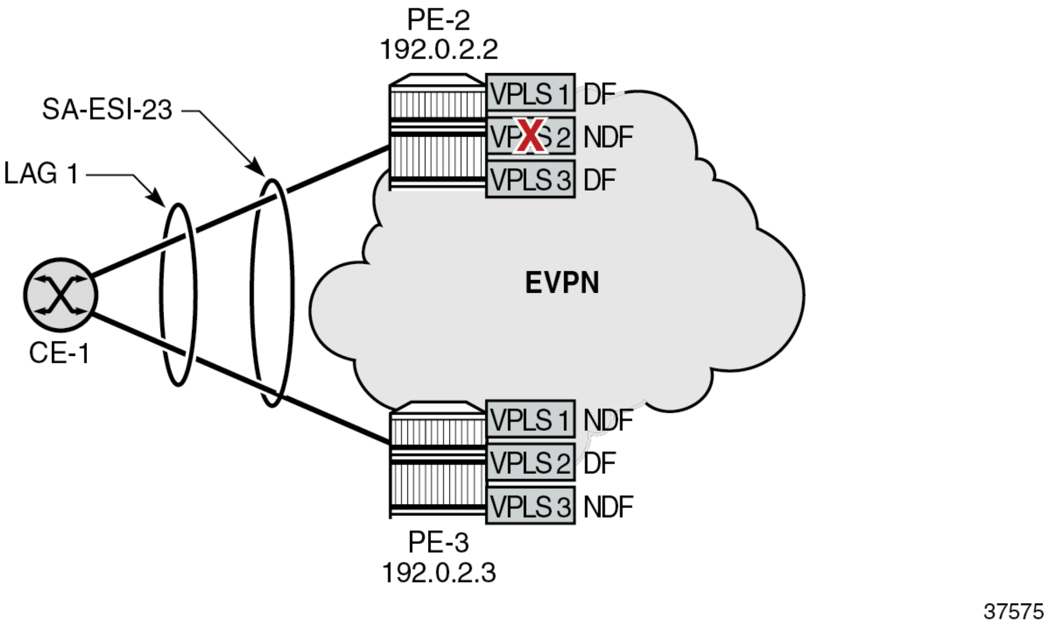
Figure 3: PE-2 is DF on single-active ES for three VPLSs



On NDF PE-3, the ES is inactive which causes the operational group in the ES to go down. LAG 1 monitors this operational group, so the LAG goes standby on NDF PE-3. LAG 1 has LACP standby-signaling enabled (default). On CE-1, only the LAG port to DF PE-2 is up and all traffic for the VPLSs goes via PE-2.

When the single-active ES has the default AC-DF setting (**ac-df-capability include**), a failure (or an unintended shutdown) on SAP lag-1:2 in VPLS 2 (or on the VPLS 2 service) on PE-2 can have an impact on all three services that share LAG 1. Figure 4: AC failure in VPLS 2 on PE-2 causes PE-3 to become DF for VPLS 2 shows that such an AC failure in VPLS 2 on PE-2 causes PE-3 to become the DF for VPLS 2 (after receiving an AD per-EVI withdrawal from PE-2).

Figure 4: AC failure in VPLS 2 on PE-2 causes PE-3 to become DF for VPLS 2

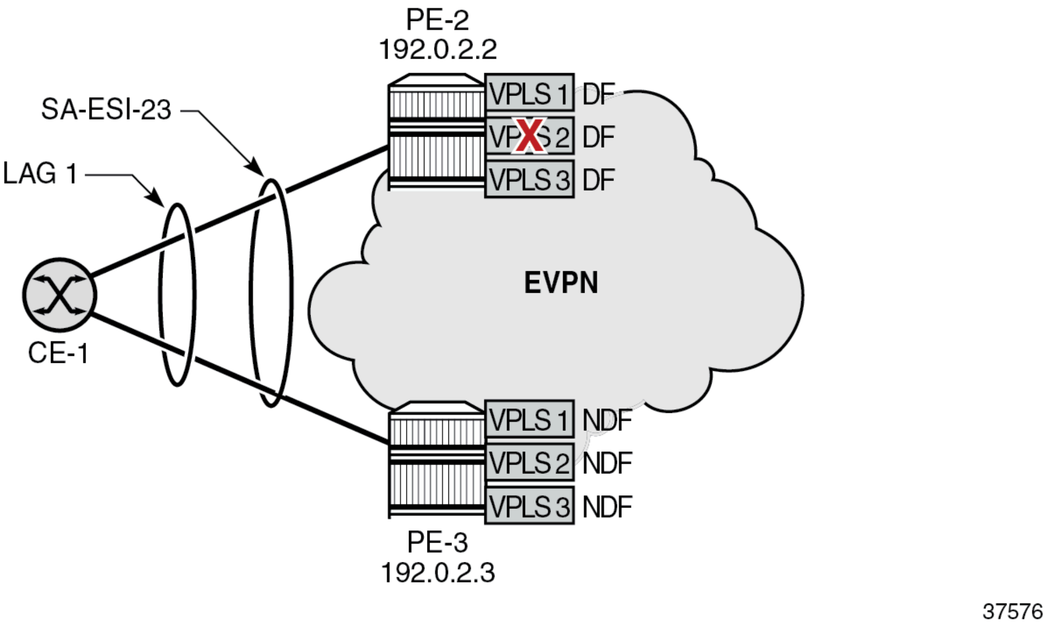


When PE-3 is the DF for VPLS 2, the ES operational group on PE-3 goes up. Therefore, the monitoring LAG is up on PE-3. On CE-1, both LAG ports to PE-2 and PE-3 are up. CE-1 can now send all VPLS traffic via either LAG port: DF PE-2 forwards the VPLS 1 and VPLS 3 traffic whereas NDF PE-3 drops it. PE-3 accepts VPLS 2 traffic, but PE-2 drops it. Approximately 50% of the traffic is lost.

AC-DF capability disabled

Nokia recommends disabling the AC-DF capability in ESs where the operational group is monitored by the LAG. Figure 5: AC failure in VPLS 2 on PE-2 has no impact on DF election shows the situation with the AC-DF disabled (**ac-df-capability exclude**): the PEs ignore the AD per-EVI withdrawal and PE-2 remains the DF for VPLS 2.

Figure 5: AC failure in VPLS 2 on PE-2 has no impact on DF election

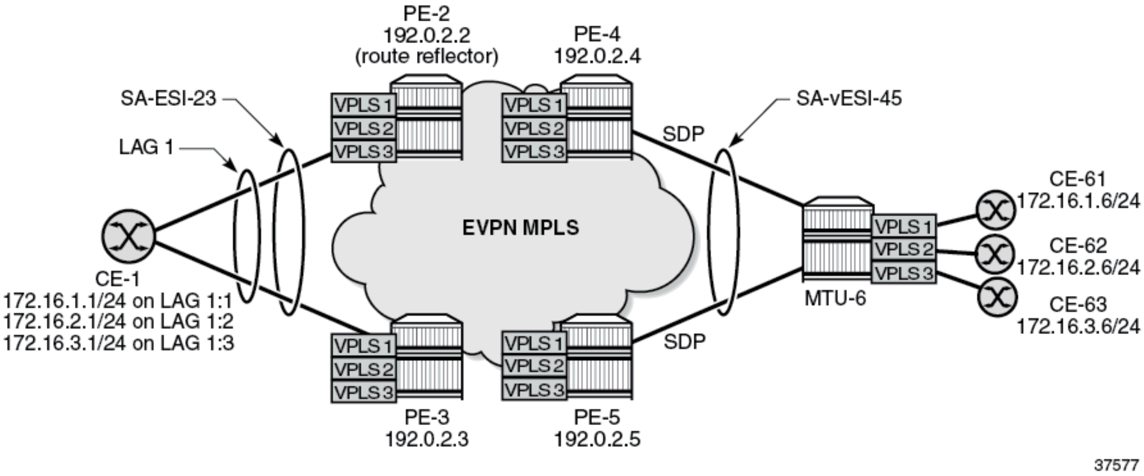


VPLS 2 traffic is dropped by PE-2, but the other services are not impacted.

Configuration

Figure 6: Example topology shows the example topology with four PEs in an EVPN-MPLS network.

Figure 6: Example topology



The initial configuration includes:

- cards, MDAs, ports

- router interfaces on the PEs and on MTU-6
- IS-IS on the router interfaces (alternatively, OSPF can be configured)
- LDP on the router interfaces

On the PEs, BGP is configured for the EVPN address family. In this example, PE-2 is the Route Reflector (RR) with the following BGP configuration:

```
# on PE-2:
configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            rapid-update evpn
            group "internal"
                family evpn
                cluster 192.0.2.2
                peer-as 64500
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
        exit
```

The BGP configuration on the clients PE-3, PE-4, and PE-5 is as follows:

```
# on PE-3, PE-4, PE-5:
configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            rapid-update evpn
            group "internal"
                family evpn
                peer-as 64500
                neighbor 192.0.2.2
                exit
            exit
        exit
```

AC-DF capability enabled – default

On PE-2 and PE-3, operational group "op-grp-sa-es-23" is configured. This operational group is assigned to the single-active ES "SA-ESI-23" and monitored on LAG 1.

On PE-2, LAG 1 is configured as follows. The LAG configuration on PE-3 is similar, but with port 1/1/1 instead.

```
# on PE-2:
configure
    lag 1 name "lag-1"
        mode access
        encap-type dot1q
        monitor-oper-group "op-grp-sa-es-23"
        port 1/1/2
        lacp active administrative-key 1 system-id 00:00:00:00:23:01
        standby-signaling lacp      # default
        no shutdown
```

On PE-2 and PE-3, three VPLS services are configured with SAPs from LAG 1, which is associated with single-active ES "SA-ESI-23". This ES is configured with the operational group "op-grp-sa-es-23" that is monitored by LAG 1. The operational group triggers the LACP standby signaling from the NDF PE to CE-1 to avoid attracting traffic.

The service configuration on PE-2 and PE-3 is similar; only the preference value for the service carving in the ES is different.

Note:
When an operational group is associated with an ES, the hold timers for the operational group must be zero (the default value).

```
# on PE-2:
configure
    service
        oper-group "op-grp-sa-es-23" create
            hold-time
                group down 0    # default
                group up 0
            exit
        exit
        system
            bgp-evpn
                ethernet-segment "SA-ESI-23" create
                    esi 01:00:00:00:00:23:01:00:00:01
                    service-carving
                        mode manual
                        manual
                            preference non-revertive create
                                value 200         # on PE-3: preference value 100
                            exit
                        exit
                    exit
                    multi-homing single-active
                    ac-df-capability include    # default
                    lag 1
                    oper-group "op-grp-sa-es-23
                    no shutdown
                exit
            exit
        exit
        vpls 1 name "VPLS 1" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
```

```
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap lag-1:1 create
                no shutdown
            exit
            no shutdown
        exit
        vpls 2 name "VPLS 2" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 2
                mpls bgp 1
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap lag-1:2 create
                no shutdown
            exit
            no shutdown
        exit
        vpls 3 name "VPLS 3" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 3
                mpls bgp 1
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap lag-1:3 create
                no shutdown
            exit
            no shutdown
        exit
```

On PE-4 and PE-5, single-active virtual ES "SA-vESI-45" is configured. No operational group is configured here. The service configuration on PE-4 is as follows. The configuration on PE-5 is similar, but with a different SDP and a different preference value for service carving.

```
# on PE-4:
configure
    service
        sdp 46 mpls create        # on PE-5: sdp 56
            far-end 192.0.2.6
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        system
            bgp-evpn
                ethernet-segment "SA-vESI-45" virtual create
                    esi 01:00:00:00:00:45:01:00:00:01
                    service-carving
                        mode manual
                        manual
                            preference create
                                value 200        # on PE-5: value 100
                            exit
                        exit
                    exit
                    multi-homing single-active
                    ac-df-capability include    # default
                    sdp 46
                    vc-id-range 1 to 3
                    no shutdown
                exit
            exit
        exit
        vpls 1 name "VPLS 1" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            spoke-sdp 46:1 create        # on PE-5: spoke-sdp 56:1
                no shutdown
            exit
            no shutdown
        exit
        vpls 2 name "VPLS 2" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 2
                mpls bgp 1
```

```
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            spoke-sdp 46:2 create        # on PE-5: spoke-sdp 56:2
                no shutdown
            exit
            no shutdown
        exit
        vpls 3 name "VPLS 3" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 3
                mpls bgp 1
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            spoke-sdp 46:3 create        # on PE-5: spoke-sdp 56:3
                no shutdown
            exit
            no shutdown
        exit
```

By default, **ac-df-capability include** is used. With the AC-DF capability enabled, the PEs send ES routes with **AC:1** in the extended community for DF election. The following ES route is received by PE-3 from PE-2:

```
53 2022/06/03 08:49:51.298 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 71
    Flag: 0x90 Type: 14 Len: 34 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-ETH-SEG Len: 23 RD: 192.0.2.2:0 ESI: 01:00:00:00:00:23:01:00:00:01, IP-Len:
 4 Orig-IP-Addr: 192.0.2.2
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        df-election::DF-Type:Preference/DP:1/DF-Preference:200/AC:1
        target:00:00:00:00:23:01
```

The remainder of the chapter focuses on PE-2 and PE-3, where an AC failure in one of the services can have an impact on the other services using the same LAG.

DF election

PE-2 is the highest-preference PE in the ES and becomes the DF (preference value 200 on PE-2 versus preference value 100 on PE-3). In case of equal preference value between PE-2 and PE-3, the Don't Preempt (DP) bit is the tiebreaker (DP = 1 for non-revertive wins over DP = 0); if that is also a tie, the lowest PE IP address is the tiebreaker.

The following command shows that PE-2 is the DF for all three VPLSs. The candidate list contains both PE-2 and PE-3 for each of these VPLSs.

```
*A:PE-2# show service system bgp-evpn ethernet-segment name "SA-ESI-23" all

===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : SA-ESI-23
Eth Seg Type            : None
Admin State             : Enabled            Oper State          : Up
ESI                     : 01:00:00:00:00:23:01:00:00:01
Oper ESI                : 01:00:00:00:00:23:01:00:00:01
Auto-ESI Type           : None
AC DF Capability        : Include
Multi-homing            : singleActive       Oper Multi-homing  : singleActive
ES SHG Label            : 524282
Source BMAC LSB         : None
Lag Id                  : 1
ES Activation Timer     : 3 secs (default)
Oper Group              : op-grp-sa-es-23
Svc Carving             : manual             Oper Svc Carving   : manual
Cfg Range Type          : lowest-pref

-------------------------------------------------------------------------------
DF Pref Election Information
-------------------------------------------------------------------------------
Preference     Preference     Last Admin Change        Oper Pref      Do No
Mode           Value                                   Value          Preempt
-------------------------------------------------------------------------------
non-revertive  200            06/03/2022 08:49:51      200            Enabled
-------------------------------------------------------------------------------
EVI Ranges: <none>
ISID Ranges: <none>
===============================================================================

===============================================================================
EVI Information
===============================================================================
EVI                 SvcId               Actv Timer Rem      DF
-------------------------------------------------------------------------------
1                   1                   0                   yes
2                   2                   0                   yes
3                   3                   0                   yes
-------------------------------------------------------------------------------
Number of entries: 3
===============================================================================

-------------------------------------------------------------------------------
DF Candidate list
-------------------------------------------------------------------------------
EVI                                     DF Address
-------------------------------------------------------------------------------
1                                       192.0.2.2
```

```
1                                       192.0.2.3
2                                       192.0.2.2
2                                       192.0.2.3
3                                       192.0.2.2
3                                       192.0.2.3
-------------------------------------------------------------------------------
Number of entries: 6
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
---snip---
```

The same command on PE-3 shows that PE-3 is NDF for the three VPLSs and the DF candidate list is identical to the one on PE-2:

```
*A:PE-3# show service system bgp-evpn ethernet-segment name "SA-ESI-23" all

===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : SA-ESI-23
Eth Seg Type            : None
Admin State             : Enabled            Oper State         : Up
ESI                     : 01:00:00:00:00:23:01:00:00:01
Oper ESI                : 01:00:00:00:00:23:01:00:00:01
Auto-ESI Type           : None
AC DF Capability        : Include
Multi-homing            : singleActive       Oper Multi-homing  : singleActive
ES SHG Label            : 524282
Source BMAC LSB         : None
Lag Id                  : 1
ES Activation Timer     : 3 secs (default)
Oper Group              : op-grp-sa-es-23
Svc Carving             : manual             Oper Svc Carving   : manual
Cfg Range Type          : lowest-pref

-------------------------------------------------------------------------------
DF Pref Election Information
-------------------------------------------------------------------------------
Preference     Preference     Last Admin Change        Oper Pref      Do No
Mode           Value                                   Value          Preempt
-------------------------------------------------------------------------------
non-revertive  100            06/03/2022 08:49:42      100            Enabled
-------------------------------------------------------------------------------
EVI Ranges: <none>
ISID Ranges: <none>
===============================================================================

===============================================================================
EVI Information
===============================================================================
EVI                 SvcId               Actv Timer Rem      DF
-------------------------------------------------------------------------------
1                   1                   0                   no
2                   2                   0                   no
3                   3                   0                   no
-------------------------------------------------------------------------------
Number of entries: 3
===============================================================================

-------------------------------------------------------------------------------
DF Candidate list
-------------------------------------------------------------------------------
EVI                                     DF Address
```

```
-------------------------------------------------------------------------------
1                                       192.0.2.2
1                                       192.0.2.3
2                                       192.0.2.2
2                                       192.0.2.3
3                                       192.0.2.2
3                                       192.0.2.3
-------------------------------------------------------------------------------
Number of entries: 6
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
---snip---
```

Operational group status

PE-2 is the DF, so the ES "SA-ESI-23" is active, the operational group "op-grp-sa-es-23" is operationally up, and the monitoring LAG 1 is operationally up.

```
*A:PE-2# show service oper-group "op-grp-sa-es-23" detail

===============================================================================
Service Oper Group Information
===============================================================================
Oper Group       : op-grp-sa-es-23
Creation Origin  : manual                          Oper Status: up
Hold DownTime    : 0 secs                          Hold UpTime: 0 secs
Members          : 1                               Monitoring : 1
===============================================================================

===============================================================================
Member Ethernet-Segment for OperGroup: op-grp-sa-es-23
===============================================================================
Ethernet-Segment                        Status
-------------------------------------------------------------------------------
SA-ESI-23                               Active
-------------------------------------------------------------------------------
Ethernet-Segment Entries found: 1
===============================================================================

===============================================================================
Monitoring LAG for OperGroup: op-grp-sa-es-23
===============================================================================
Lag-id         Adm       Opr       Weighted  Threshold Up-Count  Act/Stdby
    name
-------------------------------------------------------------------------------
1              up        up        No        0         1         N/A
    lag-1
-------------------------------------------------------------------------------
LAG Entries found: 1
===============================================================================
port option not supported with monitoring
```

PE-3 is NDF, so the ES "SA-ESI-23" is inactive, the operational group "op-grp-sa-es-23" is operationally down, and the monitoring LAG 1 is operationally down:

```
*A:PE-3# show service oper-group "op-grp-sa-es-23" detail

===============================================================================
Service Oper Group Information
===============================================================================
Oper Group       : op-grp-sa-es-23
Creation Origin  : manual                          Oper Status: down
Hold DownTime    : 0 secs                          Hold UpTime: 0 secs
Members          : 1                               Monitoring : 1
===============================================================================

===============================================================================
Member Ethernet-Segment for OperGroup: op-grp-sa-es-23
===============================================================================
Ethernet-Segment                        Status
-------------------------------------------------------------------------------
SA-ESI-23                               Inactive
-------------------------------------------------------------------------------
Ethernet-Segment Entries found: 1
===============================================================================

===============================================================================
Monitoring LAG for OperGroup: op-grp-sa-es-23
===============================================================================
Lag-id         Adm       Opr       Weighted  Threshold Up-Count  Act/Stdby
    name
-------------------------------------------------------------------------------
1              up        down      No        0         0         N/A
    lag-1
-------------------------------------------------------------------------------
LAG Entries found: 1
===============================================================================
port option not supported with monitoring
```

LAG port status

On DF PE-2, LAG port 1/1/2 toward CE-1 is operationally up:

```
*A:PE-2# show lag "lag-1" port

===============================================================================
Lag Port States
LACP Status: e - Enabled, d - Disabled
===============================================================================
Name
Id     Port-id        Adm   Act/     Opr   Primary Sub-group      Forced Prio
                            Stdby
-------------------------------------------------------------------------------
lag-1
1(e)   1/1/2          up    active   up    yes     1              -      32768
===============================================================================
```

On NDF PE-3, LAG port 1/1/1 toward CE-1 is operationally down:

```
*A:PE-3# show lag "lag-1" port

===============================================================================
Lag Port States
LACP Status: e - Enabled, d - Disabled
===============================================================================
Name
Id     Port-id        Adm   Act/     Opr   Primary Sub-group      Forced Prio
                            Stdby
-------------------------------------------------------------------------------
```

```
lag-1
1(e)   1/1/1          up    active   down  yes     1             -      32768
===============================================================================
```

On CE-1, LAG port 1/1/1 toward DF PE-2 is operationally up while LAG port 1/1/2 toward NDF PE-3 is down:

```
*A:CE-1# show lag "lag-1" port

===============================================================================
Lag Port States
LACP Status: e - Enabled, d - Disabled
===============================================================================
Name
Id     Port-id        Adm   Act/     Opr   Primary Sub-group      Forced Prio
                            Stdby
-------------------------------------------------------------------------------
lag-1
1(e)   1/1/1          up    active   up    yes     1             -      32768
       1/1/2          up    active   down          1             -      32768
===============================================================================
```

AD per-EVI route withdrawal

A failure is simulated by disabling SAP lag-1:2 in VPLS 2 on PE-2:

```
# on PE-2:
configure
    service
        vpls "VPLS 2"
            sap lag-1:2
                shutdown
```

PE-2 withdraws the EVPN-AD per-EVI route. The following withdrawal is received by PE-3:

```
101 2022/06/03 08:54:07.346 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 69
    Flag: 0x90 Type: 15 Len: 65 Multiprotocol Unreachable NLRI:
        Address Family EVPN
        Type: EVPN-MAC Len: 33 RD: 192.0.2.2:2 ESI: ESI-0, tag: 0, mac len: 48 mac:
 00:00:00:00:02:01, IP len: 0, IP: NULL, label1: 0
        Type: EVPN-AD Len: 25 RD: 192.0.2.2:2 ESI: 01:00:00:00:00:23:01:00:00:01, tag: 0 Label:
 0 (Raw Label: 0x0) PathId:
"
```

The following command on PE-3 shows that the list of DF candidates no longer includes PE-2 in the DF candidate list for VPLS 2 and that PE-3 is the DF for VPLS 2, while remaining the NDF for VPLS 1 and VPLS 3.

```
*A:PE-3# show service system bgp-evpn ethernet-segment name "SA-ESI-23" all | match "EVI
 Information" pre-lines 2 post-lines 25

===============================================================================
EVI Information
```

```
===============================================================================
EVI                 SvcId               Actv Timer Rem      DF
-------------------------------------------------------------------------------
1                   1                   0                   no
2                   2                   0                   yes
3                   3                   0                   no
-------------------------------------------------------------------------------
Number of entries: 3
===============================================================================

-------------------------------------------------------------------------------
DF Candidate list
-------------------------------------------------------------------------------
EVI                                     DF Address
-------------------------------------------------------------------------------
1                                       192.0.2.2
1                                       192.0.2.3
2                                       192.0.2.3
3                                       192.0.2.2
3                                       192.0.2.3
-------------------------------------------------------------------------------
Number of entries: 5
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
```

When PE-3 becomes the DF for one of the services, the ES "SA-ESI-23" is active and the operational group "op-grp-sa-es-23" and LAG 1 are up, as follows:

```
*A:PE-3# show service oper-group "op-grp-sa-es-23" detail

===============================================================================
Service Oper Group Information
===============================================================================
Oper Group       : op-grp-sa-es-23
Creation Origin  : manual                          Oper Status: up
Hold DownTime    : 0 secs                          Hold UpTime: 0 secs
Members          : 1                               Monitoring : 1
===============================================================================

===============================================================================
Member Ethernet-Segment for OperGroup: op-grp-sa-es-23
===============================================================================
Ethernet-Segment                        Status
-------------------------------------------------------------------------------
SA-ESI-23                               Active
-------------------------------------------------------------------------------
Ethernet-Segment Entries found: 1
===============================================================================

===============================================================================
Monitoring LAG for OperGroup: op-grp-sa-es-23
===============================================================================
Lag-id         Adm       Opr       Weighted  Threshold Up-Count  Act/Stdby
    name
-------------------------------------------------------------------------------
1              up        up        No        0         1         N/A
    lag-1
-------------------------------------------------------------------------------
LAG Entries found: 1
===============================================================================
port option not supported with monitoring
```

On PE-3, LAG port 1/1/1 toward CE-1 is up:

```
*A:PE-3# show lag "lag-1" port

===============================================================================
Lag Port States
LACP Status: e - Enabled, d - Disabled
===============================================================================
Name
Id     Port-id        Adm   Act/     Opr   Primary Sub-group      Forced Prio
                            Stdby
-------------------------------------------------------------------------------
lag-1
1(e)   1/1/1          up    active   up    yes     1              -      32768
===============================================================================
```

PE-2 remains the DF for VPLS 1 and VPLS 3:

```
*A:PE-2# show service system bgp-evpn ethernet-segment name "SA-ESI-23" all | match "EVI
 Information" pre-lines 2 post-lines 25

===============================================================================
EVI Information
===============================================================================
EVI                 SvcId               Actv Timer Rem      DF
-------------------------------------------------------------------------------
1                   1                   0                   yes
2                   2                   0                   no
3                   3                   0                   yes
-------------------------------------------------------------------------------
Number of entries: 3
===============================================================================

-------------------------------------------------------------------------------
DF Candidate list
-------------------------------------------------------------------------------
EVI                                     DF Address
-------------------------------------------------------------------------------
1                                       192.0.2.2
1                                       192.0.2.3
2                                       192.0.2.3
3                                       192.0.2.2
3                                       192.0.2.3
-------------------------------------------------------------------------------
Number of entries: 5
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
```

On PE-2, ES "SA-ESI-23" remains active, so the operational group "op-grp-sa-es-23" is up and the monitoring LAG is also up:

```
*A:PE-2# show service oper-group "op-grp-sa-es-23" detail

===============================================================================
Service Oper Group Information
===============================================================================
Oper Group       : op-grp-sa-es-23
Creation Origin  : manual                          Oper Status: up
Hold DownTime    : 0 secs                          Hold UpTime: 0 secs
Members          : 1                               Monitoring : 1
===============================================================================
===============================================================================
Member Ethernet-Segment for OperGroup: op-grp-sa-es-23
===============================================================================
Ethernet-Segment                        Status
-------------------------------------------------------------------------------
SA-ESI-23                               Active
-------------------------------------------------------------------------------
Ethernet-Segment Entries found: 1
===============================================================================

===============================================================================
Monitoring LAG for OperGroup: op-grp-sa-es-23
===============================================================================
Lag-id         Adm       Opr       Weighted  Threshold Up-Count  Act/Stdby
    name
-------------------------------------------------------------------------------
1              up        up        No        0         1         N/A
    lag-1
-------------------------------------------------------------------------------
LAG Entries found: 1
===============================================================================
port option not supported with monitoring
```

The following commands on PE-2 shows that SAP lag-1:1 in VPLS 1 is up, SAP lag-1:2 in VPLS 2 is down (as it might be due to a failure or misconfiguration), and SAP lag-1:3 in VPLS 3 is up:

```
*A:PE-2# show service id 1 sap

===============================================================================
SAP(Summary), Service 1
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
lag-1:1                         1          1     none    1     none   Up   Up
-------------------------------------------------------------------------------
Number of SAPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

```
*A:PE-2# show service id 2 sap

===============================================================================
SAP(Summary), Service 2
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
lag-1:2                         2          1     none    1     none   Down Down
-------------------------------------------------------------------------------
Number of SAPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

```
*A:PE-2# show service id 3 sap

===============================================================================
SAP(Summary), Service 3
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
lag-1:3                         3          1     none    1     none   Up   Up
-------------------------------------------------------------------------------
Number of SAPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

On PE-3, lag-1:2 is up while lag-1:1 and lag-1:3 are down, as follows:

```
*A:PE-3# show service sap-using sap lag-1

===============================================================================
Service Access Points
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
lag-1:1                         1          1     none    1     none   Up   Down
lag-1:2                         2          1     none    1     none   Up   Up
lag-1:3                         3          1     none    1     none   Up   Down
-------------------------------------------------------------------------------
Number of SAPs : 3
-------------------------------------------------------------------------------
===============================================================================
```

On CE-1, both ports in LAG 1 are up:

```
*A:CE-1# show lag "lag-1" port

===============================================================================
Lag Port States
LACP Status: e - Enabled, d - Disabled
===============================================================================
Name
Id     Port-id        Adm   Act/     Opr   Primary Sub-group      Forced Prio
                            Stdby
-------------------------------------------------------------------------------
lag-1
1(e)   1/1/1          up    active   up    yes     1              -      32768
       1/1/2          up    active   up            1              -      32768
===============================================================================
```

All traffic can take either LAG port, but PE-2 only forwards traffic for VPLS 1 and VPLS 3, while PE-3 only forwards traffic for VPLS 2. Traffic from VPLS 1 or VPLS 3 via port 1/1/2 to PE-3 is dropped by PE-3 because it is the NDF for VPLS 1 and VPLS 3. VPLS 2 traffic via LAG port 1/1/1 to PE-2 is dropped because SAP lag-1:2 is down (failure). This means that approximately 50% of the traffic is lost.

Potential loss on a single service under maintenance is acceptable but affecting other services on the same node is not acceptable. The solution is to disable the AC-DF capability.

AC-DF capability disabled

The default use of the AC-DF capability in SR OS is disabled on PE-2 and PE-3:

```
# on PE-2, PE-3:
configure
    service
        system
            bgp-evpn
                ethernet-segment "SA-ESI-23"
                    shutdown
                    ac-df-capability exclude
                    no shutdown
```

With AC-DF disabled, ES routes contain AC:0 in the DF-election extended community, as follows:

```
# on PE-3:
156 2022/06/03 08:55:53.529 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 71
    Flag: 0x90 Type: 14 Len: 34 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-ETH-SEG Len: 23 RD: 192.0.2.2:0 ESI: 01:00:00:00:00:23:01:00:00:01, IP-Len:
 4 Orig-IP-Addr: 192.0.2.2
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        df-election::DF-Type:Preference/DP:1/DF-Preference:200/AC:0
        target:00:00:00:00:23:01
"
```

With the AC-DF capability disabled, the withdrawal of EVPN-AD routes does not influence the DF election. In this example, PE-2 remains the DF for all services, including VPLS 2, even when traffic for that service is dropped by PE-2. The following command shows that the DF candidate list on PE-3 contains six entries: even for VPLS 2, PE-2 is included in the list. PE-3 is the NDF for all three services.

```
*A:PE-3# show service system bgp-evpn ethernet-segment name "SA-ESI-23" all

===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : SA-ESI-23
Eth Seg Type            : None
Admin State             : Enabled            Oper State         : Up
ESI                     : 01:00:00:00:00:23:01:00:00:01
Oper ESI                : 01:00:00:00:00:23:01:00:00:01
Auto-ESI Type           : None
AC DF Capability        : Exclude
Multi-homing            : singleActive       Oper Multi-homing  : singleActive
ES SHG Label            : 524275
Source BMAC LSB         : None
Lag Id                  : 1
ES Activation Timer     : 3 secs (default)
```

```
Oper Group              : op-grp-sa-es-23
Svc Carving             : manual             Oper Svc Carving   : manual
Cfg Range Type          : lowest-pref

-------------------------------------------------------------------------------
DF Pref Election Information
-------------------------------------------------------------------------------
Preference     Preference     Last Admin Change        Oper Pref      Do No
Mode           Value                                   Value          Preempt
-------------------------------------------------------------------------------
non-revertive  100            06/03/2022 08:49:42      100            Enabled
-------------------------------------------------------------------------------
EVI Ranges: <none>
ISID Ranges: <none>
===============================================================================

===============================================================================
EVI Information
===============================================================================
EVI                SvcId              Actv Timer Rem      DF
-------------------------------------------------------------------------------
1                  1                  0                   no
2                  2                  0                   no
3                  3                  0                   no
-------------------------------------------------------------------------------
Number of entries: 3
===============================================================================

-------------------------------------------------------------------------------
DF Candidate list
-------------------------------------------------------------------------------
EVI                                   DF Address
-------------------------------------------------------------------------------
1                                     192.0.2.2
1                                     192.0.2.3
2                                     192.0.2.2
2                                     192.0.2.3
3                                     192.0.2.2
3                                     192.0.2.3
-------------------------------------------------------------------------------
Number of entries: 6
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
---snip---
```

On NDF PE-3, the single-active ES "SA-ESI-23" is inactive and the ES operational group is down. The monitoring LAG is also operationally down.

On CE-1, LAG port 1/1/2 toward PE-3 is down:

```
*A:CE-1# show lag "lag-1" port

===============================================================================
Lag Port States
LACP Status: e - Enabled, d - Disabled
===============================================================================
Name
Id     Port-id        Adm   Act/     Opr   Primary Sub-group      Forced Prio
                            Stdby
-------------------------------------------------------------------------------
lag-1
1(e)   1/1/1          up    active   up    yes     1              -      32768
```

```
       1/1/2          up    active   down          1              -      32768
===============================================================================
```

CE-1 sends all traffic via LAG port 1/1/1 to PE-2. VPLS 1 and VPLS 3 traffic is forwarded by DF PE-2, whereas VPLS 2 traffic is dropped. Therefore, the failure does not have an impact on the other services.

On PE-2, SAP lag-1:1 in VPLS 1 and SAP lag-1:3 in VPLS 3 are operationally up:

```
*A:PE-2# show service id 1 sap

===============================================================================
SAP(Summary), Service 1
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
lag-1:1                         1          1     none    1     none   Up   Up
-------------------------------------------------------------------------------
Number of SAPs : 1
-------------------------------------------------------------------------------
===============================================================================

*A:PE-2# show service id 3 sap

===============================================================================
SAP(Summary), Service 3
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
lag-1:3                         3          1     none    1     none   Up   Up
-------------------------------------------------------------------------------
Number of SAPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

On PE-3, all SAPs in the VPLSs are down:

```
*A:PE-3# show service id 2 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 2                   Vpn Id            : 0
Service Type      : VPLS
MACSec enabled    : no
Name              : VPLS 2
---snip---

Admin State       : Up                  Oper State        : Up
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:lag-1:2                              q-tag        1518    1518    Up   Down
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

```
*A:PE-3# show service id 1 sap

===============================================================================
SAP(Summary), Service 1
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
lag-1:1                         1          1     none    1     none   Up   Down
-------------------------------------------------------------------------------
Number of SAPs : 1
-------------------------------------------------------------------------------
===============================================================================


*A:PE-3# show service id 3 sap

===============================================================================
SAP(Summary), Service 3
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
lag-1:3                         3          1     none    1     none   Up   Down
-------------------------------------------------------------------------------
Number of SAPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

Conclusion

By default, the AC-DF capability is enabled. Disabling the AC-DF capability is recommended in ESs that use an operational group monitored by the access LAG to signal standby LACP or power-off.

ARP-ND Host Routes in Data Centers

This chapter provides information about ARP-ND Host Routes in Data Centers.

Topics in this chapter include:

Applicability

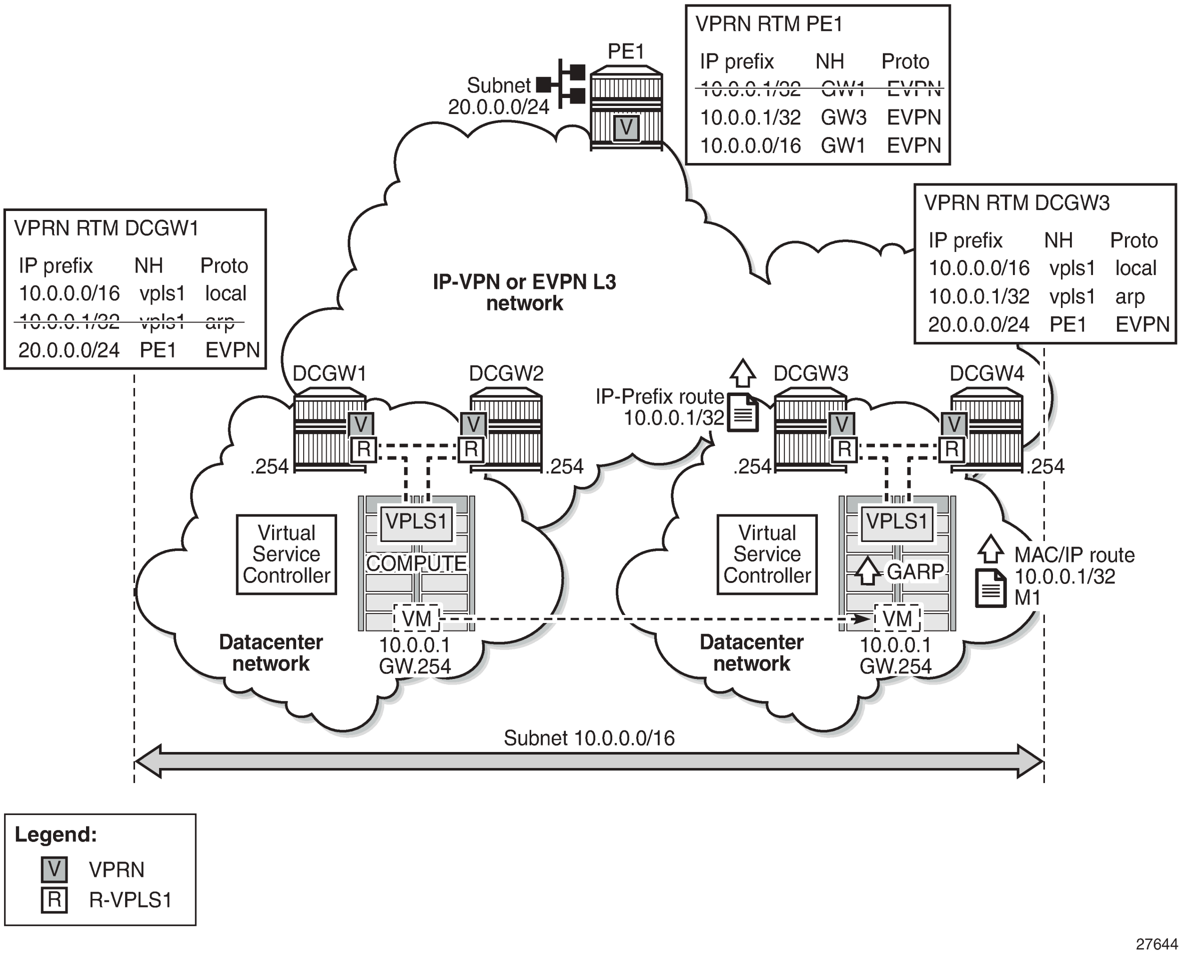
This chapter was initially written based on SR OS Release 16.0.R1, but the CLI in the current edition is based on SR OS Release 21.10.R3. Address Resolution Protocol - Neighbor Discovery (ARP-ND) host routes in VPRN and base router interfaces are supported in SR OS Release 15.0.R6 and later, but Nokia recommends using the feature in SR OS Release 15.0.R9, or later.

Chapters EVPN for MPLS Tunnels, EVPN for VXLAN Tunnels (Layer 2), EVPN for VXLAN Tunnels (Layer 3), and EVPN for MPLS Tunnels in Routed VPLS are prerequisite reading.

Overview

Inter-subnet forwarding (or simply routing) for a tenant domain in a Data Center (DC) must be efficient and avoid forwarding over the same path as arriving, known as tromboning or hairpinning. Figure 7: L2 broadcast domain extension across DCs shows an L2 broadcast domain (VPLS 1) extended across two DCs. This example is used to explain the requirement of upstream and downstream efficiency.

Figure 7: L2 broadcast domain extension across DCs



In Figure 7: L2 broadcast domain extension across DCs, subnet 10.0.0.0/16 is extended across two DCs and four DC Gateways (DCGWs), using VPLS 1 or R-VPLS 1 in the network nodes. The DCGWs are connected to the users of subnet 10.0.20.0/24 on PE1 via IP-VPN (or EVPN). In this scenario, there are two network characteristics that allow an efficient upstream and downstream routing:

- Anycast gateways
- ARP-ND host routes

Anycast Gateways provide upstream routing efficiency for the hosts connected to subnet 10.0.0.0/16, regardless of the DCGW to which they are connected. For example, if host 10.0.0.1 is in DC-1 and needs to forward traffic to subnet 10.0.20.0, DCGW1 and DCGW2 should be able to route the traffic upstream, without the need to go to DCGW3 or DCGW4. In the same way, if host 10.0.0.1 moves to DC-2, the upstream traffic to subnet 10.0.20.0 must be routed by the local DCGWs without changing the existing host default gateway IP and MAC configuration. To achieve this local default gateway routing, all the DCGWs of the extended broadcast domain need to have the same IP and MAC addresses in the R-VPLS interface (Integrated Routing and Bridging (IRB) interface in industry-standard terminology).

Anycast Gateways are implemented in SR OS by using passive VRRP. See the EVPN for MPLS Tunnels in Routed VPLS chapter for more information about passive VRRP.

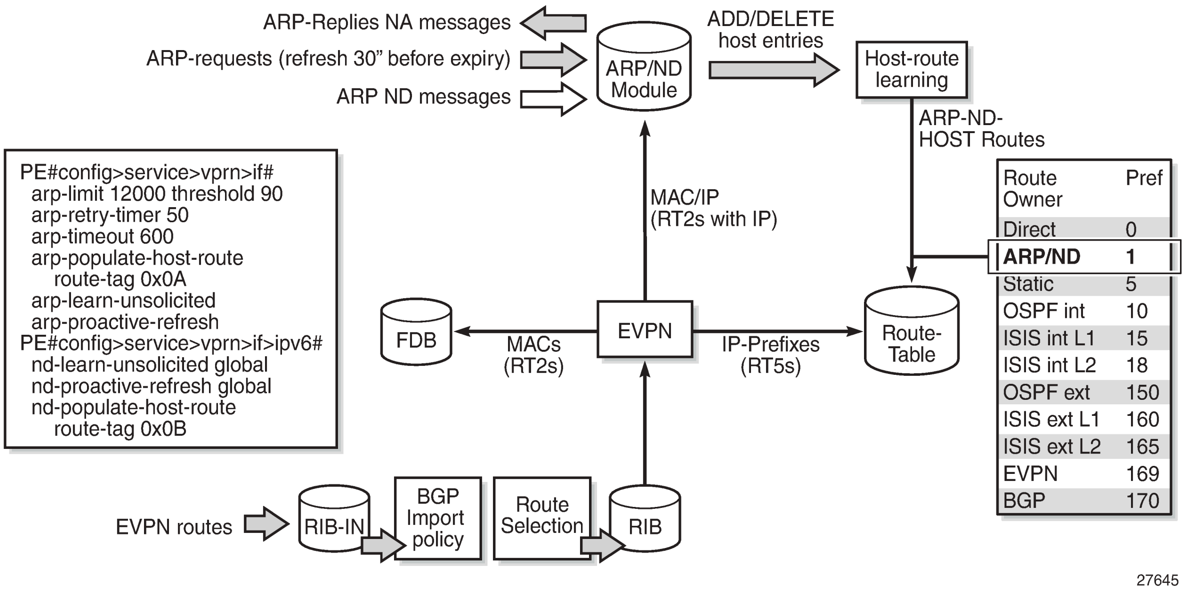
ARP-ND host routes learning and advertising are required to provide an efficient downstream routing from remote subnets to the hosts in the extended broadcast domain. Assuming virtual machine VM 10.0.0.1 (in Figure 7: L2 broadcast domain extension across DCs) is connected to DC-1 (left-side DC), when PE1 needs to send traffic to host 10.0.0.1, it will do a Longest Prefix Match (LPM) lookup on the VPRN route table. If the only IP prefix advertised by the four DCGWs were 10.0.0.0/16, PE1 could send the packets to a DC where the VM is not present. This would result in unnecessary tromboning; for example, PE1 could send the traffic to DCGW3, then DCGW3 would send it to DCGW2 to get to VM 10.0.0.1. However, PE1 could have forwarded directly to DCGW2.

To provide efficient downstream routing to the DC where the VM is located, DCGW1 and DCGW2 need to generate host routes for the VMs to which they are attached. Furthermore, when the VM moves to the other DC, DCGW3 and DCGW4 must be able to learn the VM host route and advertise it to PE1. Also, DCGW1 and DCGW2 will have to withdraw the route for 10.0.0.1, because the VM is no longer in the local DC.

To address this and other use cases, SR OS can learn the VM host route from the ARP or ND messages that it generates when it boots or when it moves. The host route can also be learned from EVPN routes type 2 (MAC/IP routes) that are installed in the ARP/ND caches, or in general, any ARP/ND entry can generate an ARP/ND host route.

A route owner type called "ARP-ND" is supported in the base router or a VPRN route table. The ARP-ND host routes have a preference of 1 and they are automatically created out of the ARP or ND Neighbor entries in the router instance. Figure 8: ARP-ND module and generated ARP-ND host routes shows how the ARP/ND software modules can generate ARP-ND host routes in the route table.

Figure 8: ARP-ND module and generated ARP-ND host routes



```
PE#config>service>vprn>if#
  arp-limit 12000 threshold 90
  arp-retry-timer 50
  arp-timeout 600
  arp-populate-host-route
     route-tag 0x0A
  arp-learn-unsolicited
  arp-proactive-refresh
PE#config>service>vprn>if>ipv6#
  nd-learn-unsolicited global
  nd-proactive-refresh global
  nd-populate-host-route
     route-tag 0x0B
```

Route Owner	Pref
Direct	0
ARP/ND	**1**
Static	5
OSPF int	10
ISIS int L1	15
ISIS int L2	18
OSPF ext	150
ISIS ext L1	160
ISIS ext L2	165
EVPN	169
BGP	170

When **config>service>vprn/ies>interface>arp-host-route>populate [static | dynamic | evpn]** is enabled, the static, dynamic, and EVPN ARP entries of the routing context will create ARP-ND host routes in the route table. In the same way, ARP-ND host routes are created in the IPv6 route table out of static, dynamic, and EVPN neighbor entries, if **config>service>vprn/ies>interface>ipv6>nd-host-route>populate [static | dynamic | evpn]** is enabled.

Figure 8: ARP-ND module and generated ARP-ND host routes shows how the ARP/ND module populates its database from the usual dynamic and static entries, as well as from EVPN routes type 2 that include an IP address. Through the host-route learning action, ARP-ND host routes are handed over to the route table.

Figure 8: ARP-ND module and generated ARP-ND host routes also shows that the preference assigned to ARP-ND host routes is 1, which means that ARP-ND routes will be preferred over any other route owner, except for direct routes. For example, if the same host route gets to the route table from ARP-ND and VPN-IPv4 or EVPN, the ARP-ND host route will be preferred and added to the route table. Although they are added to the route table and advertised to routing protocols, ARP-ND host routes are never installed in the FIB. That helps preserve the FIB scale in the router.

The **arp/nd-host-route populate [static | dynamic | evpn]** commands are typically used along with other features:

- A route tag can be added to ARP-ND hosts by the command **route-tag**. This tag can be matched on BGP **vrf-export** and peer export policies.
- The ARP-ND host route will be kept in the route table while the corresponding ARP or Neighbor entry is active. The commands **arp-proactive-refresh** and **nd-proactive-refresh** help keep the entries active (even if there is no traffic destined to them) by sending an ARP refresh 30 seconds before the **arp-timeout** or starting Neighbor Unreachable Detection (NUD) when the **stale-time** expires.
- To speed up the learning of the ARP-ND host routes, the commands **arp-learn-unsolicited** and **nd-learn-unsolicited** can be configured. When **arp-learn-unsolicited** is enabled, received unsolicited ARP messages (typically, Gratuitous Address Resolution Protocol (GARP) messages) create an ARP entry, and therefore an ARP-ND route if **arp-host-route>populate [static | dynamic | evpn]** is added. Similarly, unsolicited Neighbor Advertisement messages will create a "stale" neighbor. If **nd-host-route>populate [static | dynamic | evpn]** is enabled, a confirmation message (NUD) is sent for all the neighbor entries created as stale, and, if confirmed, the corresponding ARP-ND routes are added to the route table.

In the example of Figure 7: L2 broadcast domain extension across DCs, **arp-host-route>populate [static | dynamic | evpn]** on the DCGWs allows them to learn/advertise the ARP-ND host route 10.0.0.1/32 when the VM is locally connected, and remove/withdraw it when the VM is no longer present in the local DC.

The following sections describe three typical DC scenarios in which the use of Anycast gateways and ARP-ND host routes is needed. The examples are focused on IPv4 and ARP; however, there is equivalent functionality for IPv6 and ND.

Configuration

The initial configuration includes the following:

- Cards, MDAs, ports
- Router interfaces
- IS-IS as an IGP

The following three scenarios are configured and presented in this document:

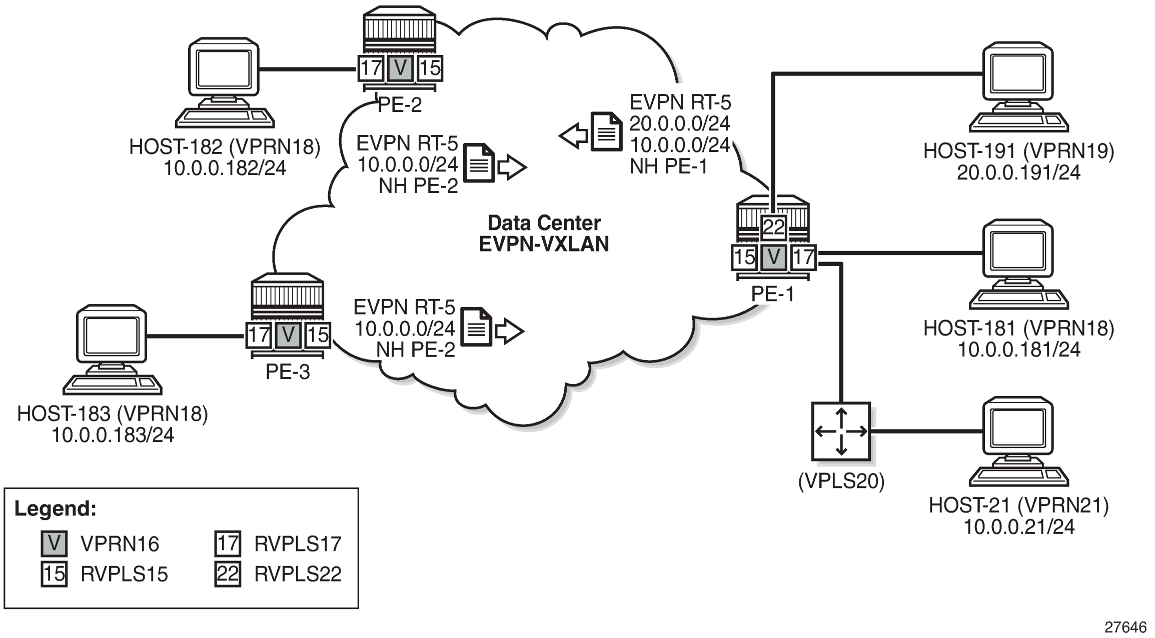
- DC inter-subnet forwarding with Anycast GWs (and no ARP-ND hosts)
- DC inter-subnet forwarding with Anycast GWs and ARP-ND hosts
- Data Center Interconnect (DCI) inter-subnet forwarding with Anycast GWs and ARP-ND hosts

DC inter-subnet forwarding with Anycast GWs

Figure 9: DC inter-subnet forwarding with Anycast GWs shows a typical DC network, where PE-1, PE-2, and PE-3 are leaf switches that use EVPN-VXLAN services to provide connectivity between two subnets of a tenant domain. Those two subnets are 10.0.0.0/24 and 10.0.20.0/24, respectively, and while the three PEs are attached to hosts in the 10.0.0.0/24 subnet, only PE-1 is attached to the 10.0.20.0/24 subnet. Subnet 10.0.0.0/24 uses R-VPLS 17 in the three PEs and subnet 10.0.20.0/24 uses R-VPLS 22 in PE-1. The distribution of the R-VPLS services does not have to be uniform in all the PEs, and those R-VPLS services are only created if there are hosts attached to them.

To provide inter-subnet forwarding for the tenant, each PE must be configured with a VPRN instance (VPRN 16) that has an interface to the subnet R-VPLS. In industry-standard terms, VPRN 16 represents the IP-VRF for the tenant, and R-VPLS 17 and R-VPLS 22 are user Broadcast Domains (BDs). R-VPLS 15 is not a user BD, but rather a backhaul R-VPLS that provides EVPN connectivity among the VPRN instances.

Figure 9: DC inter-subnet forwarding with Anycast GWs



The BGP configuration in the PEs is similar. As an example, the BGP configuration in PE-1 is as follows:

```
# on PE-1:
configure
    router Base
        bgp
            family evpn
            vpn-apply-import
            vpn-apply-export
            rapid-withdrawal
            rapid-update evpn
            group "dc"
                type internal
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
            no shutdown
        exit
```

PE-2 has the following service configuration. The service configuration on PE-3 is similar.

```
# on PE-2:
configure
    service
        vpls 15 name "sbd-15" customer 1 create
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 15 create
            exit
            bgp
            exit
            bgp-evpn
                no mac-advertisement
                ip-route-advertisement
                evi 15
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            no shutdown
        exit
        vprn 16 name "ip-vrf-16" customer 1 create
            ecmp 2
            interface "evi-15" create
                mac 00:00:00:00:00:02
                vpls "sbd-15"
                    evpn-tunnel
                exit
            exit
            interface "evi-17" create
                address 10.0.0.2/24
                vrrp 1 passive
                    backup 10.0.0.254
                    ping-reply
                    traceroute-reply
                exit
                vpls "evi-17"
                exit
            exit
            no shutdown
        exit
        vpls 17 name "evi-17" customer 1 create
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 17 create
            exit
            bgp
            exit
            bgp-evpn
                evi 17
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            sap pxc-10.a:17 create
```

```
                no shutdown
            exit
            no shutdown
        exit
```

R-VPLS 17, "evi-17" in the configuration, is the BD used by subnet 10.0.0.0/24 in all the PEs. On the evi-17 interface in VPRN 16, a real IP address as well as a virtual (passive VRRP) IP address are configured. The real IP address is a unique address across the three PEs in R-VPLS 17 (10.0.0.2 in PE-2). This IP address will not be used by the R-VPLS 17 hosts as a default gateway, but rather will be used for troubleshooting purposes (ICMP or similar).

The backup IP address in the passive VRRP instance (10.0.0.254) is the Anycast GW IP address, and the same IP address is configured in all the PEs attached to R-VPLS 17. Because the virtual MAC is auto-derived from the VRRP instance, all the PEs will also have the same virtual MAC for this Anycast GW:

```
*A:PE-2# show router 16 vrrp instance interface "evi-17"

===============================================================================
VRRP Instances for interface "evi-17"
===============================================================================
-------------------------------------------------------------------------------
VRID 1
-------------------------------------------------------------------------------
Owner               : No                  VRRP State        : Master
Passive             : Yes
Primary IP of Master: 10.0.0.2 (Self)
Primary IP          : 10.0.0.2            Standby-Forwarding: Disabled
VRRP Backup Addr    : 10.0.0.254
Admin State         : Up                  Oper State        : Up
Up Time             : 02/18/2022 14:52:45 Virt MAC Addr     : 00:00:5e:00:01:01
---snip---
```

```
*A:PE-3# show router 16 vrrp instance interface "evi-17"

===============================================================================
VRRP Instances for interface "evi-17"
===============================================================================
-------------------------------------------------------------------------------
VRID 1
-------------------------------------------------------------------------------
Owner               : No                  VRRP State        : Master
Passive             : Yes
Primary IP of Master: 10.0.0.3 (Self)
Primary IP          : 10.0.0.3            Standby-Forwarding: Disabled
VRRP Backup Addr    : 10.0.0.254
Admin State         : Up                  Oper State        : Up
Up Time             : 02/18/2022 14:52:53 Virt MAC Addr     : 00:00:5e:00:01:01
---snip---
```

```
*A:PE-1# show router 16 vrrp instance interface "evi-17"

===============================================================================
VRRP Instances for interface "evi-17"
===============================================================================
-------------------------------------------------------------------------------
VRID 1
-------------------------------------------------------------------------------
Owner               : No                  VRRP State        : Master
Passive             : Yes
Primary IP of Master: 10.0.0.1 (Self)
Primary IP          : 10.0.0.1            Standby-Forwarding: Disabled
VRRP Backup Addr    : 10.0.0.254
Admin State         : Up                  Oper State        : Up
Up Time             : 02/18/2022 14:52:38 Virt MAC Addr     : 00:00:5e:00:01:01
---snip---
```

All the hosts attached to R-VPLS 17, such as host-181, host-182, and host-183, are configured with the Anycast GW as default gateway (10.0.0.254). The use of passive VRRP (or Anycast GW in standard terminology) has the following benefits:

- All the hosts use the same default gateway configuration, regardless of what PE they are attached to.
- When the hosts send traffic destined to a remote subnet, the local PE can route it directly, without any tromboning.
- In the case of a host moving to a different leaf switch, the host does not need to change its IP or default gateway, or even its ARP cache.

For completeness, the service configuration in PE-1 follows:

```
# on PE-1:
configure
    service
        vpls 15 name "sbd-15" customer 1 create
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 15 create
            exit
            bgp
            exit
            bgp-evpn
                no mac-advertisement
                ip-route-advertisement
                evi 15
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
        vprn 16 name "ip-vrf-16" customer 1 create
            ecmp 2
            interface "evi-15" create
                mac 00:00:00:00:00:01
                vpls "sbd-15"
                    evpn-tunnel
                exit
            exit
            interface "evi-17" create
                address 10.0.0.1/24
                vrrp 1 passive
                    backup 10.0.0.254
                    ping-reply
                    traceroute-reply
                exit
                vpls "evi-17"
                exit
            exit
            interface "evi-22" create
                address 10.0.20.1/24
                vrrp 1 passive
                    backup 10.0.20.254
                    ping-reply
                    traceroute-reply
                exit
                vpls "evi-22"
                exit
            exit
            no shutdown
        exit
        vpls 17 name "evi-17" customer 1 create
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 17 create
            exit
            bgp
            exit
            bgp-evpn
                evi 17
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap pxc-10.a:17 create
                no shutdown
            exit
            sap pxc-10.b:20 create
                no shutdown
            exit
            no shutdown
        exit
        vpls 22 name "evi-22" customer 1 create
            allow-ip-int-bind
            exit
            stp
                shutdown
            exit
            sap pxc-10.b:19 create
                no shutdown
            exit
            no shutdown
        exit
```

See the EVPN for VXLAN Tunnels (Layer 3) chapter for more information about the EVPN-related configuration in the R-VPLS services. When there is no need for a recursive resolution of the EVPN IP prefix routes to a MAC/IP route, **no mac-advertisement** is used in the R-VPLS 15, compared to the examples in EVPN for VXLAN Tunnels (Layer 3).

With the described configuration, as an example, the intra-subnet and inter-subnet forwarding connectivity from host-182 is tested (host-182 is simulated with VPRN 18 that is connected to R-VPLS 17 via PXC SAP):

```
*A:PE-2# traceroute router 18 10.0.0.183 source 10.0.0.182
traceroute to 10.0.0.183 from 10.0.0.182, 30 hops max, 40 byte packets
  1  10.0.0.183 (10.0.0.183)    2.49 ms  2.31 ms  2.41 ms
```

```
*A:PE-2# traceroute router 18 10.0.20.191 source 10.0.0.182
traceroute to 10.0.20.191 from 10.0.0.182, 30 hops max, 40 byte packets
  1  10.0.0.2 (10.0.0.2)    0.979 ms  0.890 ms  0.875 ms
  2  10.0.20.1 (10.0.20.1)    2.02 ms  1.98 ms  1.92 ms
  3  10.0.20.191 (10.0.20.191)    2.61 ms  2.73 ms  2.74 ms
```

When host-182 sends traffic to host-191, it will ARP for the Anycast GW IP and will receive the virtual MAC as a reply. The virtual MAC is always associated with the local CPM on the local PE; therefore, the local PE can always route the traffic directly while it has a route for the IP destination.

Host-182 (VPRN 18) resolves the Anycast GW to the virtual MAC:

```
*A:PE-2# show router 18 arp 10.0.0.254

===============================================================================
ARP Table (Service: 18)
===============================================================================
IP Address      MAC Address       Expiry    Type   Interface
-------------------------------------------------------------------------------
10.0.0.254      00:00:5e:00:01:01 03h59m18s Dyn[I] local
===============================================================================
```

In PE-2, the virtual MAC is associated with a local IP interface:

```
*A:PE-2# show service id 17 fdb mac 00:00:5e:00:01:01

===============================================================================
Forwarding Database, Service 17
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
17         00:00:5e:00:01:01 cpm                     Intf     02/18/22 14:52:45
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The following route table of VPRN 16 on PE-2 shows that subnet 10.0.20.0/24 from host-191 is learned via EVPN:

```
*A:PE-3# show router 16 route-table

===============================================================================
Route Table (Service: 16)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.0.0/24                                   Local   Local     00h03m08s  0
       evi-17                                                       0
10.0.20.0/24                                  Remote  EVPN-IFF  00h03m07s  169
       evi-15 (ET-00:00:00:00:00:01)                                0
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

DC inter-subnet forwarding with Anycast GWs and ARP-ND host routes

While the configuration shown in the preceding section is common in DCs, there is a variation that eliminates the flooding among PEs that are attached to the same BD, typically caused by ARP messages and ND. The configuration described in this section is recommended only if all the following conditions are met:

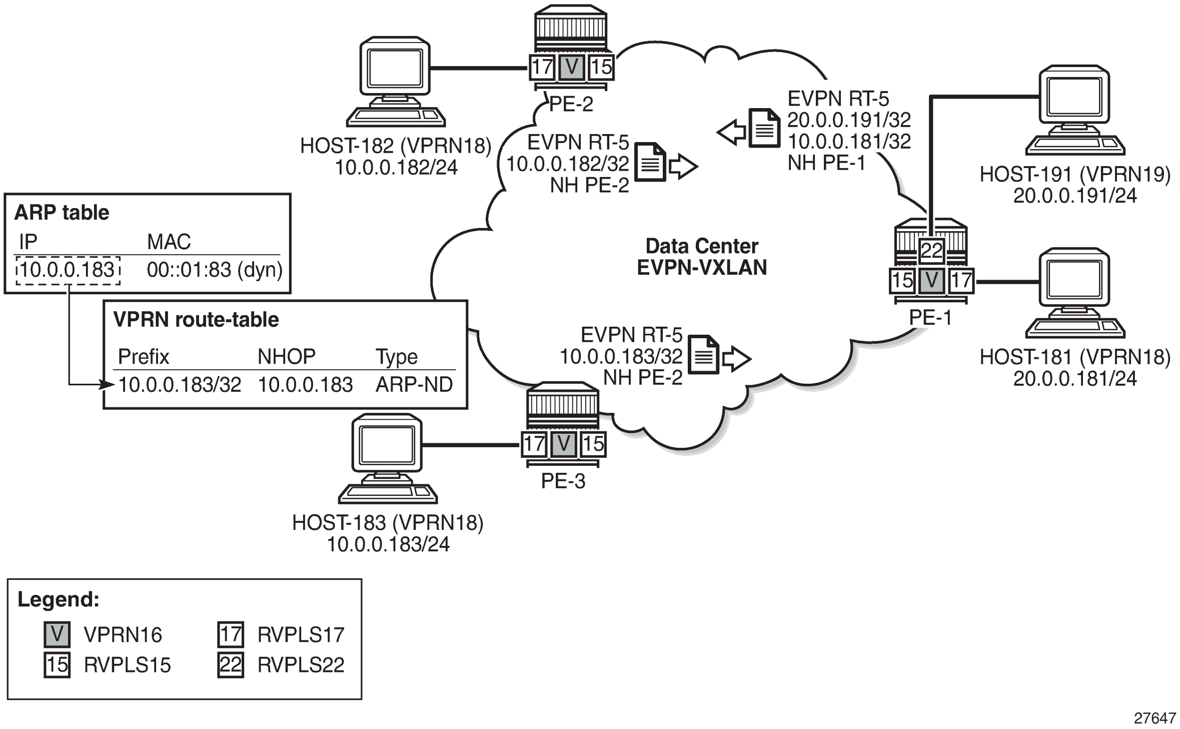
- All the hosts are directly connected to the leaf switches (PEs in Figure 9: DC inter-subnet forwarding with Anycast GWs).
- All the hosts announce themselves by issuing a GARP (or unsolicited NA for IPv6) whenever they boot up or move to a different leaf switch.

  Note: This is the case for virtual machines.
- All the traffic among hosts is IP unicast or non-IP unicast (if the hosts are in the same BD), and there is no Broadcast, Unknown unicast, or Multicast (BUM) traffic from the hosts in the tenant domain, other than ARP/ND.

If the preceding conditions are true, the ARP-ND host route feature can help eliminate BUM traffic completely.

Figure 10: DC inter-subnet forwarding with Anycast GWs and ARP-ND host routes shows the scenario used in this section.

Figure 10: DC inter-subnet forwarding with Anycast GWs and ARP-ND host routes



Compared to the configuration used in the preceding section, VPRN 16 is modified in the three PEs as follows (changes in bold):

```
# on PE-2:
configure
    service
        vprn "ip-vrf-16"
            ecmp 2
            interface "evi-15" create
                mac 00:00:00:00:00:02
                vpls "sbd-15"
                    evpn-tunnel
                exit
            exit
            interface "evi-17" create
                address 10.0.0.2/24
                arp-host-route
                    populate static
                    populate dynamic
                    populate evpn
                exit
                arp-timeout 300
                arp-learn-unsolicited
                arp-proactive-refresh
                vrrp 1 passive
                    backup 10.0.0.254
                    ping-reply
                    traceroute-reply
                exit
                remote-proxy-arp
                vpls "evi-17"
                exit
            exit
            no shutdown
```

```
# on PE-3:
configure
    service
        vprn "ip-vrf-16"
            ecmp 2
            interface "evi-15" create
                mac 00:00:00:00:00:03
                vpls "sbd-15"
                    evpn-tunnel
                exit
            exit
            interface "evi-17" create
                address 10.0.0.3/24
                arp-host-route
                    populate static
                    populate dynamic
                    populate evpn
                exit
                arp-timeout 300
                arp-learn-unsolicited
                arp-proactive-refresh
                vrrp 1 passive
                    backup 10.0.0.254
                    ping-reply
                    traceroute-reply
                exit
                remote-proxy-arp
                vpls "evi-17"
                exit
            exit
            no shutdown
```

```
# on PE-1:
configure
    service
        vprn "ip-vrf-16"
            ecmp 2
            interface "evi-15" create
                mac 00:00:00:00:00:01
                vpls "sbd-15"
                    evpn-tunnel
                exit
            exit
            interface "evi-17" create
                address 10.0.0.1/24
                arp-host-route
                    populate static
                    populate dynamic
                    populate evpn
                exit
                arp-timeout 300
                arp-learn-unsolicited
                arp-proactive-refresh
                vrrp 1 passive
                    backup 10.0.0.254
                    ping-reply
                    traceroute-reply
                exit
                remote-proxy-arp
                vpls "evi-17"
                exit
            exit
            interface "evi-22" create
                address 10.0.20.1/24
                arp-host-route
                    populate static
                    populate dynamic
                    populate evpn
                exit
                arp-timeout 300
                arp-learn-unsolicited
                arp-proactive-refresh
                vrrp 1 passive
                    backup 10.0.20.254
                    ping-reply
                    traceroute-reply
                exit
                remote-proxy-arp
                vpls "evi-22"
                exit
            exit
            no shutdown
```

The behavior due to the newly added commands is as follows:

- **arp-host-route>populate [static | dynamic | evpn]** makes the router create an ARP-ND host route per ARP entry in the route table of VPRN "ip-vrf-16".
- **arp-learn-unsolicited** makes the router learn ARP entries for the hosts out of the GARP messages that they send when they boot up or move. Without this command, ARP entries are only created after the

router receives packets with the host as the destination, issues an ARP request, and the host replies to this solicited ARP request.

- **arp-proactive-refresh** makes the router refresh every dynamic ARP entry even if there is no traffic destined to the owner. Without the command, host IP addresses will not be maintained in the ARP cache unless they receive traffic from remote hosts.
- **arp-timeout 300** is the timeout selected in this example (in seconds). The ARP timeout has an impact on how often the router will try to refresh an entry (30 seconds before the timeout expires). In environments where the hosts are subject to mobility (VMs moving between leaves), having a shorter ARP timeout will speed up the removal of the old ARP entry, that is, the old ARP-ND host route entry. However, in scaled environments with tens of thousands of ARP entries, Nokia does not recommend lowering the ARP timeout under 10 minutes.
- **remote-proxy-arp** allows the router to reply to any ARP request looking for an IP address in the same subnet as the source, with its virtual MAC (00:00:5e:00:01:01), and route the traffic, as long as there is a route for the destination in the route table.

In addition, the following commands will be executed in the three PEs:

```
# on PE-1, PE-2, PE-3:
configure
    service
        vpls "evi-17"
            bgp-evpn
                vxlan
                    shutdown
                exit
                no ingress-repl-inc-mcast-advertisement
                vxlan
                    no shutdown
                exit
            exit
```

By disabling the advertisement of the Inclusive Multicast Ethernet Tag (IMET) route in R-VPLS 17, the PEs will not create a VXLAN BUM destination among each other, preventing the exchange of BUM traffic. Only known unicast traffic can be now exchanged in the context of R-VPLS 17. The three PEs will show VXLAN destinations that have Mcast "-", as opposed to "BUM":

```
*A:PE-3# show service id 17 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
Instance    VTEP Address                            Egress VNI  EvpnStatic Num
 Mcast       Oper State                              L2 PBR     SupBcasDom MACs
-------------------------------------------------------------------------------
1           192.0.2.1                               17          evpn       3
 -           Up                                      No          No
1           192.0.2.2                               17          evpn       2
 -           Up                                      No          No
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 2
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
```

```
No Matching Entries
===============================================================================
```

With the described configuration, when the hosts boot up and generate a GARP message, the ARP entries will be created, and subsequently ARP-ND hosts and EVPN IP-prefix advertisements for them. The host bootup is simulated by disabling and re-enabling the VPRN that emulates the host. As an example, some debug commands are used to see the behavior when host-181 boots up and sends a GARP:

```
*A:PE-1# configure service vprn 18 shutdown
*A:PE-1# configure service vprn 18 no shutdown

1 2022/02/18 14:58:30.128 UTC MINOR: DEBUG #2001 vprn18 PIP
"PIP: ARP
instance 3 (18), interface index 7 (local),
ARP egressing on local
   Who has 10.0.0.181 ? Tell 10.0.0.181
"

2 2022/02/18 14:58:30.129 UTC MINOR: DEBUG #2001 vprn16 PIP
"PIP: ARP
instance 2 (16), interface index 5 (evi-17),
ARP ingressing on evi-17
   Who has 10.0.0.181 ? Tell 10.0.0.181
"

4 2022/02/18 14:58:30.129 UTC MINOR: DEBUG #2001 vprn21 PIP
"PIP: ARP
instance 5 (21), interface index 9 (local),
ARP ingressing on local
   Who has 10.0.0.181 ? Tell 10.0.0.181
"
```

The GARP creates an ARP entry and, subsequently, an ARP-ND host route in the route table of VPRN 16. Host-181 MAC/IP and IP-prefix routes are advertised too:

```
3 2022/02/18 14:58:30.129 UTC MINOR: DEBUG #2001 vprn16 PIP
"PIP: ROUTE
instance 2 (16), RTM ADD event
   New Route Info
      prefix: 10.0.0.181/32 (0x11952c690)  preference: 1   metric: 0
              backup metric: 0  owner: ARP-ND ownerId: 0
      1 ecmp hops  0 backup hops:
         hop 0:  10.0.0.181 @ if 5, weight 0
"

5 2022/02/18 14:58:30.129 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 81
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.1
        Type: EVPN-MAC Len: 33 RD: 192.0.2.1:17 ESI: ESI-0, tag: 0, mac len: 48
              mac: 00:00:00:00:01:81, IP len: 0, IP: NULL, label1: 17
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:17
        bgp-tunnel-encap:VXLAN
"
```

```
6 2022/02/18 14:58:30.129 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 90
    Flag: 0x90 Type: 14 Len: 45 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.1
        Type: EVPN-IP-PREFIX Len: 34 RD: 192.0.2.1:15, tag: 0,
                   ip_prefix: 10.0.0.181/32 gw_ip 0.0.0.0 Label: 15 (Raw Label: 0xf)
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:15
        mac-nh:00:00:00:00:00:01
        bgp-tunnel-encap:VXLAN
"
```

As an example, following are the ARP and route tables in PE-1:

```
*A:PE-1# show router 16 arp

===============================================================================
ARP Table (Service: 16)
===============================================================================
IP Address      MAC Address       Expiry    Type   Interface
-------------------------------------------------------------------------------
10.0.0.1        00:00:00:00:1e:17 00h00m00s Oth[I] evi-17
10.0.0.2        00:00:00:00:2e:17 00h00m00s Evp[I] evi-17
10.0.0.3        00:00:00:00:3e:17 00h00m00s Evp[I] evi-17
10.0.0.181      00:00:00:00:01:81 00h03m09s Dyn[I] evi-17
10.0.0.254      00:00:5e:00:01:01 00h00m00s Oth[I] evi-17
10.0.20.1       00:00:00:00:1e:22 00h00m00s Oth[I] evi-22
10.0.20.254     00:00:5e:00:01:01 00h00m00s Oth[I] evi-22
-------------------------------------------------------------------------------
No. of ARP Entries: 7
===============================================================================
```

```
*A:PE-1# show router 16 route-table

===============================================================================
Route Table (Service: 16)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.0.0/24                                   Local   Local     00h03m09s  0
       evi-17                                                       0
10.0.0.2/32                                   Remote  ARP-ND    00h03m09s  1
       10.0.0.2                                                     0
10.0.0.3/32                                   Remote  ARP-ND    00h03m09s  1
       10.0.0.3                                                     0
10.0.0.181/32                                 Remote  ARP-ND    00h02m55s  1
       10.0.0.181                                                   0
10.0.0.182/32                                 Remote  EVPN-IFF  00h02m58s  169
       evi-15 (ET-00:00:00:00:00:02)                                0
10.0.0.183/32                                 Remote  EVPN-IFF  00h02m56s  169
       evi-15 (ET-00:00:00:00:00:03)                                0
10.0.20.0/24                                  Local   Local     00h03m09s  0
       evi-22                                                       0
-------------------------------------------------------------------------------
```

```
No. of Routes: 7
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

As discussed, the ARP-ND host routes are installed in the route table, but not in the FIB:

```
*A:PE-1# show router 16 fib 1

===============================================================================
FIB Display
===============================================================================
Prefix [Flags]                                              Protocol
  NextHop
-------------------------------------------------------------------------------
10.0.0.0/24                                                 LOCAL
  10.0.0.0 (evi-17)
10.0.0.182/32                                               EVPN-IFF
  (evi-15 (ET-00:00:00:00:00:02))
10.0.0.183/32                                               EVPN-IFF
  (evi-15 (ET-00:00:00:00:00:03))
10.0.20.0/24                                                LOCAL
  10.0.20.0 (evi-22)
-------------------------------------------------------------------------------
Total Entries : 4
-------------------------------------------------------------------------------
===============================================================================
```

A side effect of this scenario is that traffic between hosts in the same BD (R-VPLS 17) is routed instead of switched. This can be shown on the traceroute from host-181 to host-182 (there are three hops instead of two), or the TTL on the ping packets (62 instead of 64):

```
*A:PE-1# traceroute router 18 10.0.0.182
traceroute to 10.0.0.182, 30 hops max, 40 byte packets
  1  10.0.0.1 (10.0.0.1)    2.02 ms  2.29 ms  2.26 ms
  2  10.0.0.2 (10.0.0.2)    3.35 ms  3.39 ms  3.17 ms
  3  10.0.0.182 (10.0.0.182)    4.10 ms  3.95 ms  3.56 ms
```

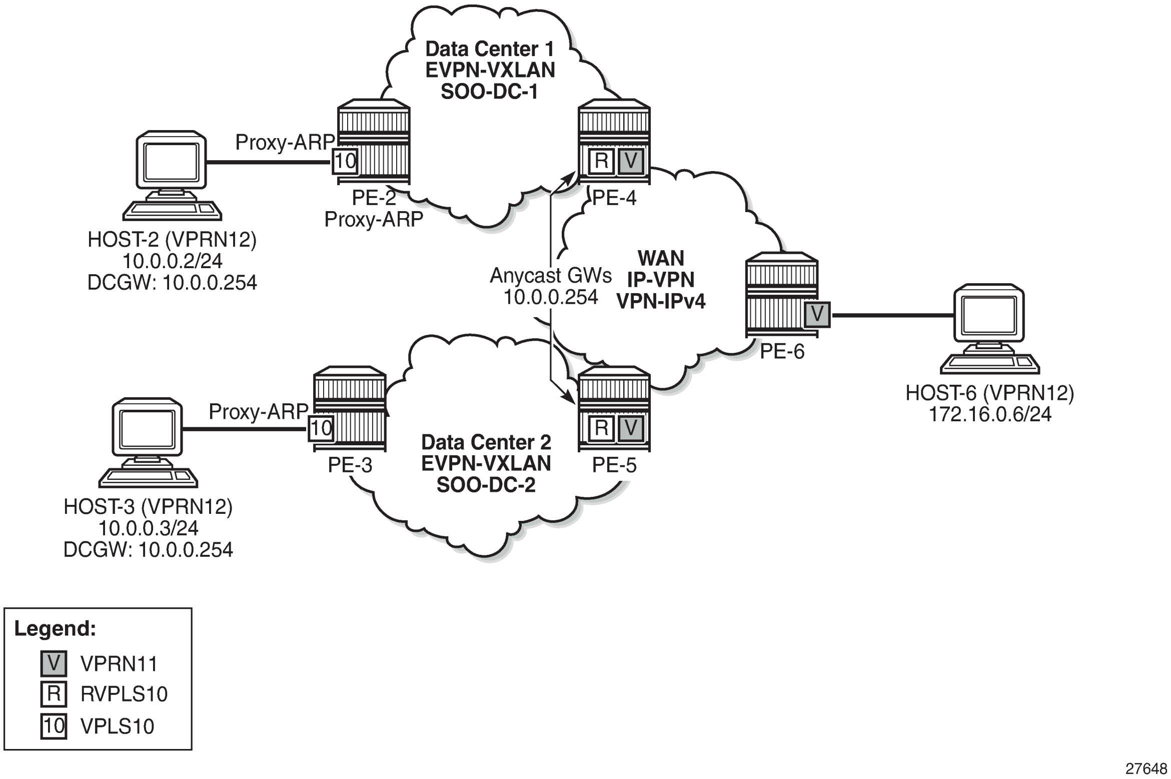
```
*A:PE-1# ping router 18 10.0.0.182 source 10.0.0.181
PING 10.0.0.182 56 data bytes
64 bytes from 10.0.0.182: icmp_seq=1 ttl=62 time=3.25ms.
64 bytes from 10.0.0.182: icmp_seq=2 ttl=62 time=3.49ms.
64 bytes from 10.0.0.182: icmp_seq=3 ttl=62 time=3.41ms.
64 bytes from 10.0.0.182: icmp_seq=4 ttl=62 time=3.49ms.
64 bytes from 10.0.0.182: icmp_seq=5 ttl=62 time=3.54ms.

---- 10.0.0.182 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 3.25ms, avg = 3.43ms, max = 3.54ms, stddev = 0.101ms
```

This extension of a subnet across a pure routing domain is compliant with the virtual subnet concept described in RFC 7814.

DCI inter-subnet forwarding with Anycast GWs and ARP-ND hosts

Figure 11: DCI inter-subnet forwarding with Anycast GWs and ARP-ND host routes shows a DCI scenario where the use of Anycast GWs, ARP-ND hosts, and some additional configuration provide efficient inter-subnet forwarding within the tenant domain.

Figure 11: DCI inter-subnet forwarding with Anycast GWs and ARP-ND host routes



In this example, VPLS 10 is extended across DC-1 and DC-2, via PE-4 and PE-5 (which are DC GWs). PE-4 and PE-5 are also connected to the WAN and use IP-VPN for inter-subnet forwarding connectivity to the remote host-6. In this network, PE-4 and PE-5 provide the Anycast GW functionality to host-2 and host-3, so that they can move between the two DCs without having to change their IP/MAC/default GW or ARP cache, and efficient upstream forwarding is provided.

PE-4 and PE-5 learn the ARP-ND host route of their respective host and advertise it to the WAN, so that downstream routing from PE-6 can be efficient and without tromboning.

To avoid unnecessary ARP flooding between DCs, proxy-ARP is used in PE-2 and PE-3. The configuration of VPLS 10 in the PE-2 and PE-3 is as follows:

```
# on PE-2:
configure
    service
        vpls 10 name "centralized-gw-bd" customer 1 create
            vxlan instance 1 vni 10 create
            exit
            bgp
            exit
            bgp-evpn
                evi 10
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap pxc-10.a:10 create
                no shutdown
            exit
            proxy-arp
                send-refresh 120
                no unknown-arp-request-flood-evpn
                dynamic-arp-populate
                no garp-flood-evpn
                evpn-route-tag 1
                no shutdown
            exit
            no shutdown
        exit
```

```
# on PE-3:
configure
    service
       vpls 10 name "centralized-gw-bd" customer 1 create
            vxlan instance 1 vni 10 create
            exit
            bgp
            exit
            bgp-evpn
                evi 10
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap pxc-10.a:10 create
                no shutdown
            exit
            proxy-arp
                send-refresh 120
                no unknown-arp-request-flood-evpn
                dynamic-arp-populate
                no garp-flood-evpn
                evpn-route-tag 1
                no shutdown
            exit
            no shutdown
        exit
```

Because the hosts are directly connected to PE-2 and PE-3, and they announce themselves to the network through a GARP when they boot up or move, the proxy-ARP configuration includes the parameters **no unknown-arp-request-flood-evpn** and **no garp-flood-evpn**. Those two commands prevent unnecessary ARP flooding between DCs.

The two PEs also include the **proxy-arp evpn-route-tag 1** command. This command allows the proxy-ARP module to tag the routes when sent to BGP for advertisement of a MAC/IP route with non-zero IP. In this example, the tag is used in an export policy to add a Site-Of-Origin (SOO) extended community to the

MAC/IP routes with non-zero IP. This, for example, allows PE-4 to accept MAC/IP routes from its own DC-1 and drop MAC/IP routes from DC-2 so that PE-4 only advertises ARP-ND host routes attached to DC-1. Vice versa for PE-5. The MAC/IP routes with zero-IP (that are also sent for every MAC) will not be tagged with the SOO and, therefore, will be imported by all the PEs in VPLS 10. This allows normal L2 connectivity among the four PEs, while the ARP-ND routes are only generated for the local hosts.

On PE-2, BGP is configured as follows:

```
# on PE-2:
configure
    router Base
        autonomous-system 64500
        policy-options
            begin
            community "SOO-DC-1"
                members "origin:64500:1"
            exit
            policy-statement "export-add-SOO"
                entry 10
                    from
                        tag 1
                    exit
                    action accept
                        community add "SOO-DC-1"
                    exit
                exit
            exit
            policy-statement "import-prefer-DC-1"
                entry 10
                    from
                        community "SOO-DC-1"
                    exit
                    action accept
                        local-preference 200
                    exit
                exit
            exit
            commit
        exit
        bgp
            family vpn-ipv4 vpn-ipv6 evpn
            vpn-apply-import
            vpn-apply-export
            import "import-prefer-DC-1"
            export "export-add-SOO"
            rapid-withdrawal
            rapid-update evpn
            group "dc"
                type internal
                neighbor 192.0.2.3
                exit
            exit
            group "dcgws"
                type internal
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
        exit
```

On PE-3, BGP is configured as follows:

```
# on PE-3:
configure
    router
        autonomous-system 64500
        policy-options
            begin
            community "SOO-DC-2"
                members "origin:64500:2"
            exit
            policy-statement "export-add-SOO"
                entry 10
                    from
                        tag 1
                    exit
                    action accept
                        community add "SOO-DC-2"
                    exit
                exit
            exit
            policy-statement "import-prefer-DC-2"
                entry 10
                    from
                        community "SOO-DC-2"
                    exit
                    action accept
                        local-preference 200
                    exit
                exit
            exit
            commit
        exit
        bgp
            family vpn-ipv4 vpn-ipv6 evpn
            vpn-apply-import
            vpn-apply-export
            import "import-prefer-DC-2"
            export "export-add-SOO"
            rapid-withdrawal
            rapid-update evpn
            group "dc"
                type internal
                neighbor 192.0.2.2
                exit
            exit
            group "dcgws"
                type internal
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
        exit
```

As an example, the following **show** commands prove that PE-2 does not add an SOO to MAC/IP routes with zero-IP, but it does add SOO-DC-1 for MAC/IP routes with non-zero IP:

```
*A:PE-2# show router bgp routes evpn mac rd 192.0.2.2:10 hunt
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
---snip---

Network        : n/a
Nexthop        : 192.0.2.2
Path Id        : None
To             : 192.0.2.3
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : n/a
Connector      : None
Community      : origin:64500:1 target:64500:10
                 bgp-tunnel-encap:VXLAN
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.3
Origin         : IGP
AS-Path        : No As-Path
EVPN type      : MAC
ESI            : ESI-0
Tag            : 0
IP Address     : 10.0.0.2
Route Dist.    : 192.0.2.2:10
Mac Address    : 00:00:00:00:00:02
MPLS Label1    : VNI 10                 MPLS Label2    : n/a
Route Tag      : 0
Neighbor-AS    : n/a
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0

Network        : n/a
Nexthop        : 192.0.2.2
Path Id        : None
To             : 192.0.2.3
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : n/a
Connector      : None
Community      : target:64500:10 bgp-tunnel-encap:VXLAN
                 mac-mobility:Seq:0/Static
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.3
Origin         : IGP
AS-Path        : No As-Path
EVPN type      : MAC
ESI            : ESI-0
Tag            : 0
IP Address     : n/a
```

```
Route Dist.    : 192.0.2.2:10
Mac Address    : 02:13:ff:00:03:3a
MPLS Label1    : VNI 10                 MPLS Label2    : n/a
Route Tag      : 0
Neighbor-AS    : n/a
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0

---snip---
```

The VPLS 10 configuration on PE-4 and the corresponding import policy to drop non-local SOO follow. PE-5 has a similar configuration (not shown), including the same RD 64500:10 in VPLS 10 as PE-4. The policy will drop routes tagged with SOO-DC-1 instead of SOO-DC-2.

```
# on PE-4:
configure
    service
        vpls 10 name "centralized-gw-bd" customer 1 create
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 10 create
            exit
            bgp
                route-distinguisher 64500:10
            exit
            bgp-evpn
                evi 10
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
```

On PE-4, the BGP configuration is as follows:

```
# on PE-4:
configure
    router Base
        autonomous-system 64500
        policy-options
            begin
            community "SOO-DC-1"
                members "origin:64500:1"
            exit
            community "SOO-DC-2"
                members "origin:64500:2"
            exit
            policy-statement "export-add-SOO"
                entry 10
                    from
                    exit
                    action accept
                        community add "SOO-DC-1"
                    exit
                exit
            exit
            policy-statement "import-drop-DC-2"
                entry 10
```

```
                    from
                        community "SOO-DC-2"
                    exit
                    action drop
                    exit
                exit
            exit
            commit
        exit
        bgp
            family vpn-ipv4 vpn-ipv6 evpn
            vpn-apply-import
            vpn-apply-export
            import "import-drop-DC-2"
            export "export-add-SOO"
            rapid-withdrawal
            rapid-update evpn
            group "dc"
                type internal
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
            group "wan"
                type internal
                neighbor 192.0.2.5
                exit
                neighbor 192.0.2.6
                exit
            exit
            no shutdown
        exit
```

There is another aspect for which policies are used: on PE-2 and PE-3, two MAC/IP routes with the Anycast GW virtual MAC are received (one from PE-4 and another from PE5). To provide efficient upstream routing with no tromboning, it is important that PE-2 prefers the PE-4 virtual MAC route (its own DGW) over that of PE-5, and vice versa for PE-3. This is achieved by:

- Configuring the same RD on PE-4 and PE-5 for VPLS10.
- Configuring an import policy on PE-2 and PE-3 that modifies the local preference of the routes, so that each one prefers the local DGW.

PE-2 and PE-3 could have dropped the routes from the non-local DCGW, but with this configuration, DCGW redundancy is provided in case of failure:

```
*A:PE-2# show router policy "import-prefer-DC-1"
    entry 10
        from
            community "SOO-DC-1"
        exit
        action accept
            local-preference 200
        exit
    exit
```

```
*A:PE-2# show router bgp routes evpn mac community target:64500:10
                                                  mac-address 00:00:5e:00:01:01
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  64500:10            00:00:5e:00:01:01 ESI-0
      0                   Static            VNI 10
                          10.0.0.254
                          192.0.2.4

*i    64500:10            00:00:5e:00:01:01 ESI-0
      0                   Static            VNI 10
                          10.0.0.254
                          192.0.2.5

-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

```
*A:PE-3# show router policy "import-prefer-DC-2"
    entry 10
        from
            community "SOO-DC-2"
        exit
        action accept
            local-preference 200
        exit
    exit
```

```
*A:PE-3# show router bgp routes evpn mac community target:64500:10
                                                   mac-address 00:00:5e:00:01:01
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  64500:10            00:00:5e:00:01:01 ESI-0
      0                   Static            VNI 10
                          10.0.0.254
                          192.0.2.5

*i    64500:10            00:00:5e:00:01:01 ESI-0
      0                   Static            VNI 10
```

```
                          10.0.0.254
                          192.0.2.4

-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

Finally, the VPRN 11 configuration on PE-4 and PE-5 is as follows:

```
# on PE-4:
configure
    service
        vprn 11 name "wan-ip-vpn" customer 1 create
            interface "evi-10" create
                address 10.0.0.4/16
                mac 00:00:00:00:00:04
                arp-host-route
                    populate static route-tag 1
                    populate dynamic route-tag 1
                    populate evpn route-tag 1
                exit
                arp-timeout 600
                arp-learn-unsolicited
                vrrp 1 passive
                    backup 10.0.0.254
                    ping-reply
                    traceroute-reply
                exit
                vpls "centralized-gw-bd"
                exit
            exit
            bgp-ipvpn
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    route-distinguisher auto-rd
                    vrf-target target:64500:11
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

```
# on PE-5:
configure
    service
        vprn 11 name "wan-ip-vpn" customer 1 create
            interface "evi-10" create
                address 10.0.0.5/16
                mac 00:00:00:00:00:05
                arp-host-route
                    populate static route-tag 1
                    populate dynamic route-tag 1
                    populate evpn route-tag 1
                exit
                arp-timeout 600
                arp-learn-unsolicited
                vrrp 1 passive
                    backup 10.0.0.254
                    ping-reply
                    traceroute-reply
                exit
                vpls "centralized-gw-bd"
                exit
            exit
            bgp-ipvpn
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    route-distinguisher auto-rd
                    vrf-target target:64500:11
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

The passive VRRP commands, as well as the ARP commands, have already been discussed in preceding sections. The only new command in the configuration is **route-tag 1**. This command tags all the ARP-ND host routes learned on the interface, so that export policies can match on that tag and modify the routes before they are advertised. The command is included for completeness, however, in this configuration, there is no export policy using this tag.

When the configuration is in place and the hosts are connected, the FDBs, proxy-ARP, ARP caches, and route tables are checked with the following commands (example for host-2 and host-6).

When host-2 ARPs for its default gateway (10.0.0.254), PE-2 will reply with the information from its proxy-ARP table:

```
*A:PE-2# show service id 10 proxy-arp detail 10.0.0.254
-------------------------------------------------------------------------------
Proxy Arp
-------------------------------------------------------------------------------
Admin State       : enabled
Dyn Populate      : enabled
Age Time          : disabled            Send Refresh      : 120 secs
Table Size        : 250                 Total             : 5
Static Count      : 0                   EVPN Count        : 4
Dynamic Count     : 1                   Duplicate Count   : 0

Dup Detect
-------------------------------------------------------------------------------
Detect Window     : 3 mins              Num Moves         : 5
Hold down         : 9 mins
Anti Spoof MAC    : None

EVPN
-------------------------------------------------------------------------------
Garp Flood        : disabled            Req Flood         : disabled
Static Black Hole : disabled
EVPN Route Tag    : 1
-------------------------------------------------------------------------------

===============================================================================
VPLS Proxy Arp Entries
===============================================================================
IP Address          Mac Address         Type      Status    Last Update
-------------------------------------------------------------------------------
10.0.0.254          00:00:5e:00:01:01   evpn      active    02/18/2022 15:06:57
-------------------------------------------------------------------------------
Number of entries : 1
===============================================================================
```

When host-2 sends traffic to the virtual MAC, it will forward it to PE-4 based on a lookup on the FDB:

```
*A:PE-2# show service id 10 fdb mac 00:00:5e:00:01:01

===============================================================================
Forwarding Database, Service 10
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
10         00:00:5e:00:01:01 vxlan-1:                EvpnS:P  02/18/22 15:06:57
                             192.0.2.4:10
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

If PE-4 receives packets with MAC Destination Address (DA) equal to the virtual MAC and IP DA of host-6 (172.16.0.6), the forwarding is based on the information in the R-VPLS FDB first, and afterward on the VPRN 11 route table, as follows.

```
*A:PE-4# show service id 10 fdb mac 00:00:5e:00:01:01

===============================================================================
Forwarding Database, Service 10
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
10         00:00:5e:00:01:01 cpm                     Intf     02/18/22 15:06:57
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

```
*A:PE-4# show router 11 route-table

===============================================================================
Route Table (Service: 11)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.0.0/16                                   Local   Local     00h06m07s  0
       evi-10                                                       0
10.0.0.2/32                                   Remote  ARP-ND    00h06m06s  1
       10.0.0.2                                                     0
172.16.0.0/24                                 Remote  BGP VPN   00h05m33s  170
       192.0.2.6 (tunneled)                                         10
-------------------------------------------------------------------------------
No. of Routes: 3
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

When the traffic goes back from host-6 to host-2, PE-6 will forward to PE-4 due to a Longest Prefix Match (LPM) lookup on the VPRN route table. The advertisement of the ARP-ND routes on PE-4 and PE-6 ensures that PE-6 can forward downstream traffic to the correct PE:

```
*A:PE-6# show router 11 route-table
```

```
===============================================================================
Route Table (Service: 11)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.0.0/16                                   Remote  BGP VPN   00h06m57s  170
       192.0.2.4 (tunneled)                                         10
10.0.0.0/16                                   Remote  BGP VPN   00h06m57s  170
       192.0.2.5 (tunneled)                                         10
10.0.0.2/32                                   Remote  BGP VPN   00h06m57s  170
       192.0.2.4 (tunneled)                                         10
10.0.0.3/32                                   Remote  BGP VPN   00h06m57s  170
       192.0.2.5 (tunneled)                                         10
172.16.0.0/24                                 Local   Local     00h07m01s  0
       local                                                        0
-------------------------------------------------------------------------------
No. of Routes: 5
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

Traceroute commands from host-6 provide information about the path to each remote host (VPRN 12 on PE-6 simulates host-6):

```
*A:PE-6# traceroute router 12 10.0.0.2
traceroute to 10.0.0.2, 30 hops max, 40 byte packets
  1  172.16.0.254 (172.16.0.254)    3.09 ms  2.23 ms  2.31 ms
  2  10.0.0.4 (10.0.0.4)    3.27 ms  3.24 ms  3.28 ms
  3  10.0.0.2 (10.0.0.2)    5.64 ms  5.77 ms  5.86 ms
```

```
*A:PE-6# traceroute router 12 10.0.0.3
traceroute to 10.0.0.3, 30 hops max, 40 byte packets
  1  172.16.0.254 (172.16.0.254)    1.96 ms  2.19 ms  2.20 ms
  2  10.0.0.5 (10.0.0.5)    3.44 ms  3.27 ms  3.04 ms
  3  10.0.0.3 (10.0.0.3)    8.40 ms  5.63 ms  5.48 ms
```

Communication between host-2 and host-3 uses regular L2 switching, as expected, because there are EVPN-VXLAN destinations created between PE-2 and PE-3 for VPLS 10:

```
*A:PE-2# show service id 10 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
Instance    VTEP Address                            Egress VNI  EvpnStatic Num
 Mcast       Oper State                              L2 PBR     SupBcasDom MACs
-------------------------------------------------------------------------------
1           192.0.2.3                               10          evpn       2
 BUM         Up                                      No          No
1           192.0.2.4                               10          evpn       2
 BUM         Up                                      No          No
1           192.0.2.5                               10          evpn       1
 BUM         Up                                      No          No
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 3
-------------------------------------------------------------------------------
===============================================================================
===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                      Num. Macs     Last Change
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

```
*A:PE-2# ping router 12 10.0.0.3
PING 10.0.0.3 56 data bytes
64 bytes from 10.0.0.3: icmp_seq=1 ttl=64 time=9.23ms.
64 bytes from 10.0.0.3: icmp_seq=2 ttl=64 time=3.69ms.
64 bytes from 10.0.0.3: icmp_seq=3 ttl=64 time=3.46ms.
64 bytes from 10.0.0.3: icmp_seq=4 ttl=64 time=3.42ms.
64 bytes from 10.0.0.3: icmp_seq=5 ttl=64 time=3.48ms.

---- 10.0.0.3 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 3.42ms, avg = 4.65ms, max = 9.23ms, stddev = 2.29ms
```

```
*A:PE-2# traceroute router 12 10.0.0.3
traceroute to 10.0.0.3, 30 hops max, 40 byte packets
  1  10.0.0.3 (10.0.0.3)    3.76 ms  3.69 ms  3.67 ms
```

Troubleshooting and debugging

The following commands can be used when troubleshooting these scenarios:

- **show router** *<id>* **route table** and **show router** *<id>* **fib** *<slot>* (and their corresponding commands for IPv6)
- **show router** *<id>* **arp / neighbor**
- **show service** *<id>* **fdb detail**
- **show service** *<id>* **proxy-arp/nd detail**
- **show router bgp routes evpn / vpn-ipv4 / vpn-ipv6**

The following debug commands are also important to analyze the scenarios:

```
debug
    router "Base"
        bgp
            update
        exit
    exit
    router service-name "ip-vrf-16"
        ip
            arp
            route-table
        exit
    exit
    router service-name "VM-test-anycast-gw"
        ip
            arp
        exit
    exit
    service
```

```
        id 10
            proxy-arp
                 all
            exit
        exit
    exit
exit
```

Conclusion

ARP-ND host routes are generated out of ARP-ND entries in a router context. These ARP-ND host routes, along with passive VRRP (for Anycast GWs), provide the correct solution for efficient inter-subnet forwarding in DCs and DCI networks.

Auto-Learn MAC Protect in EVPN

This chapter provides information about Auto-Learn MAC Protect in EVPN.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS Release 14.0.R5, but the CLI in the current edition is based on SR OS Release 21.2.R1. Auto-Learn MAC Protect (ALMP) is supported for EVPN in SR OS Release 14.0.R1, and later.

Overview

MAC protection is needed in Layer 2 services to safeguard business-critical MAC addresses against the possibility of being learned on the wrong SAP/SDP-binding. When a MAC address is learned on the wrong SAP/SDP-binding, traffic would be diverted from its intended destination. This could be caused by misconfiguration or by a malicious source launching a Denial of Service (DoS) attack. MAC protect can also be used to prevent loops in certain topologies.

Chapter EVPN for VXLAN Tunnels (Layer 2) describes MAC protection for static MAC addresses that are configured on SAPs or spoke-SDPs. The command to configure static MAC addresses in a VPLS service is as follows:

```
*A:PE-2>conf>serv>vpls>static-mac$ mac ?
  - mac <ieee-address> [create] black-hole
  - mac <ieee-address> [create] sap <sap-id> monitor {fwd-status}
  - no mac <ieee-address>
  - mac <ieee-address> [create] spoke-sdp <sdp-id:vc-id> monitor {fwd-status}
---snip---
```

Configuring static MAC addresses is not scalable if large numbers of MAC addresses need to be protected. Also, configuring static MAC addresses is not an option when the MAC addresses are unknown. Auto-Learn MAC Protect (ALMP) offers the same protection for learned MAC addresses in services such as EVPN VPLS and EVPN R-VPLS. However, ALMP is not supported for PBB-EVPN.

ALMP can be enabled with the **auto-learn-mac-protect** command in EVPN with VXLAN or MPLS bindings on the following:

- SAPs
- Mesh-SDPs

- Spoke-SDPs
- Pseudowire (PW) templates
- Split Horizon Groups (SHGs)
- SHGs in PW templates

When enabled, all MAC addresses learned on those objects become protected.

The following commands can be used to enable ALMP on objects in VPLS 1:

```
configure
    service
        pw-template 1 name "PW1" create
            auto-learn-mac-protect
        exit
        vpls 1 name "VPLS 1" customer 1 create
            split-horizon-group "SHG1" create
                auto-learn-mac-protect
            exit
            sap 1/2/1:1 create
                auto-learn-mac-protect
                no shutdown
            exit
            spoke-sdp 23:1 create
                auto-learn-mac-protect
                no shutdown
            exit
            mesh-sdp 24:1 create
                auto-learn-mac-protect
                no shutdown
            exit
```

When enabled on an SHG, it is only applicable to the SAPs within the SHG, not to spoke-SDPs. If ALMP is required on spoke-SDPs in the SHG, the parameter must be configured on each spoke-SDP individually. All MAC Source Addresses (SAs) learned on these objects will be protected and advertised with the sticky bit set. The sticky bit indicates that these MAC addresses should be treated as protected on the remote PEs, where these protected MAC addresses are considered to have been learned on the EVPN MPLS/ VXLAN destinations. The remote EVPN peers then use the MAC protection functionality in the same way as the local peer to protect the MAC address.

ALMP enables an implicit **restrict-protected-src discard-frame** (RPS-DF) by default on SAPs and spoke/ mesh-SDPs. When enabled, frames with a protected MAC SA are discarded if received on objects where they were not learned and protected. This configuration is the default and cannot be configured on objects where MAC addresses are learned, such as SAPs, spoke/mesh-SDPs, and SHGs.

However, RPS-DF can optionally be configured on destinations in EVPN MPLS or EVPN VXLAN, where data plane MAC learning is never performed for incoming traffic. For EVPN MPLS, the RPS-DF configuration is in the BGP EVPN context, as follows:

```
*A:PE-2>config>service>vpls>bgp-evpn>mpls# restrict-protected-src ?
  - no restrict-protected-src
  - restrict-protected-src discard-frame

 <discard-frame>      : keyword - discard frame and  trap on a protected MAC
```

For EVPN VXLAN, the RPS-DF configuration is in the VXLAN context, as follows:

```
*A:PE-2>config>service>vpls>vxlan$ restrict-protected-src ?
  - no restrict-protected-src
  - restrict-protected-src discard-frame
```

Instead of discarding the frame, the SAP or spoke/mesh-SDP can be brought operationally down when a frame is received with a protected MAC SA that has not been learned on the object, by configuring **restrict-protected-src** (RPS) without any parameter on the object in EVPN services. After the object has been brought down, an operator needs to disable (**shutdown**) and enable (**no shutdown**) the object in order to make it operational again.

RPS can be enabled without any parameter on SAPs, spoke/mesh-SDPs, SHGs, and PW templates, but not on EVPN MPLS/VXLAN destinations, using following commands:

```
configure
    service
        pw-template 1 name "PW1" create
            restrict-protected-src
        exit
        pw-template 2 name "PW2" create
            split-horizon-group "SHG1"
                restrict-protected-src
            exit
        exit
        vpls "VPLS 1"
            split-horizon-group "SHG1"
                restrict-protected-src
            exit
            sap 1/2/1:1
                restrict-protected-src
            exit
            spoke-sdp 23:1
                restrict-protected-src
            exit
            mesh-sdp 24:1
                restrict-protected-src
            exit
```

The operation for an object is reverted to **restrict-protected-src discard-frame** after configuring the **no restrict-protected-src** command.

RPS cannot be configured without any parameter on EVPN MPLS destinations; if attempted, the following error will be raised:

```
*A:PE-2>config>service>vpls>bgp-evpn>mpls# restrict-protected-src
                                                                  ^
Error: Missing parameter
```

Likewise, RPS cannot be configured without any parameter on EVPN VXLAN destinations; if attempted, the following error will be raised:

```
*A:PE-2>config>service>vpls>vxlan# restrict-protected-src
                                                          ^
Error: Missing parameter
```

Note:
The configuration of restrict-protected-src alarm-only and restrict-unprotected-dst are not allowed in EVPN.

Protection is provided at the point where a MAC address first enters the EVPN part of the network. Therefore, the preference for an auto-learned protected MAC address is higher than that of a MAC address received in a BGP update with the sticky bit set.

The following list shows the MAC learning priority, with the highest priority first:

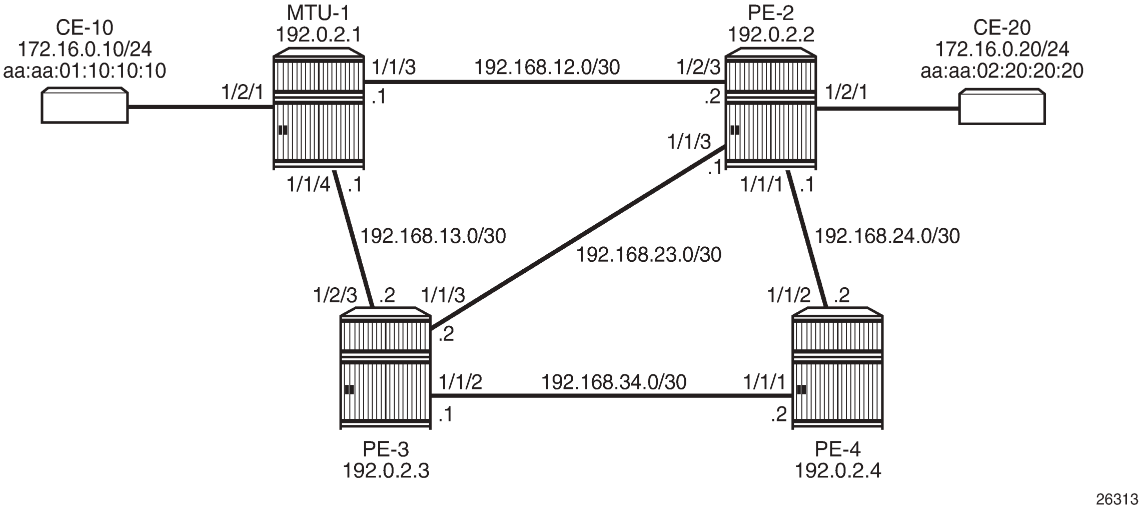
1. Local MAC address (including AS-MAC without static-black-hole, es-bmac, src-bmac, OAM, and so on)
2. Conditional static MAC address (including AS-MAC with static-black-hole)
3. **Auto-Learn Protected MAC address**
4. EVPN MAC address with sticky/static bit set
5. Data plane learned MAC address (regular learning on SAP/SDP-binding)
6. EVPN MAC address without sticky/static bit set

Note:
ALMP MAC addresses have a higher priority but do not overwrite EVPN static MAC addresses.

Configuration

Figure 12: Example topology - no LAG shows the example topology with one MTU and three PEs.

Figure 12: Example topology - no LAG



- Cards, MDAs
- The ports between the PEs are configured as network ports; the other ports are access ports. No LAG is configured initially.
- IGP (IS-IS is used in this example) between the PEs
- LDP between the PEs
- BGP with address family EVPN on the PEs

PE-2 is the BGP route reflector. The BGP configuration on the PEs is similar. BGP is configured on PE-3 as follows:

```
# on PE-3:
configure
    router
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family evpn
                peer-as 64500
                neighbor 192.0.2.2
                exit
            exit
        exit
```

VPLS 1 is configured on all nodes. Initially, ALMP is disabled. On MTU-1, the VPLS 1 contains three SAPs: one toward CE-10, one toward PE-2, and one toward PE-3.

On PE-2, VPLS 1 is configured with EVPN MPLS and contains a SAP toward CE-20 and a SAP toward MTU-1, as follows:

```
# on PE-2:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    ingress-replication-bum-label
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/2/1:1 create
                no shutdown
            exit
            sap 1/2/3:1 create
                no shutdown
            exit
            no shutdown
        exit
```

On PE-3, VPLS 1 is configured with EVPN MPLS and contains a SAP toward MTU-1, as follows:

```
# on PE-3:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
```

```
            bgp
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    ingress-replication-bum-label
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/2/3:1 create
                no shutdown
            exit
            no shutdown
        exit
```

The following use cases will be described in this section:

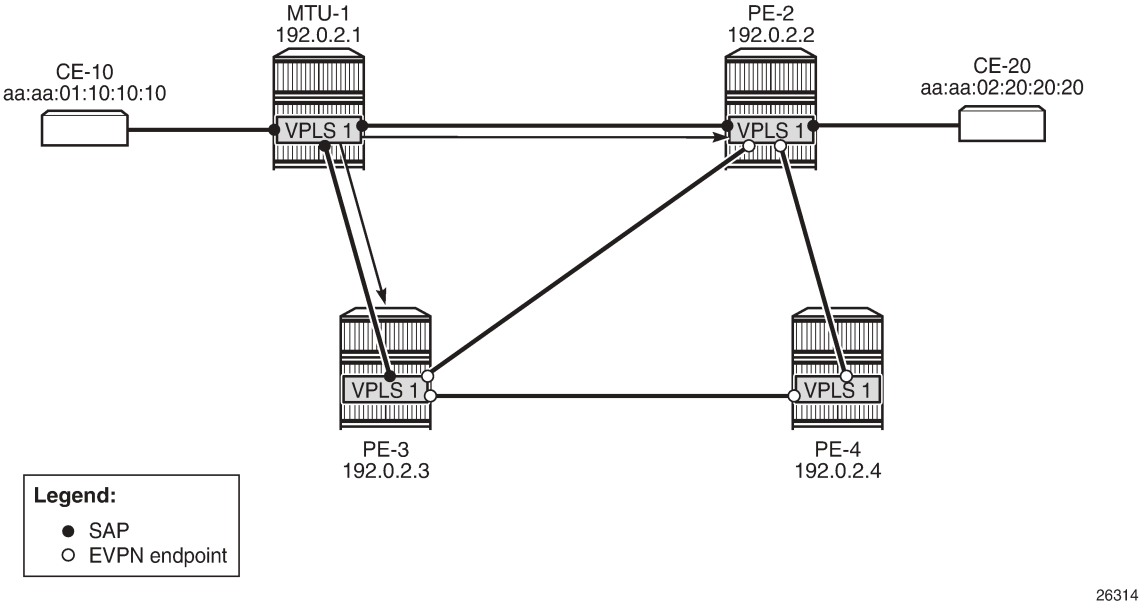
- EVPN MPLS without multi-homing.
  - Default behavior: no ALMP on SAPs, no protected MAC addresses
  - No ALMP on SAPs, RPS-DF on EVPN MPLS destinations
  - ALMP and implicit RPS-DF on SAPs.
    - RPS-DF on EVPN MPLS destinations, MAC first learned on PE-2
    - RPS-DF on EVPN MPLS destinations, MAC simultaneously learned on PE-2 and PE-3
    - No RPS-DF on EVPN MPLS destinations, MAC simultaneously learned on PE-2 and PE-3
  - ALMP and RPS on SAPs.
    - RPS-DF on EVPN MPLS destinations, MAC first learned on PE-2
    - RPS-DF on EVPN MPLS destinations, MAC simultaneously learned on PE-2 and PE-3
    - No RPS-DF on EVPN MPLS destinations, MAC simultaneously learned on PE-2 and PE-3
- EVPN MPLS with ALMP in all-active multi-homing.
  - RPS-DF on SAPs, RPS-DF on EVPN MPLS destinations

Default behavior: no protected MAC addresses

The following example is not a recommended configuration because it causes a loop. By default, ALMP is disabled and no static MAC addresses are configured. As described in chapter EVPN for VXLAN Tunnels (Layer 2), duplicate MAC addresses are detected in BGP EVPN and the MAC address will be put in a hold-down state on the EVPN destinations after a configurable threshold is reached. This applies to EVPN-MPLS as well as to EVPN-VXLAN. By default, the maximum number of MAC address moves is five in a time window of 3 minutes.

Figure 13: MAC address learned simultaneously on SAPs on PE-2 and PE-3 shows that the MAC address from CE-10 is learned simultaneously on the SAPs in VPLS 1 on PE-2 and PE-3.

Figure 13: MAC address learned simultaneously on SAPs on PE-2 and PE-3



CE-10 sends frames to CE-20 with MAC Destination Address (DA) aa:aa:02:20:20:20. MTU-1 has not learned that MAC DA, so the frames are flooded to PE-2 and PE-3, where they enter the SAPs simultaneously. PE-2 and PE-3 have not learned the MAC DA either, so the frames are flooded to all potential destinations. The frames received on PE-2 will be sent (among others) to PE-3, and vice versa. These frames are forwarded back out of the SAP toward MTU-1. This causes a loop.

Both PEs send a BGP update for the MAC SA aa:aa:01:10:10:10 to the other PEs with no sticky bit set. That MAC SA is learned, but not protected on the destination to the other PE. The stream of frames will cause the learned MAC SA to oscillate between the SAP and EVPN destinations on PE-2 and PE-3, and between the EVPN destinations on PE-4.

After a configurable number of BGP EVPN MAC address moves in a time span (by default, after five MAC address moves in a period of 3 minutes), the MAC address is put in a hold-down state on the EVPN destinations for a specific duration (until the next MAC address duplication detection retry; by default, after 9 minutes).

The following message in log 99 on PE-2 (and also on PE-3) indicates that duplicate MAC addresses have been detected:

```
74 2021/03/19 08:14:13.100 UTC MINOR: SVCMGR #2331 Base
"VPLS Service 1 has MAC(s) detected as duplicates by EVPN mac-duplication detection."
```

The following shows the settings for EVPN MAC address duplication detection, which are the default. It also lists the detected duplicate MAC addresses of CE-10 and CE-20:

```
*A:PE-3# show service id 1 bgp-evpn

===============================================================================
BGP EVPN Table
===============================================================================
MAC Advertisement  : Enabled            Unknown MAC Route  : Disabled
CFM MAC Advertise  : Disabled
```

```
Creation Origin    : manual
MAC Dup Detn Moves : 5                  MAC Dup Detn Window: 3
MAC Dup Detn Retry : 9                  Number of Dup MACs : 2
MAC Dup Detn BH    : Disabled
IP Route Advert    : Disabled
Sel Mcast Advert   : Disabled

EVI                : 1
Ing Rep Inc McastAd: Enabled
Accept IVPLS Flush : Disabled

-------------------------------------------------------------------------------
Detected Duplicate MAC Addresses             Time Detected
-------------------------------------------------------------------------------
aa:aa:01:10:10:10                            03/19/2021 08:14:13
aa:aa:02:20:20:20                            03/19/2021 08:14:13
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN MPLS Information
===============================================================================
Admin Status       : Enabled            Bgp Instance       : 1
Force Vlan Fwding  : Disabled
Route NextHop Type : system-ipv4
Control Word       : Disabled
Max Ecmp Routes    : 1
Entropy Label      : Disabled
Default Route Tag  : none
Split Horizon Group: (Not Specified)
Ingress Rep BUM Lbl: Enabled
Ingress Ucast Lbl  : 524284             Ingress Mcast Lbl  : 524283
RestProtSrcMacAct  : none
Evpn Mpls Encap    : Enabled            Evpn MplsoUdp      : Disabled
Oper Group         :
===============================================================================

===============================================================================
BGP EVPN MPLS Auto Bind Tunnel Information
===============================================================================
Allow-Flex-Algo-Fallback  : false
Resolution                : any                Strict Tnl Tag   : false
Max Ecmp Routes           : 1
Bgp Instance              : 1
Filter Tunnel Types       : (Not Specified)
===============================================================================
```

RPS is disabled (by default) on the EVPN destinations (**RestProtSrcMacAct** : *none*).

The MAC addresses are in a hold-down state on the EVPN destinations and no MAC address moves take place until the next MAC address duplication detection retry after 9 minutes. After 9 minutes, the EVPN MAC address duplication alarm is cleared, but after the next five MAC address moves within a time span of 3 minutes, the alarm is raised again and this threshold is reached soon after the alarm has been cleared.

The MAC addresses of both CEs are learned on the SAP of PE-3 (CE-20's MAC address is also learned on the SAP toward MTU-1), not on the EVPN destinations, because of the MAC address duplication detection and hold-down state in EVPN, as follows:

```
*A:PE-3# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          aa:aa:01:10:10:10 sap:1/2/3:1             L/0      03/19/21 08:14:13
1          aa:aa:02:20:20:20 sap:1/2/3:1             L/0      03/19/21 08:14:13
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

A similar output can be shown for PE-2.

Both PE-2 and PE-3 learn the MAC addresses locally and send BGP EVPN MAC address route updates to their BGP peers. PE-3 received the following BGP EVPN MAC address routes from PE-2, with the MAC address mobility sequence number representing the number of MAC address moves:

```
*A:PE-3# show router bgp routes evpn mac
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.2:1         aa:aa:01:10:10:10 ESI-0
      0                   Seq:4             LABEL 524284
                          n/a
                          192.0.2.2

u*>i  192.0.2.2:1         aa:aa:02:20:20:20 ESI-0
      0                   Seq:4             LABEL 524284
                          n/a
                          192.0.2.2

-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

PE-3 does not use these BGP EVPN MAC address routes in its FDB, because locally learned MAC addresses are preferred.

The remote PE (PE-4) received the following BGP EVPN MAC routes from PE-2 and PE-3:

```
*A:PE-4# show router bgp routes evpn mac
===============================================================================
 BGP Router ID:192.0.2.4        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
```

```
===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.2:1         aa:aa:01:10:10:10 ESI-0
      0                   Seq:4             LABEL 524284
                          n/a
                          192.0.2.2

u*>i  192.0.2.2:1         aa:aa:02:20:20:20 ESI-0
      0                   Seq:4             LABEL 524284
                          n/a
                          192.0.2.2

u*>i  192.0.2.3:1         aa:aa:01:10:10:10 ESI-0
      0                   Seq:3             LABEL 524284
                          n/a
                          192.0.2.3

u*>i  192.0.2.3:1         aa:aa:02:20:20:20 ESI-0
      0                   Seq:5             LABEL 524284
                          n/a
                          192.0.2.3

-------------------------------------------------------------------------------
Routes : 4
===============================================================================
```

In the preceding output, MAC aa:aa:01:10:10:10 is learned from BGP peer 192.0.2.3 with MAC mobility sequence number 3, and from BGP peer 192.0.2.2 with sequence number 4. MAC aa:aa:02:20:20:20 is learned from BGP peer 192.0.2.2 with sequence number 4 and from BGP peer 192.0.2.3 with sequence number 5. The FDB for VPLS 1 on PE-4 contains the MAC addresses learned from BGP EVPN MAC updates with the highest MAC mobility sequence number, as follows:

```
*A:PE-4# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          aa:aa:01:10:10:10 mpls:                   Evpn     03/19/21 08:14:13
                             192.0.2.2:524284
           ldp:65538
1          aa:aa:02:20:20:20 mpls:                   Evpn     03/19/21 08:14:13
                             192.0.2.3:524284
           ldp:65537
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

VPLS 1 on MTU-1 does not have EVPN configured and no MAC address duplication detection mechanism implemented. The MAC address from CE-10 is last learned on the SAP toward PE-3 (it might equally have been the SAP toward PE-2) instead of the SAP toward CE-10, resulting from the loop, as follows:

```
*A:MTU-1# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          aa:aa:01:10:10:10 sap:1/1/4:1             L/0      03/19/21 08:18:05
1          aa:aa:02:20:20:20 sap:1/1/3:1             L/0      03/19/21 08:14:13
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

No ALMP on SAPs, RPS-DF on EVPN destinations

When there are no protected MAC addresses (ALMP is disabled and no static MAC addresses are configured), the behavior is as described earlier. RPS-DF discards frames with protected MAC addresses that were not learned on the object, but there are no protected MAC addresses, because ALMP is not configured. RPS-DF does not discard frames with MAC SAs that are not protected.

RPS-DF is enabled on EVPN destinations on all PEs, as follows:

```
# on PE-2, PE-3, PE-4:
configure
    service
        vpls "VPLS 1"
            bgp-evpn
                mpls bgp 1
                    restrict-protected-src discard-frame
                exit
            exit
```

The state of RPS is now "discard-frame" instead of "none", as follows:

```
*A:PE-3# show service id 1 bgp-evpn | match RestProtSrcMacAct
RestProtSrcMacAct  : Discard-frame
```

It is also allowed to configure RPS without parameters on the SAPs, but that does not change the behavior when ALMP is disabled and there are no protected MAC addresses. RPS will not bring down a SAP after receiving a frame with an unprotected MAC SA.

ALMP and implicit RPS-DF on SAPs

ALMP is enabled on the SAPs in PE-2 as follows:

```
# on PE-2:
configure
    service
```

```
        vpls "VPLS 1"
            sap 1/2/1:1          # SAP toward CE-20
                auto-learn-mac-protect
                no shutdown
            exit
            sap 1/2/3:1          # SAP toward MTU-1
                auto-learn-mac-protect
                no shutdown
            exit
```

The configuration is similar on PE-3.

The following shows that ALMP is enabled on the SAP and that the default RPS-DF is used:

```
*A:PE-2# show service id 1 sap 1/2/3:1 detail

===============================================================================
Service Access Points(SAP)
===============================================================================
Service Id         : 1
SAP                : 1/2/3:1                  Encap             : q-tag
Description        : (Not Specified)
Admin State        : Up                       Oper State        : Up
Flags              : None
---snip---

Restr MacUnpr Dst  : Disabled
Auto Learn Mac Prot: Enabled
ALMP Exclude List  : <none>
RestMacProtSrc Act : none (oper: Discard-frame)
---snip---
```

ALMP and RPS-DF on SAPs, RPS-DF on EVPN MPLS destinations, MAC first learned on PE-2

Initially, the SAP on PE-3 is disabled to ensure that the MAC address will first be learned on PE-2, then on PE-3, as follows:

```
# on PE-3:
configure
    service
        vpls "VPLS 1"
            sap 1/2/3:1
                shutdown
```

Each learned MAC address on the SAPs on PE-2 will be protected; therefore, a BGP update with the static/sticky bit set will be sent to the BGP EVPN peers. In this example, the MAC aa:aa:01:10:10:10 of CE-10 is learned first on SAP 1/2/3:1 on PE-2, and MAC aa:aa:02:20:20:20 is learned on SAP 1/2/1:1 on PE-2. Consequently, PE-2 sends BGP updates with the static/sticky bit set to PE-3 for both MAC aa:aa:01:10:10:10 and MAC aa:aa:02:20:20:20, as follows:

```
67 2021/03/19 08:23:22.782 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 89
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-MAC Len: 33 RD: 192.0.2.2:1 ESI: ESI-0, tag: 0, mac len: 48
                       mac: aa:aa:01:10:10:10, IP len: 0, IP: NULL, label1: 8388544
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:1
        bgp-tunnel-encap:MPLS
        mac-mobility:Seq:0/Static
"
```

```
69 2021/03/19 08:23:22.783 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 89
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-MAC Len: 33 RD: 192.0.2.2:1 ESI: ESI-0, tag: 0, mac len: 48
                       mac: aa:aa:02:20:20:20, IP len: 0, IP: NULL, label1: 8388544
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:1
        bgp-tunnel-encap:MPLS
        mac-mobility:Seq:0/Static
"
```

Note:

The MPLS label is label1 in the BGP update divided by 16 (2^4), as follows:

$$\frac{8388544}{16} = 524284$$

PE-2 sends similar BGP EVPN updates to peer PE-4.

After these BGP EVPN updates have been sent to PE-3 (and PE-4), the SAP on PE-3 is enabled again, as follows:

```
# on PE-3:
configure
    service
        vpls "VPLS 1"
            sap 1/2/3:1
                no shutdown
```

The MAC addresses in the FDB on PE-2, where these MAC addresses are learned, get the indication "L" for learned and "P" for protected MAC address, as follows:

```
*A:PE-2# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          aa:aa:01:10:10:10 sap:1/2/3:1             LP/180   03/19/21 08:23:23
1          aa:aa:02:20:20:20 sap:1/2/1:1             LP/180   03/19/21 08:23:23
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The MAC addresses in the FDB on PE-3 are learned from the BGP EVPN updates and get the indication "S" for static (sticky bit) and "P" for protected MAC address, as follows

```
*A:PE-3# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          aa:aa:01:10:10:10 mpls:                   EvpnS:P  03/19/21 08:23:23
                             192.0.2.2:524284
           ldp:65537
1          aa:aa:02:20:20:20 mpls:                   EvpnS:P  03/19/21 08:23:23
                             192.0.2.2:524284
           ldp:65537
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The FDB on the remote PE (PE-4) looks similar, as follows:

```
*A:PE-4# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          aa:aa:01:10:10:10 mpls:                   EvpnS:P  03/19/21 08:23:23
                             192.0.2.2:524284
           ldp:65538
1          aa:aa:02:20:20:20 mpls:                   EvpnS:P  03/19/21 08:23:23
                             192.0.2.2:524284
           ldp:65538
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

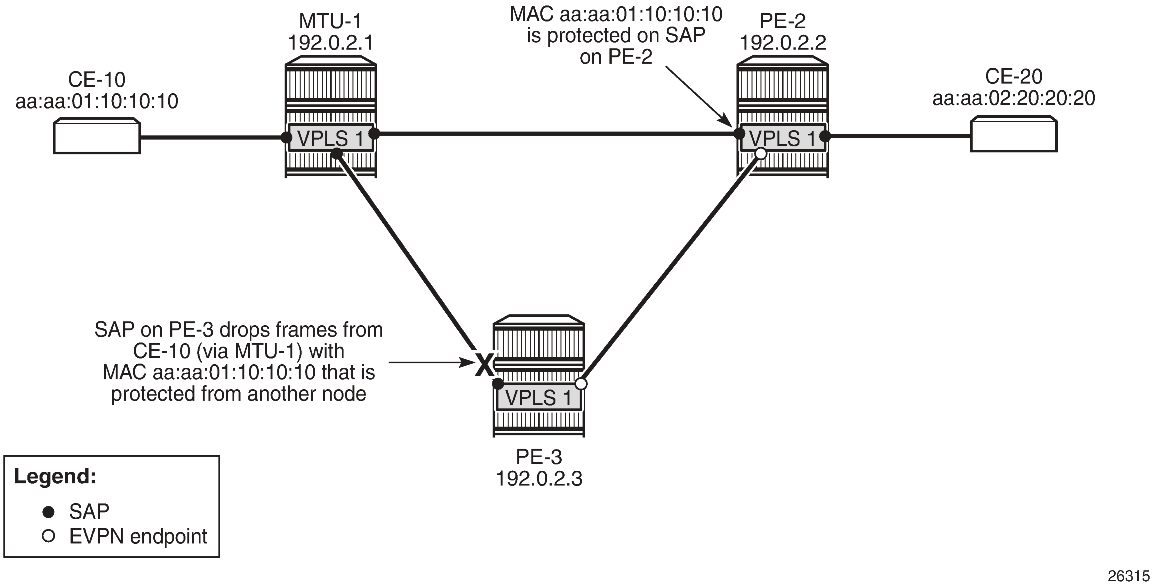
The BGP EVPN MAC address routes on PE-3 have MAC address mobility equal to "Static", as follows:

```
*A:PE-3# show router bgp routes evpn mac
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.2:1         aa:aa:01:10:10:10 ESI-0
      0                   Static            LABEL 524284
                          n/a
                          192.0.2.2

u*>i  192.0.2.2:1         aa:aa:02:20:20:20 ESI-0
      0                   Static            LABEL 524284
                          n/a
                          192.0.2.2

-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

The BGP EVPN MAC routes on PE-4 are similar.

When a stream of frames with MAC SA aa:aa:01:10:10:10 enters the SAP on PE-3, these frames will be dropped by this SAP because of the implicit RPS-DF behavior in the SAP for protected MAC addresses, as shown in Figure 14: Default RPS-DF on SAPs - MAC learned and protected on SAP on PE-2.

Figure 14: Default RPS-DF on SAPs - MAC learned and protected on SAP on PE-2



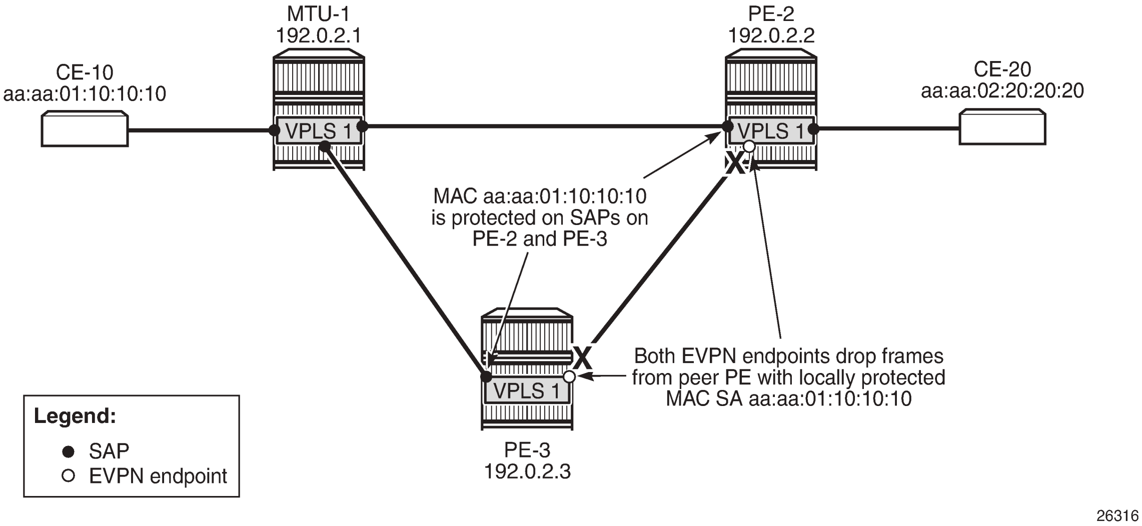
Because the MAC address was protected on the SAP on PE-2 and the BGP EVPN MAC route update had been received by PE-3 before any frame was received with this MAC SA, there will be no temporary loop. The frames with the protected MAC SA will be discarded at the SAP on PE-3, not on the EVPN MPLS destination on PE-2. In this case, there is no need to configure RPS-DF on the EVPN MPLS destinations, but it will make a difference when the MAC address is learned on both SAPs simultaneously.

ALMP and RPS-DF on SAPs, RPS-DF on EVPN MPLS destinations, MAC simultaneously learned on PE-2 and PE-3

In the preceding example, the MAC addresses of CE-10 and CE-20 were first learned and protected on PE-2 and received on PE-3's SAP after the BGP update with static/sticky bit was received by PE-3. However, when the MAC address of CE-10 is learned simultaneously on both PEs, for example, because the MAC DA aa:aa:02:20:20:20 is unknown, there is a temporary loop until the MAC addresses are protected. Initially, the frames enter a SAP, are forwarded to the EVPN peer, and forwarded out of the remote SAP.

After the MAC addresses are learned and protected on the SAPs on both PEs, new frames received on a SAP with the protected MAC address will be sent to the other PE. However, they will be discarded due to RPS-DF on destination, as shown in Figure 15: MAC learned and protected simultaneously on PEs - RPS-DF on EVPN endpoints, because the destination PE has that same MAC address protected on its local SAP. This prevents a loop. BGP updates with the static/sticky bit set are sent to the BGP EVPN peer, but the locally learned and protected MAC address is preferred to the MAC address in a BGP update. Therefore, the FDB contains the locally learned MAC address aa:aa:01:10:10:10, not the BGP EVPN MAC address update for MAC address aa:aa:01:10:10:10.

Figure 15: MAC learned and protected simultaneously on PEs - RPS-DF on EVPN endpoints



The MAC addresses of the CEs are cleared from the FDBs on all nodes, as follows:

```
clear service id 1 fdb mac aa:aa:01:10:10:10
clear service id 1 fdb mac aa:aa:02:20:20:20
```

This clear command for the FDB only works for auto-learned MAC addresses, not for BGP EVPN MAC address updates. BGP EVPN MAC address withdraw updates need to be sent. In this example, BGP is configured with **rapid-update evpn**, as shown previously.

When traffic is sent from CE-10 to CE-20, MAC address aa:aa:01:10:10:10 of CE-10 is learned simultaneously on SAP 1/2/3:1 in PE-2 and PE-3 and protected on both SAPs. MAC address aa:aa:02:20:20:20 is, in this case, first learned via MAC address learning on PE-2 and advertised via a BGP EVPN MAC address route update. However, it might happen that it was learned and protected on the SAP on PE-3 first, before the MAC address was learned and protected on PE-2 and the BGP EVPN MAC address route update sent by PE-2 was received at PE-3. In the latter case, both MAC address aa:aa:01:10:10:10 and MAC address aa:aa:02:20:20:20 are learned and protected on the SAPs on both PE-2 and PE-3, and RPS-DF on the EVPN-MPLS destinations prevents loops.

However, in the present case, MAC address aa:aa:02:20:20:20 is only protected on the SAP on PE-2, because PE-3 received the EVPN MAC address update before it received a frame with MAC SA aa:aa:02:20:20:20. Therefore, the SAP on PE-3 will discard any frames with MAC SA aa:aa:02:20:20:20.

The FDB for VPLS 1 on PE-2 shows that both MAC addresses are learned locally and protected, as follows:

```
*A:PE-2# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          aa:aa:01:10:10:10 sap:1/2/3:1             LP/0     03/19/21 08:33:32
1          aa:aa:02:20:20:20 sap:1/2/1:1             LP/0     03/19/21 08:33:32
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The FDB for VPLS 1 on PE-3 shows that MAC address aa:aa:01:10:10:10 is learned and protected locally, but MAC address aa:aa:02:20:20:20 is protected on PE-2, which has been advertised by PE-2 in a BGP EVPN MAC update, as follows:

```
*A:PE-3# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          aa:aa:01:10:10:10 sap:1/2/3:1             LP/16    03/19/21 08:33:32
1          aa:aa:02:20:20:20 mpls:                   EvpnS:P  03/19/21 08:33:32
                             192.0.2.2:524284
           ldp:65537
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

Both PE-2 and PE-3 send BGP EVPN MAC updates to their BGP peers for each locally learned and protected MAC address. The following BGP EVPN MAC update is sent by PE-2 to PE-3 for MAC address aa:aa:01:10:10:10:

```
# on PE-2:
```

```
72 2021/03/19 08:33:32.068 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 89
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-MAC Len: 33 RD: 192.0.2.2:1 ESI: ESI-0, tag: 0, mac len: 48
                       mac: aa:aa:01:10:10:10, IP len: 0, IP: NULL, label1: 8388544
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:1
        bgp-tunnel-encap:MPLS
        mac-mobility:Seq:0/Static
"
```

Similar BGP EVPN updates are sent to the remote PE (PE-4). The FDB for VPLS 1 on PE-4 only contains entries learned from BGP EVPN updates, as follows:

```
*A:PE-4# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          aa:aa:01:10:10:10 mpls:                   EvpnS:P  03/19/21 08:33:32
                             192.0.2.2:524284
           ldp:65538
1          aa:aa:02:20:20:20 mpls:                   EvpnS:P  03/19/21 08:33:32
                             192.0.2.2:524284
           ldp:65538
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

PE-4 received BGP EVPN MAC address route updates from PE-2 and PE-3, but only installs the MAC address routes to PE-2 in its FDB, based on the lowest next-hop IP of the EVPN NLRI (192.0.2.2).

ALMP and RPS-DF on SAPs, no RPS-DF on EVPN MPLS destinations, MAC simultaneously learned on PE-2 and PE-3

RPS-DF is disabled on the EVPN MPLS destinations on the PEs, as follows:

```
# on PE-2, PE-3, PE-4:
configure
    service
        vpls "VPLS 1"
            bgp-evpn
                mpls bgp 1
                    no restrict-protected-src
```

When a frame is received at SAP 1/2/3:1 on PE-3 with protected MAC SA aa:aa:01:10:10:10, it is not dropped by the SAP, because this MAC SA has been learned and protected on this SAP on PE-3. The frame is forwarded to PE-2 where it will not be discarded by the EVPN MPLS destination because RPS-DF is disabled. The frame will be forwarded to other objects in the VPLS in PE-2. For BUM traffic, there will be a loop, because all frames will be flooded to all objects in VPLS 1 on PE-2, including the SAP toward MTU-1.

ALMP and RPS on SAPs

When ALMP is enabled on an object, the default behavior is that frames with a protected MAC SA are discarded (RPS-DF). However, it is possible to configure RPS without any parameter on the object, in this case on the SAPs on PE-2 and PE-3, as follows:

```
# on PE-2, PE-3:
configure
    service
        vpls "VPLS 1"
            sap 1/2/3:1
                restrict-protected-src
```

Instead of discarding frames with MAC SAs that are protected on another object or node, the entire object (here: SAP) can be brought operationally down after a frame has been received with a MAC SA that is protected on another node.

The RPS configuration on the SAP can be shown as follows. The SAP has not been brought down yet.

```
*A:PE-2# show service id 1 sap 1/2/3:1 detail

===============================================================================
Service Access Points(SAP)
===============================================================================
Service Id         : 1
SAP                : 1/2/3:1                  Encap             : q-tag
Description        : (Not Specified)
Admin State        : Up                       Oper State        : Up
Flags              : None
---snip---

Restr MacUnpr Dst  : Disabled
Auto Learn Mac Prot: Enabled
ALMP Exclude List  : <none>
RestMacProtSrc Act : SAP-oper-down
---snip---
```

The **RestMacProtSrc Act** parameter is set to *SAP-oper-down*, meaning that RPS is configured without any parameter, which causes the system to bring down the SAP when a duplicate MAC address is received that is protected on another object or node. When a SAP is brought down because of this, the *RxProtSrcMAC* flag will be raised and can be shown in the detailed SAP show output.

ALMP and RPS on SAPs, RPS-DF on EVPN MPLS destinations, MAC first learned on PE-2

RPS-DF is enabled on the EVPN MPLS destinations on the PEs, as follows:

```
# on PE-2, PE-3, PE-4:
configure
    service
        vpls "VPLS 1"
            bgp-evpn
                mpls bgp 1
                    restrict-protected-src discard-frame
```

To simulate a scenario where the MAC addresses are first learned on PE-2, the SAP on PE-3 is disabled until the BGP EVPN MAC route updates are sent, as follows:

```
# on PE-3:
configure
    service
        vpls "VPLS 1"
            sap 1/2/3:1
                shutdown
```

The FDBs are cleared on the nodes, as follows:

```
clear service id 1 fdb mac aa:aa:01:10:10:10
clear service id 1 fdb mac aa:aa:02:20:20:20
```

Traffic is sent between CE-10 and CE-20, and the MAC addresses are learned and protected on the SAP on PE-2, as follows:

```
*A:PE-2# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          aa:aa:01:10:10:10 sap:1/2/3:1             LP/30    03/19/21 08:40:39
1          aa:aa:02:20:20:20 sap:1/2/1:1             LP/30    03/19/21 08:40:39
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

No MAC learning took place on the SAP on PE-3, and the FDB contains the MAC addresses from the BGP EVPN updates, as follows:

```
*A:PE-3# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
```

```
1          aa:aa:01:10:10:10 mpls:                   EvpnS:P  03/19/21 08:40:39
                             192.0.2.2:524284
           ldp:65537
1          aa:aa:02:20:20:20 mpls:                   EvpnS:P  03/19/21 08:40:39
                             192.0.2.2:524284
           ldp:65537
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The SAP on PE-3 is enabled, as follows:

```
# on PE-3:
configure
    service
        vpls "VPLS 1"
            sap 1/2/3:1
                no shutdown
```

The operational state of the SAP is up, because no protected MAC addresses have been received yet:

```
*A:PE-3# show service id 1 sap

===============================================================================
SAP(Summary), Service 1
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
1/2/3:1                         1          1     none    1     none   Up   Up
-------------------------------------------------------------------------------
Number of SAPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

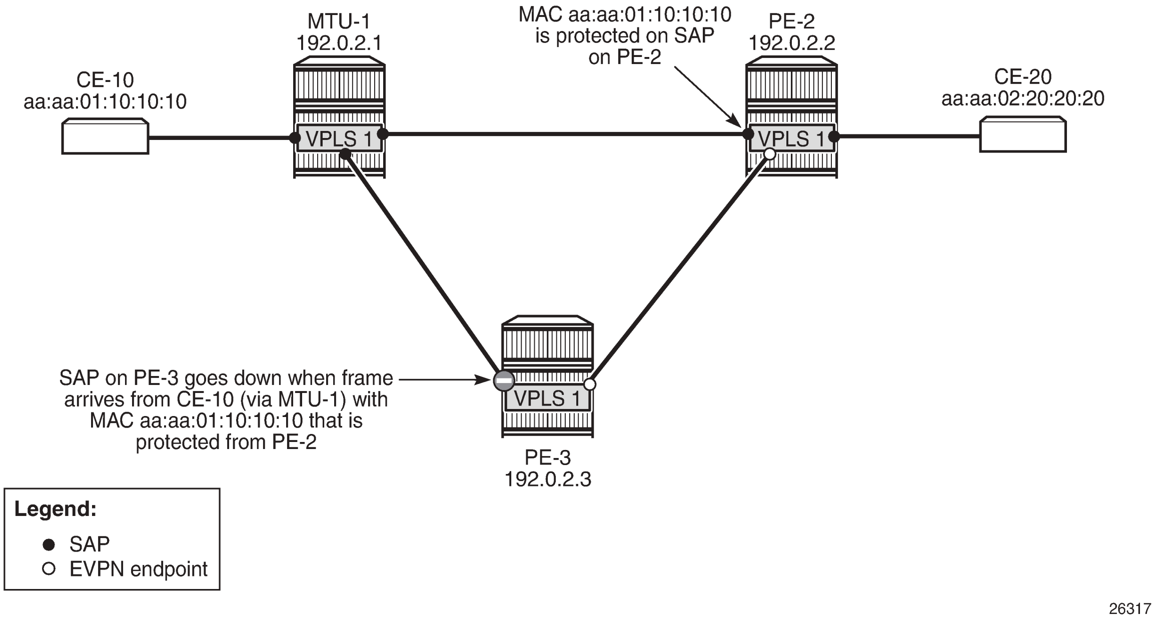
The FDB is cleared for MAC address aa:aa:02:20:20:20 on MTU-1, as follows:

```
# on MTU-1:
clear service id 1 fdb mac aa:aa:02:20:20:20
```

Traffic from CE-10 toward the unknown MAC address aa:aa:02:20:20:20 reaches the SAPs on PE-2 and PE-3. When MAC SA aa:aa:01:10:10:10, which is protected on PE-2, is received on PE-3, SAP 1/2/3:1 will be brought operationally down, as shown in Figure 16: MAC learned and protected on SAP on PE-2 - RPS enabled on SAP on PE-3, and the following alarms will be raised in log 99:

```
85 2021/03/19 08:41:17.636 UTC MINOR: SVCMGR #2208 Base
"Protected MAC aa:aa:01:10:10:10 received on SAP 1/2/3:1 in service 1. The SAP will be
 disabled."

86 2021/03/19 08:41:17.636 UTC MINOR: SVCMGR #2203 Base
"Status of SAP 1/2/3:1 in service 1 (customer 1) changed to admin=up oper=down flags=RxProtSrc
Mac "
```

Figure 16: MAC learned and protected on SAP on PE-2 - RPS enabled on SAP on PE-3



The operational state of SAP 1/2/3:1 is now down. Detailed information about this SAP shows the *RxProtSrcMAC* flag, indicating that a duplicate MAC address that is protected on a remote node has been received, as follows:

```
*A:PE-3# show service id 1 sap 1/2/3:1

===============================================================================
Service Access Points(SAP)
===============================================================================
Service Id         : 1
SAP                : 1/2/3:1                  Encap             : q-tag
Description        : (Not Specified)
Admin State        : Up                       Oper State        : Down
Flags              : RxProtSrcMac
Multi Svc Site     : None
Last Status Change : 03/19/2021 08:41:18
Last Mgmt Change   : 03/19/2021 08:40:57
===============================================================================
```

The SAP is operationally down and will not come up automatically when the FDB is cleared. To bring the SAP up, an operator needs to disable and re-enable the SAP, as follows:

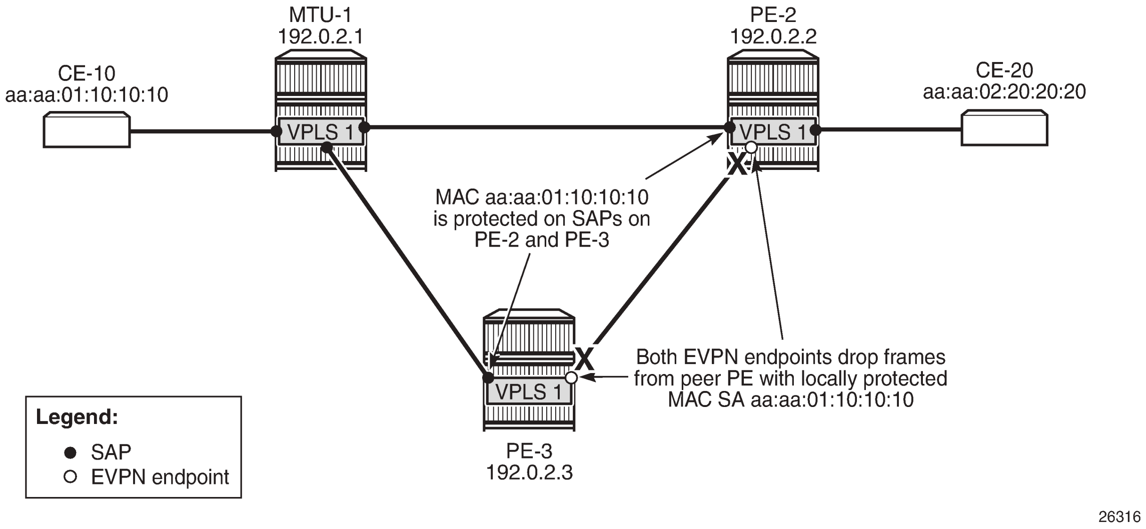
```
*A:PE-3# configure service vpls "VPLS 1" sap 1/2/3:1 shutdown
*A:PE-3# configure service vpls "VPLS 1" sap 1/2/3:1 no shutdown
*A:PE-3# show service id 1 sap

===============================================================================
SAP(Summary), Service 1
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
1/2/3:1                         1          1     none    1     none   Up   Up
```

```
-------------------------------------------------------------------------------
Number of SAPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

ALMP and RPS on SAPs, RPS-DF on EVPN MPLS destinations, MAC simultaneously learned on PE-2 and PE-3

When CE-10 sends traffic to CE-20 and the destination MAC address is unknown, MAC address aa:aa:01:10:10:10 is simultaneously learned and protected on PE-2 and PE-3. No SAP will be brought down when MAC address aa:aa:01:10:10:10 is received on PE-2 or PE-3. This scenario is identical to the one with ALMP and (default) RPS-DF on the SAPs, as shown in Figure 17: RPS enabled on SAPs - RPS-DF on EVPN endpoints, MACs learned simultaneously (which is identical to Figure 15: MAC learned and protected simultaneously on PEs - RPS-DF on EVPN endpoints).

Figure 17: RPS enabled on SAPs - RPS-DF on EVPN endpoints, MACs learned simultaneously



A temporary loop is possible until the MAC address is protected on the SAPs. Initially, the frames enter the SAP, are forwarded to the other PEs, and are forwarded out of the other SAP (unless the MAC address is protected). When the MAC address is protected, any other frames received on the SAP will be sent to the other PE (for example, from PE-3 to PE-2, or vice versa), but they will be discarded by the receiving PE, because RPS-DF is applied on the EVPN destination. BGP EVPN updates are sent to the peer PEs with the sticky bit set. This MAC route will not be installed in the FDB of PE-2 and PE-3 because the MAC address has already been learned locally, which has a higher preference.

The FDB on PE-2 contains locally learned and protected MAC addresses, as follows:

```
*A:PE-2# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
```

```
1          aa:aa:01:10:10:10 sap:1/2/3:1             LP/0     03/19/21 08:47:53
1          aa:aa:02:20:20:20 sap:1/2/1:1             LP/0     03/19/21 08:47:53
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The FDB on PE-3 contains MAC address aa:aa:01:10:10:10 that is learned locally and protected before a BGP-EVPN MAC was received and MAC address aa:aa:02:20:20:20 that is protected on PE-2, as follows.

```
*A:PE-3# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          aa:aa:01:10:10:10 sap:1/2/3:1             LP/0     03/19/21 08:47:53
1          aa:aa:02:20:20:20 mpls:                   EvpnS:P  03/19/21 08:47:53
                             192.0.2.2:524284
           ldp:65537
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

SAP 1/2/3:1 will not be brought down if frames are received with MAC address aa:aa:01:10:10:10 that is locally learned and protected. However, MAC address aa:aa:02:20:20:20 was learned and protected first on PE-2 and the BGP update was received by PE-3 before the MAC address was received on PE-3. Therefore, MAC address aa:aa:02:20:20:20 will not be learned and protected on PE-3 and, if frames with a MAC SA aa:aa:02:20:20:20 were received on SAP 1/2/3:1 on PE-3, the SAP would be brought down.

ALMP and RPS on SAPs, no RPS-DF on EVPN MPLS destinations, MAC simultaneously learned on PE-2 and PE-3

RPS-DF is disabled on the EVPN MPLS destinations on the PEs, as follows:

```
# on PE-2, PE-3, PE-4:
configure
    service
        vpls "VPLS 1"
            bgp-evpn
                mpls bgp 1
                    no restrict-protected-src
```

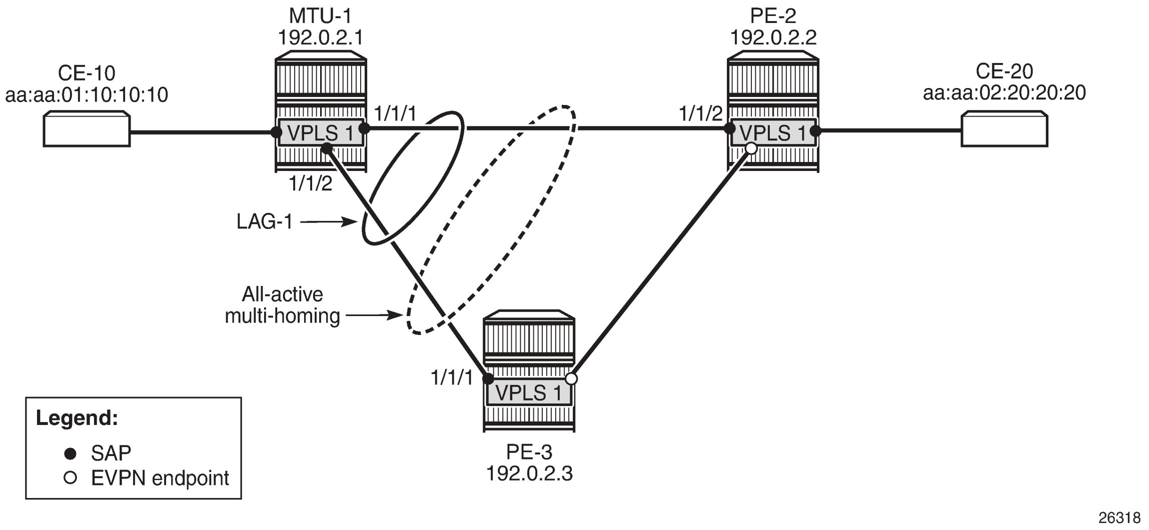
When frames are received at SAP 1/2/3:1 on PE-3 with protected MAC SA aa:aa:01:10:10:10, the SAP is not brought down, because this MAC SA has been learned and protected on this SAP. The frame is forwarded to PE-2 where it will not be discarded by the EVPN MPLS destination because RPS-DF is disabled. It will be forwarded to other objects in the VPLS. For BUM traffic, there will be a loop, because the frames will be flooded to all objects, including the SAP on PE-2 toward MTU-1.

ALMP in all-active multi-homing SAPs

All-active multi-homing for EVPN MPLS is explained in chapter EVPN for MPLS Tunnels. ALMP is not required on all-active multi-homing SAPs. The following example shows that traffic can be dropped when ALMP is enabled on the SAPs and RPS-DF is enabled on the EVPN-MPLS destinations.

Figure 18: ALMP in all-active multi-homing SAPs shows the example topology for all-active multi-homing.

Figure 18: ALMP in all-active multi-homing SAPs



VPLS is configured with SAP lag-1:1 on the three nodes in the topology, as follows:

```
# on MTU-1, PE-2, PE-3:
configure
    service
        vpls "VPLS 1"
            sap lag-1:1 create
                no shutdown
            exit
```

The SAPs used in the preceding scenarios are removed.

All-active multi-homing is configured on PE-2 and PE-3, as follows:

```
# on PE-2, PE-3:
configure
    service
        system
            bgp-evpn
                ethernet-segment "ESI-23" create
                    esi 01:00:00:00:00:23:00:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 1
                    no shutdown
                exit
            exit
```

ALMP is enabled on the SAPs on PE-2 and PE-3, as follows:

```
# on PE-2, PE-3:
configure
    service
        vpls "VPLS 1"
            sap lag-1:1
                auto-learn-mac-protect
                no shutdown
            exit
```

MAC address aa:aa:01:10:10:10 is learned and protected on PE-2 and PE-3, as follows:

```
*A:PE-2# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                        Age
-------------------------------------------------------------------------------
1          aa:aa:01:10:10:10 sap:lag-1:1             EvpnS:P  03/19/21 08:51:10
1          aa:aa:02:20:20:20 sap:1/2/1:1             LP/0     03/19/21 08:47:53
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

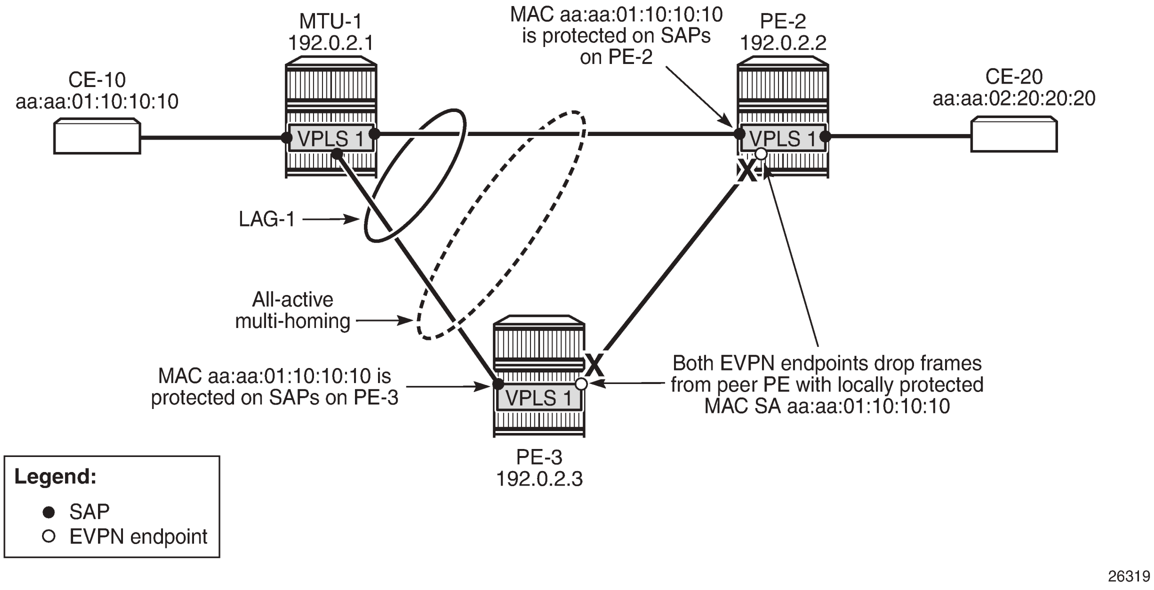
```
*A:PE-3# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                        Age
-------------------------------------------------------------------------------
1          aa:aa:01:10:10:10 sap:lag-1:1             LP/0     03/19/21 08:51:10
1          aa:aa:02:20:20:20 mpls:                   EvpnS:P  03/19/21 08:47:53
                             192.0.2.2:524284
           ldp:65537
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

ALMP in all-active multi-homing, RPS-DF on EVPN MPLS destinations

ALMP is not recommended in all-active multi-homing because it can cause traffic loss. The following example shows when frames are dropped.

Figure 19: All-active multi-homing - RPS-DF on SAPs and EVPN endpoints shows the example setup with MAC address aa:aa:01:10:10:10 protected on SAP lag-1:1 on both PE-2 and PE-3, and RPS-DF enabled on the EVPN endpoints.

Figure 19: All-active multi-homing - RPS-DF on SAPs and EVPN endpoints



When frames with MAC address aa:aa:01:10:10:10 are sent between PE-2 and PE-3, these frames will be dropped by the EVPN MPLS destination that has RPS-DF enabled.

The traffic flows from CE-10 and CE-20 are hashed over both links in the LAG. When the frames are sent out on MTU-1 on port 1/1/1 toward PE-2, the traffic reaches CE-20, and traffic can be sent back from CE-20 to CE-10 via the direct link between PE-2 and MTU-1. However, when traffic is sent out from MTU-1 on port 1/1/2 toward PE-3, the frames will be forwarded from PE-3 to PE-2, where they will be discarded at the EVPN MPLS destination on PE-2 because of RPS-DF. No traffic flow is possible for frames with the protected MAC SA aa:aa:01:10:10:10 via PE-3 to PE-2, or vice versa. If the MAC address is not protected yet on PE-2, the first few messages get through until the MAC address is protected on PE-2. Both multi-homing PEs, PE-2 and PE-3, protect the MAC address aa:aa:01:10:10:10 on their local all-active SAP. Therefore, PE-2 discards all frames with the MAC SA aa:aa:01:10:10:10 when they are received on the EVPN MPLS destination from the other multi-homing PE (PE-3).

An improved mechanism for EVPN loop protection in all-active multi-homing is black-hole MAC duplication, as described in chapter Black-hole MAC for EVPN Loop Protection.

For single-active multi-homing, this problem does not arise: only the designated forwarder in the Ethernet segment receives and forwards traffic. Therefore, the CE MAC addresses will not be learned and protected on different PEs in the same Ethernet segment.

Conclusion

For security, MAC addresses learned on objects, such as SAPs, spoke/mesh-SDPs, and SHGs in EVPN services can be protected and advertised by BGP with the sticky bit set. By default, frames with a protected MAC SA are discarded if received on objects where the MAC address was not learned. Objects can be configured to be shut down when a frame is received with a protected MAC SA that has not been learned locally.

BGP Multi-Homing for VPLS Networks

This chapter describes BGP Multi-Homing (BGP-MH) for VPLS network configurations.

Topics in this chapter include:

Applicability

Initially, the information in this chapter was based on SR OS Release 8.0.R5, with additions for SR OS Release 9.0.R1. The CLI in the current edition corresponds to SR OS Release 20.10.R2.

Overview

SR OS supports the use of Border Gateway Protocol Multi-Homing for VPLS (hereafter called BGP-MH). BGP-MH is described in *draft-ietf-bess-vpls-multihoming, BGP based Multi-homing in Virtual Private LAN Service*, and provides a network-based resiliency mechanism (no interaction from the Provider Edge routers (PEs) to Multi-Tenant Units/Customer Equipment (MTU/CE)) that can be applied on service access points (SAPs) or network (pseudowires) topologies. The BGP-MH procedures will run between the PEs and will provide a loop-free topology from the network perspective (only one logical active path will be provided per VPLS among all the objects SAPs or pseudowires which are part of the same Multi-Homing site).

Each multi-homing site connected to two or more peers is represented by a site ID (2 bytes long) which is encoded in the BGP MH Network Layer Reachability Information (NLRI). The BGP peer holding the active path for a particular multi-homing site will be named as the Designated Forwarder (DF), whereas the rest of the BGP peers participating in the BGP MH process for that site will be named as non-DF and will block the traffic (in both directions) for all the objects belonging to that multi-homing site.

BGP MH uses the following rules to determine which PE is the DF for a particular multi-homing site:

1. A BGP MH NLRI with D flag = 0 (multi-homing object up) always takes precedence over a BGP MH NLRI with D flag = 1 (multi-homing object down). If there is a tie, then:
2. The BGP MH NLRI with the highest BGP Local Preference (LP) wins. If there is a tie, then:
3. The BGP MH NLRI issued from the PE with the lowest PE ID (system address) wins.

The main advantages of using BGP-MH as opposed to other resiliency mechanisms for VPLS are:

- Flexibility: BGP-MH uses a common mechanism for access and core resiliency. The designer has the flexibility of using BGP-MH to control the active/standby status of SAPs, spoke SDPs, Split Horizon Groups (SHGs) or even mesh SDP bindings.
- The standard protocol is based on BGP, a standard, scalable, and well-known protocol.

- Specific benefits at the access:
  - It is network-based, independent of the customer CE and, as such, it does not need any customer interaction to determine the active path. Consequently, the operator will spend less effort on provisioning and will minimize both operation costs and security risks (in particular, this removes the requirement for spanning tree interaction between the PE and CE).
  - Easy load balancing per service (no service fate-sharing) on physical links.
- Specific benefits in the core:
  - It is a network-based mechanism, independent of the MTU resiliency capabilities and it does not need MTU interaction, therefore operational advantages are achieved as a result of the use of BGP-MH: less provisioning is required and there will be minimal risks of loops. In addition, simpler MTUs can be used.
  - Easy load balancing per service (no service fate-sharing) on physical links.
  - Less control plane overhead: there is no need for an additional protocol running the pseudowire redundancy when BGP is already used in the core of the network. BGP-MH just adds a separate NLRI in the L2-VPN family (AFI=25, SAFI=65).

This chapter describes how to configure and troubleshoot BGP-MH for VPLS.

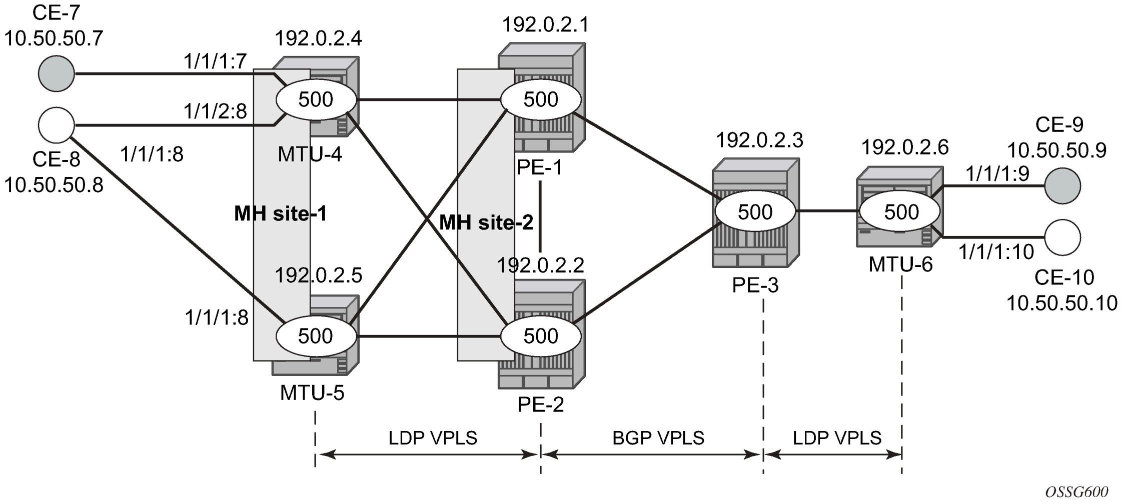
Knowledge of the LDP/BGP VPLS (RFC 4762, *Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling*, and RFC 4761, *Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling*) architecture and functionality is assumed throughout this document. For further information, see the relevant Nokia documentation.

Figure 20: Example topology shows the example topology that will be used throughout the rest of the chapter.

The initial configuration includes:

- IGP — IS-IS, Level 2 on all routers; area 49.0001
- RSVP-TE for transport tunnels
- Fast reroute (FRR) protection in the core; no FRR protection at the access.

Figure 20: Example topology



The topology consists of three core nodes (PE-1, PE-2, and PE-3) and three MTUs connected to the core.

The VPLS service 500 is configured on all the six nodes with the following characteristics:

- The core VPLS instances are connected by a full mesh of BGP-signaled pseudowires (that is, pseudowires among PE-1, PE-2, and PE-3 will be signaled by BGP VPLS).
- As shown in Figure 20: Example topology, the MTUs are connected to the BGP VPLS core by TLDP pseudowires. MTU-6 is connected to PE-3 by a single pseudowire, whereas MTU-4 and MTU-5 are dual-homed to PE-1 and PE-2. The following resiliency mechanisms are used on the dual-homed MTUs:
  - MTU-4 is dual-connected to PE-1 and PE-2 by an active/standby pseudowire (A/S pseudowire hereafter).
  - MTU-5 is dual-connected to PE-1 and PE-2 by two active pseudowires, one of them being blocked by BGP MH running between PE-1 and PE-2. The PE-1 and PE-2 pseudowires, set up from MTU-5, will be part of the BGP MH site MH-site-2.
  - MTU-4 and MTU-5 are running BGP MH, being SHG site-1 and SAP 1/1/1:8 on MTU-5 part of the same BGP MH site, MH-site-1.
- The CEs are connected to the network in the following way:
  - CE-7, CE-9, and CE-10 are single-connected to the network
  - CE-8 is dual connected to MTU-4 and MTU-5.
  - CE-7 and CE-8 are part of the split-horizon group (SHG) site-1(SAPs 1/1/4:500 and 1/1/3:500 on MTU-4). Assume that CE-7 and CE-8 have a backdoor link between them so that when MTU-5 is elected as DF, CE-7 does not get isolated. This configuration highlights the use of a SHG within a site configuration.

For each BGP MH site, MH-site-1 and MH-site-2, the BGP MH process will elect a DF, blocking the site objects for the non-DF nodes. In other words, based on the specific configuration explained throughout the chapter:

- For MH-site-1, MTU-4 will be elected as the DF. The non-DF-MTU-5 will block the SAP 1/1/1:8.
- For MH-site-2, PE-1 will be elected as the DF. The non-DF PE-1 will block the spoke-SDP to MTU-5.

Configuration

This section describes all the relevant configuration tasks for the setup shown in Figure 20: Example topology. The appropriate associated IP/MPLS configuration is out of the scope of this chapter. In this example, the following protocols will be configured beforehand:

- ISIS-TE as IGP with all the interfaces being level-2 (OSPF-TE could have been used instead).
- RSVP-TE as the MPLS protocol to signal the transport tunnels (LDP could have been used instead).
- LSPs between core PEs will be FRR protected (facility bypass tunnels) whereas LSP tunnels between MTUs and PEs will not be protected.

Note:
The designer can choose whether to protect access link failures by means of MPLS FRR or A/S pseudowire or BGP MH. Whereas FRR provides a faster convergence (around 50ms) and stability (it does not impact on the service layer, therefore, link failures do not trigger MAC flush and flooding), some interim inefficiencies can be introduced compared to A/S pseudowire or BGP MH.

Once the IP/MPLS infrastructure is up and running, the specific service configuration including the support for BGP MH can begin.

Global BGP configuration

BGP is used in this configuration guide for these purposes:

1. Auto-discovery and signaling of the pseudowires in the core, as per RFC 4761.
2. Exchange of multi-homing site NLRIs and redundancy handling from MTU-5 to the core.
3. Exchange of multi-homing site NLRIs and redundancy handling at the access for CE-7/CE-8.

A BGP route reflector (RR), PE-3, is used for the reflection of BGP updates corresponding to the preceding uses **a** and **b**.

A direct peering is established between MTU-4 and MTU-5 for use **c**. The same RR could have been used for the three cases, however, like in this example, the designer may choose to have a direct BGP peering between access devices. The reasons for this are:

- By having a direct BGP peering between MTU-4 and MTU-5, the BGP updates do not have to travel back and forth.
- On MTU-4 and MTU-5, BGP is exclusively used for multi-homing, therefore there will not be more BGP peers for either MTUs and a RR adds nothing in terms of control plane scalability.

On all nodes, the autonomous system number must be configured.

```
# on all nodes:
configure
    router Base
        autonomous-system 65000
```

The following CLI output shows the global BGP configuration required on MTU-4. The 192.0.2.5 address will be replaced by the corresponding peer or the RR system address for PE-1 and PE-2.

```
# on MTU-4:
configure
    router Base
        bgp
            family l2-vpn
            rapid-withdrawal
            rapid-update l2-vpn
            group "Multi-Homing"
                neighbor 192.0.2.5
                    peer-as 65000
                exit
            exit
```

In this example, PE-3 is the BGP RR, therefore its BGP configuration will contain a cluster with all its peers included (PE-1 and PE-2):

```
# on PE-3:
configure
    router Base
        bgp
            family l2-vpn
            rapid-withdrawal
            rapid-update l2-vpn
            group "internal"
```

```
                cluster 1.1.1.1
                neighbor 192.0.2.1
                    peer-as 65000
                exit
                neighbor 192.0.2.2
                    peer-as 65000
                exit
            exit
```

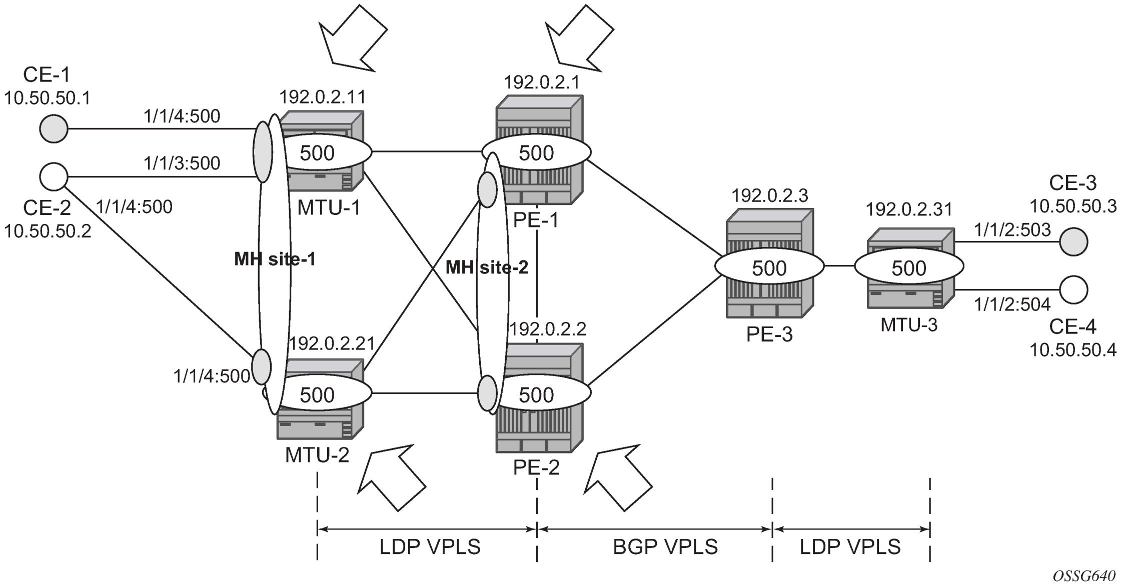
The relevant BGP commands for BGP-MH are in bold. Some considerations about those:

- It is required to specify **family l2-vpn** in the BGP configuration. That statement will allow the BGP peers to agree on the support for the family AFI=25 (Layer 2 VPN), SAFI=65 (VPLS). This family is used for BGP VPLS as well as for BGP MH and BGP AD.
- The **rapid-update l2-vpn** statement allows BGP MH to send BGP updates immediately after detecting link failures, without having to wait for the Minimum Route Advertisement Interval (MRAI) to send the updates in batches. This statement is required to guarantee a fast convergence for BGP MH.
- Optionally, **rapid-withdrawal** can also be added. In the context of BGP MH, this command is only useful if a particular multi-homing site is cleared. In that case, a BGP withdrawal is sent immediately without having to wait for the MRAI. A multi-homing site is cleared when the BGP MH site is removed or even the entire VPLS service.

Service level configuration

Once the IP/MPLS infrastructure is configured, including BGP, this section shows the configuration required at service level (VPLS 500). The focus is on the nodes involved on BGP MH, that is, MTU-4, MTU-5, PE-1, and PE-2. These nodes are highlighted in Figure 21: Nodes involved in BGP MH.

Figure 21: Nodes involved in BGP MH



Core PE service configuration

The following CLI excerpt shows the service level configuration on PE-1. The import/export policies configured on the PE nodes are identical:

```
# on PE-1:
configure
    router Base
        policy-options
            begin
            community "comm_core"
                members "target:65000:500"
            exit
            policy-statement "vsi500_export"
                entry 10
                    action accept
                        community add "comm_core"
                    exit
                exit
            exit
            policy-statement "vsi500_import"
                entry 10
                    from
                        community "comm_core"
                        family l2-vpn
                    exit
                    action accept
                    exit
                exit
                default-action drop
                exit
            exit
            commit
```

The configuration of the SDPs, PW template, and VPLS on PE-1 is as follows:

```
# on PE-1:
configure
    service
        sdp 12 mpls create
            description "SDP to transport BGP-signaled PWs"
            signaling bgp
            far-end 192.0.2.2
            lsp "LSP-PE-1-PE-2"
            path-mtu 8000
            no shutdown
        exit
        sdp 13 mpls create
            description "SDP to transport BGP-signaled PWs"
            signaling bgp
            far-end 192.0.2.3
            lsp "LSP-PE-1-PE-3"
            path-mtu 8000
            no shutdown
        exit
        sdp 14 mpls create
            far-end 192.0.2.4
            lsp "LSP-PE-1-MTU-4"
            path-mtu 8000
            no shutdown
        exit
        sdp 15 mpls create
            far-end 192.0.2.5
            lsp "LSP-PE-1-MTU-5"
            path-mtu 8000
            no shutdown
        exit
        pw-template 500 use-provisioned-sdp create
        exit
        vpls 500 name "VLPS 500" customer 1 create
            bgp
                route-distinguisher 65000:501
                vsi-export "vsi500_export"
                vsi-import "vsi500_import"
                pw-template-binding 500 split-horizon-group "CORE"
                exit
            exit
            bgp-vpls
                max-ve-id 65535
                ve-name 501
                    ve-id 501
                exit
                no shutdown
            exit
            site "MH-site-2" create
                site-id 2
                spoke-sdp 15:500
                no shutdown
            exit
            spoke-sdp 14:500 create
            exit
            spoke-sdp 15:500 create
            exit
            no shutdown
        exit
```

The following are general comments about the configuration of VPLS 500:

- As seen in the preceding CLI output for PE-1, there are four provisioned SDPs that the service VPLS 500 will use in this example. SDP 14 and SDP 15 are tunnels over which the TLDP FEC128 pseudowires for service 500 will be carried (according to RFC 4762), whereas SDP 12 and SDP 13 are the tunnels for the core BGP pseudowires (based on RFC 4761).
- The BGP context provides the general service BGP configuration that will be used by BGP VPLS and BGP MH:
  - Route distinguisher (notation chosen is based on <AS_number:500 + node_id>)
  - VSI export policies are used to add the export route-targets included in all the BGP updates sent to the BGP peers.
  - VSI import policies are used to control the NLRIs accepted in the RIB, normally based on the route targets.
  - Both VSI-export and VSI-import policies can be used to modify attributes such as the Local Preference (LP) that will be used to influence the BGP MH Designated Forwarder (DF) election (LP is the second rule in the BGP MH election process, as previously discussed). The use of these policies will be described later in the chapter.
  - The **pw-template-binding** command maps the previously defined pw-template 500 to the SHG “CORE”. In this way, all the BGP-signaled pseudowires will be part of this SHG. Although not shown in this example, the **pw-template-binding** command can also be used to instantiate pseudowires within different SHGs, based on different import route targets:

Note:
Detailed BGP-VPLS configuration is out of the scope of this chapter. For more information, see chapter *BGP-VPLS*.

```
*A:PE-1# configure service vpls 500 bgp pw-template-binding ?
  - pw-template-binding <policy-id> [split-horizon-group <group-name>]
                                      [import-rt {ext-community,...(upto 5 max)}]
  - no pw-template-binding <policy-id>

---snip---
```

- The BGP-signaled pseudowires (from PE-1 to PE-2 and PE-3) are set up according to the configuration in the **bgp** context. Beside those pseudowires, the VPLS 500 also has two more pseudowires signaled by TLDP: spoke-SDP 14:500 (to MTU-4) and spoke-SDP 15:500 (to MTU-5).

The general BGP MH configuration parameters for a particular multi-homing site are shown in the following output:

```
*A:PE-1# configure service vpls ?
  - no vpls <service-id>
  - vpls <service-id> [customer <customer-id>] [create] [vpn <vpn-id>] [m-vpls]
                                             [b-vpls|i-vpls] [etree] [name <name>]

---snip---
```

```
*A:PE-1# configure service vpls 500 site ?
  - no site <name>
  - site <name> [create]

 <name>               : [32 chars max]

 [no] boot-timer      - Configure/Override site boot-timer
      failed-thresho* - Configure threshold for the site to be declared down
 [no] mesh-sdp-bindi* - Enable/Disable application to all Mesh-SDP
 [no] monitor-oper-g* - Configure an Operational-Group to monitor
 [no] sap             - Configure a SAP for the site
 [no] shutdown        - Administratively enable/disable the site
 [no] site-activatio* - Configure/Override site activation timer
 [no] site-id         - Configure site identifier
 [no] site-min-down-* - Configure minimum down timer for the site
 [no] split-horizon-* - Configure a split-horizon-group
 [no] spoke-sdp       - Configure a spoke-SDP
```

Where:

- The **site** *name* is defined by a string of up to 32 characters.
- The **site-id** is an integer that identifies the multi-homing site and is encoded in the BGP MH NLRI. This ID must be the same one used on the peer node where the same multi-homing site is connected to. That is, MH-site-2 must use the same site-id in PE-1 and PE-2 (value = 2 in the PE-1 site configuration).
- Out of the four potential objects in a site—spoke SDP, SAP, SHG, and mesh SDP binding—only one can be used at the time on a particular site. To add more than just one SAP/spoke-SDP to the same site, an SHG composed of the SAP/spoke-SDP objects must be used in the site configuration.

Otherwise, only one object—spoke SDP, SAP, SHG, or mesh SDP binding—is allowed per site. A CLI log message warns the operator of such fact:

```
*A:PE-1>config>service>vpls>site# mesh-sdp-binding
MINOR: SVCMGR #5855 only one object is allowed per site
```

- The **failed-threshold** command defines how many objects should be down for the site to be declared down. This command is obviously only valid for multi-object sites (SHGs and mesh-SDP bindings). By default, all the objects in a site must be down for the site to be declared as operationally down.

```
*A:PE-1>config>service>vpls>site# failed-threshold ?
  - failed-threshold <[1..1000]>
  - failed-threshold all
```

- The **boot-timer** specifies for how long the service manager waits after a node reboot before running the MH procedures. The boot-timer value should be configured to allow for the BGP sessions to come up and for the NLRI information to be refreshed/exchanged. In environments with the default BGP MRAI (30 seconds), it is highly recommended to increase this value (for instance, 120 seconds for a normal configuration). The **boot-timer** is only important when a node comes back up and would become the DF. Default value: 10 seconds.

```
*A:PE-1>config>service>vpls>site# boot-timer ?
  - boot-timer <seconds>
  - no boot-timer

 <seconds>            : [0..600]
```

- The **site-activation-timer** command defines the amount of time the service manager will keep the local objects in standby (in the absence of BGP updates from remote PEs) before running the DF election algorithm to decide whether the site should be unblocked. The timer is started when one of the following events occurs only if the site is operationally up:
  - Manual site activation using the **no shutdown** command at the site-id level or at member object(s) level (SAP(s) or pseudowire(s))
  - Site activation after a failure
  - The BGP MH election procedures will be resumed upon expiration of this timer or the arrival of a BGP MH update for the multi-homing site. Default value: 2 seconds.

```
*A:PE-1>config>service>vpls>site# site-activation-timer ?
  - no site-activation-timer
  - site-activation-timer <seconds>

 <seconds>            : [0..100]
```

- When a BGP MH site goes down, it may be preferred that it stays down for a minimum time. This is configurable by the **site-min-down-timer**. When set to zero, this timer is disabled.

```
*A:PE-1>config>service>vpls>site# site-min-down-timer ?
  - no site-min-down-timer
  - site-min-down-timer <seconds>

 <seconds>            : [0..100]
```

- The **boot-timer**, **site-activation-timer**, and **site-min-down-timer** commands can be provisioned at service level or at global level. The service level settings have precedence and override the global configuration. The **no** form of the commands at global level, sets the value back to the default values. The **no** form of the commands at service level, makes the timers inherit the global values.

```
*A:PE-1# configure redundancy bgp-multi-homing ?
  - bgp-multi-homing

 [no] boot-timer      - Configure BGP multi-homing boot-timer
 [no] site-activatio* - Configure BGP multi-homing site activation timer
 [no] site-min-down-* - Configure minimum down timer for the site
```

- The **shutdown** command controls the admin state of the site. Each site has three possible states:
  - Admin state — controlled by the shutdown command.
  - Operational state — controlled by the operational status of the individual site objects.
  - Designated Forwarder (DF) state — controlled by the BGP MH election algorithm.

The following CLI output shows the three states for BGP MH site "MH-site-1" on MTU-5:

```
*A:MTU-5# show service id 500 site "MH-site-1"

===============================================================================
Site Information
===============================================================================
Site Name            : MH-site-1
-------------------------------------------------------------------------------
Site Id              : 1
Dest                 : sap:1/1/1:8          Mesh-SDP Bind    : no
Admin Status         : Enabled              Oper Status      : up
Designated Fwdr      : No
DF UpTime            : 0d 00:00:00          DF Chg Cnt       : 1
Boot Timer           : default              Timer Remaining  : 0d 00:00:00
Site Activation Timer: default              Timer Remaining  : 0d 00:00:00
Min Down Timer       : default              Timer Remaining  : 0d 00:00:00
Failed Threshold     : default(all)
Monitor Oper Grp     : (none)
===============================================================================
```

For this example, MH-site " MH-site-2" is configured in PE-1, where the site-id is 2 and the object in the site is spoke-SDP 15:500 (pseudowire established from PE-1 to MTU-5).

The following CLI shows the service configuration for PE-2. The site-id is 2, that is, the same value configured in PE-1. The object defined in PE-2's site is spoke-SDP 25:500 (pseudowire established from PE-2 to MTU-5).

```
# on PE-2:
configure
    service
        sdp 21 mpls create
            description "SDP to transport BGP-signaled PWs"
            signaling bgp
            far-end 192.0.2.1
            lsp "LSP-PE-2-PE-1"
            path-mtu 8000
            no shutdown
        exit
        sdp 23 mpls create
            description "SDP to transport BGP-signaled PWs"
            signaling bgp
            far-end 192.0.2.3
            lsp "LSP-PE-2-PE-3"
            path-mtu 8000
            no shutdown
        exit
        sdp 24 mpls create
            far-end 192.0.2.4
            lsp "LSP-PE-2-MTU-4"
            path-mtu 8000
            no shutdown
        exit
        sdp 25 mpls create
            far-end 192.0.2.5
            lsp "LSP-PE-2-MTU-5"
            path-mtu 8000
            no shutdown
        exit
        pw-template 500 use-provisioned-sdp create
        exit
        vpls 500 name "VPLS 500" customer 1 create
            bgp
                route-distinguisher 65000:502
                vsi-export "vsi500_export"
                vsi-import "vsi500_import"
                pw-template-binding 500 split-horizon-group "CORE"
                exit
            exit
            bgp-vpls
                max-ve-id 65535
                ve-name 502
                    ve-id 502
                exit
                no shutdown
            exit
            site "MH-site-2" create
                site-id 2
                spoke-sdp 25:500
                no shutdown
            exit
            spoke-sdp 24:500 create
            exit
            spoke-sdp 25:500 create
            exit
            no shutdown
        exit
```

MTU service configuration

The service configuration in MTU-4 is as follows:

```
# on MTU-4:
configure
    service
        sdp 41 mpls create
            far-end 192.0.2.1
            lsp "LSP-MTU-4-PE-1"
            path-mtu 8000
            no shutdown
        exit
        sdp 42 mpls create
            far-end 192.0.2.2
            lsp "LSP-MTU-4-PE-2"
            path-mtu 8000
            no shutdown
        exit
        vpls 500 name "VPLS 500" customer 1 create
            endpoint "CORE" create
                no suppress-standby-signaling
            exit
            split-horizon-group "site-1" create
            exit
            bgp
                route-distinguisher 65000:504
                route-target export target:65000:500 import target:65000:500
            exit
            site "MH-site-1" create
                site-id 1
                split-horizon-group site-1
                no shutdown
            exit
            sap 1/1/1:7 split-horizon-group "site-1" create
            exit
            sap 1/1/2:8 split-horizon-group "site-1" create
                eth-cfm
                    mep 48 domain 1 association 1 direction down
                        fault-propagation-enable use-if-tlv
                        ccm-enable
                        no shutdown
                    exit
                exit
            exit
            spoke-sdp 41:500 endpoint "CORE" create
                precedence primary
            exit
            spoke-sdp 42:500 endpoint "CORE" create
            exit
            no shutdown
        exit
```

MTU-4 is configured with the following characteristics:

- The BGP context provides the general BGP parameters for service 500 in MTU-4. The **route-target** command is now used instead of the **vsi-import** and **vsi-export** commands. The intent in this example is to configure only the export and import route-targets. There is no need to modify any other attribute. If the local preference is to be modified (to influence the DF election), a **vsi-policy** must be configured.
- An A/S pseudowire configuration is used to control the pseudowire redundancy towards the core.
- The multi-homing site, MH-site-1 has a site-id = 1 and an SHG as an object. The SHG site-1 is composed of SAP 1/1/1:7 and SAP 1/1/2:8. As previously discussed, the site will not be declared operationally down until the two SAPs belonging to the site are down. This behavior can be changed by the **failed-threshold** command (for instance, in order to bring the site down when only one object has failed even though the second SAP is still up).
- As an example, a Y.1731 MEP with fault-propagation has been defined in SAP 1/1/2:8. As discussed later in the chapter, this MEP will signal the status of the SAP (as a result of the BGP MH process) to CE-8.

The service configuration in MTU-5 is as follows:

```
# on MTU-5:
configure
    service
        sdp 51 mpls create
            far-end 192.0.2.1
            lsp "LSP-MTU-5-PE-1"
            path-mtu 8000
            no shutdown
        exit
        sdp 52 mpls create
            far-end 192.0.2.2
            lsp "LSP-MTU-5-PE-2"
            path-mtu 8000
            no shutdown
        exit
        vpls 500 name "VPLS 500" customer 1 create
            bgp
                route-distinguisher 65000:505
                route-target export target:65000:500 import target:65000:500
            exit
            site "MH-site-1" create
                site-id 1
                sap 1/1/1:8
                no shutdown
            exit
            sap 1/1/1:8 create
            exit
            spoke-sdp 51:500 create
            exit
            spoke-sdp 52:500 create
            exit
            no shutdown
        exit
```

Influencing the DF election

As previously explained, assuming that the sites on the two nodes taking part of the same multi-homing site are both up, the two tie-breakers for electing the DF are (in this order):

1. Highest LP
2. Lowest PE ID

The LP by default is 100 in all the routers. Under normal circumstances, if the LP in any router is not changed, MTU-4 will be elected the DF for MH-site-1, whereas PE-1 will be the DF for MH-site-2. Assume in this section that this behavior is changed for MH-site-2 to make PE-2 the DF. Because changing the system address (to make PE-2’s ID the lower of the two IDs) is usually not an easy task to accomplish, the vsi-export policy on PE-2 is modified with an LP of 150 with which the MH-site-2 NLRI is announced to PE-1. Because LP 150 is greater than the default 100 in PE-1, PE-2 will be elected as the DF for MH-site-2. The vsi-import policy remains unchanged and the vsi-export policy is modified as follows:

```
# on PE-2:
configure
    router Base
        policy-options
            begin
            community "comm_core"
                members "target:65000:500"
            exit
            policy-statement "vsi500_export"
                entry 10
                    action accept
                        community add "comm_core"
                        local-preference 150
                    exit
                exit
            exit
            policy-statement "vsi500_import"
                entry 10
                    from
                        community "comm_core"
                        family l2-vpn
                    exit
                    action accept
                    exit
                exit
                default-action drop
                exit
            exit
            commit
```

In PE-1, the import and export policies are not modified. The policies were already applied in the **bgp** context of VPLS 500, as follows:

```
# on PE-2:
configure
    service
        vpls "VPLS 500"
            bgp
                route-distinguisher 65000:502
                vsi-export "vsi500_export"
                vsi-import "vsi500_import"
                pw-template-binding 500 split-horizon-group "CORE"
                exit
            exit
---snip---
```

The DF state of PE-2 can be verified as follows:

```
*A:PE-2# show service id 500 site "MH-site-2"

===============================================================================
Site Information
===============================================================================
Site Name            : MH-site-2
-------------------------------------------------------------------------------
Site Id              : 2
Dest                 : sdp:25:500           Mesh-SDP Bind    : no
Admin Status         : Enabled              Oper Status      : up
Designated Fwdr      : Yes
DF UpTime            : 0d 00:00:29          DF Chg Cnt       : 1
Boot Timer           : default              Timer Remaining  : 0d 00:00:00
Site Activation Timer: default              Timer Remaining  : 0d 00:00:00
Min Down Timer       : default              Timer Remaining  : 0d 00:00:00
Failed Threshold     : default(all)
Monitor Oper Grp     : (none)
===============================================================================
```

The import and export policies are applied at service 500 level, which means that the LP changes for all the potential multi-homing sites configured under service 500. Therefore, load balancing can be achieved on a per-service basis, but not within the same service.

These policies are applied on the VPLS 500 for all the potential BGP applications: BGP VPLS, BGP MH, and BGP AD. In the example, the LP for the PE-2 BGP updates for BGP MH and BGP VPLS will be set to 150. However, this has no impact on BGP VPLS because a PE cannot receive two BGP VPLS NLRIs with the same VE-ID, which implies that a different VE-ID per PE within the same VPLS is required.

The vsi-export policy is restored to its original settings on PE-2, as follows:

```
# on PE-2:
configure
    router Base
        policy-options
            begin
            policy-statement "vsi500_export"
                entry 10
                    action accept
                        community add "comm_core"
                        no local-preference
                    exit
                exit
            exit
            commit
```

In all the PE nodes, the import and export policies applied in the **bgp** context of VPLS 500 have identical settings again, and PE-1 is the DF.

Black-hole avoidance

SR OS supports the appropriate MAC flush mechanisms for BGP MH, regardless of the protocol being used for the pseudowire signaling:

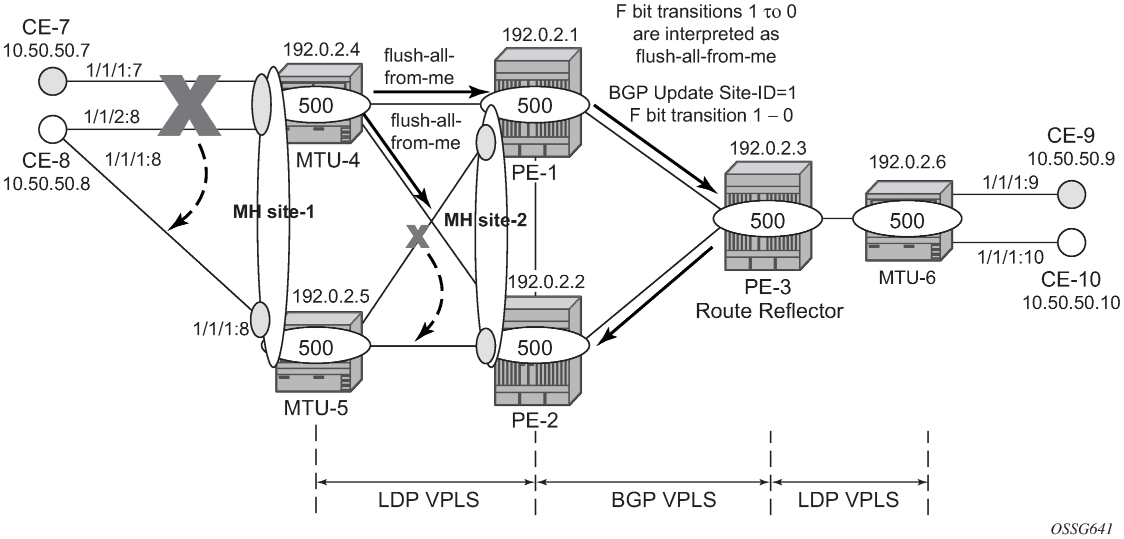
- LDP VPLS — The PE that contains the old DF site (the site that just experienced a DF to non-DF transition) always sends a LDP MAC flush-all-from-me to all LDP pseudowires in the VPLS, including the LDP pseudowires associated with the new DF site. No specific configuration is required.
- BGP VPLS — The remote BGP VPLS PEs interpret the F bit transitions from 1 to 0 as an implicit MAC flush-all-from-me indication. If a BGP update with the flag F=0 is received from the previous DF PE, the remote PEs perform MAC flush-all-from-me, flushing all the MACs associated with the pseudowire to the old DF PE. No specific configuration is required.

Double flushing will not happen because it is expected that between any pair of PEs there will exist only one type of pseudowires—either BGP or LDP pseudowire—, but not both types.

In the example, assuming MTU-4 and PE-1 are the DF nodes:

- When MH-site-1 is brought operationally down on MTU-4 (so by default, the two SAPs must go down unless the **failed-threshold** parameter is changed so that the site is down when only one SAP is brought down), MTU-4 will issue a flush-all-from-me message.
- When MH-site-2 is brought operationally down on PE-1, a BGP update with F=0 and D=1 is issued by PE-1. PE-2 and PE-3 will receive the update and will flush the MAC addresses learned on the pseudowire to PE-1.

Figure 22: MAC flush for BGP MH



Node failures implicitly trigger a MAC flush on the remote nodes, because the TLDP/BGP session to the failed node goes down.

Access CE/PE signaling

BGP MH works at service level, therefore no physical ports are torn down on the non-DF, but rather the objects are brought down operationally, while the physical port will stay up and used for any other services existing on that port. Due to this reason, there is a need for signaling the standby status of an object to the remote PE or CE.

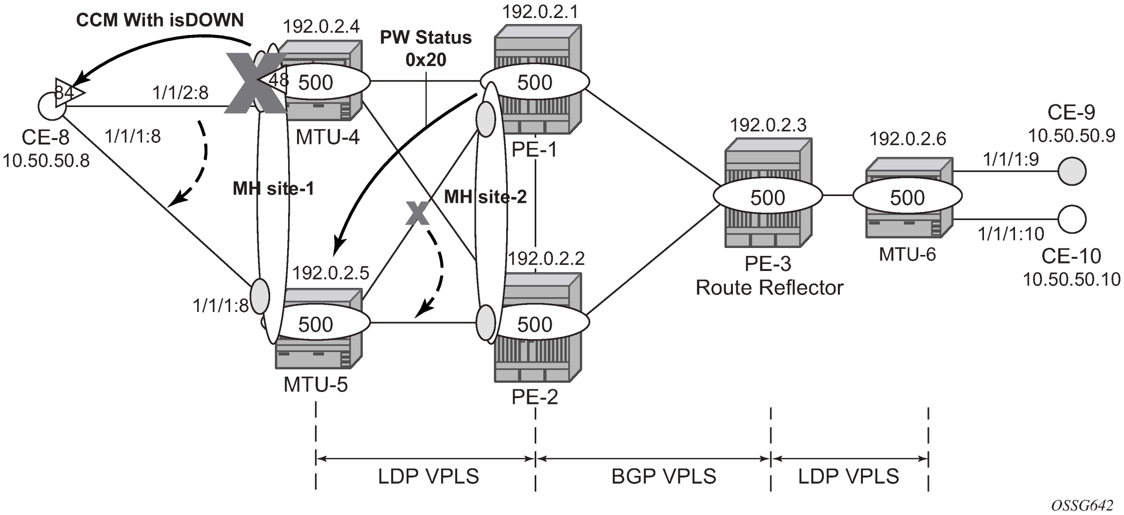
- Access PEs running BGP MH on spoke SDPs and elected non-DF, will signal pseudowire standby status (0x20) to the other end. If no pseudowire status is supported on the remote MTU, a label withdrawal is performed. If there is more than one spoke SDP on the site (part of the same SHG), the signaling is sent for all the pseudowires of the site.

Note:
The **configure service vpls x spoke-sdp y:z no pw-status-signaling** parameter allows to send a TLDP label-withdrawal instead of pseudowire status bits, even though the peer supports pseudowire status.

- Multi-homed CEs connected through SAPs to the PEs running BGP MH, are signaled by the PEs using Y.1731 CFM, either by stopping the transmission of CCMs or by sending CCMs with isDown (interface status down encoding in the interface status TLV).

In this example, down MEPs on MTU-4 SAP 1/1/2:8 and CE-8 SAP 1/1/2:8 are configured. In a similar way, other MEPs can be configured on MTU-4 SAP 1/1/1:7, MTU-5 SAP 1/1/1:8, and CE-8 SAP 1/1/1:7 and SAP 1/1/1:8. Figure 23: Access PE/CE signaling shows the MEPs on MTU-4 SAP 1/1/2:8 and CE-8. Upon failure on the MTU-4 site MH-site-1, the MEP 48 will start sending CCMs with interface status down.

Figure 23: Access PE/CE signaling



The CFM configuration required at SAP 1/1/2:8 is as follows. Down MEPs will be configured on CE-8 and MTU-5 SAPs in the same way, but in a different association. The option **fault-propagation-enable use-if-tlv** must be added. In case the CE does not understand the CCM interface status TLV, the **fault-propagation-enable suspend-ccm** option can be enabled instead. This will stop the transmission of CCMs upon site failures. Detailed configuration guidelines for Y.1731 are beyond the scope of this chapter.

```
# on MTU-4:
configure
    eth-cfm
        domain 1 format none level 3 admin-name "domain-1"
            association 1 format icc-based name "Association48" admin-name "assoc-1"
                bridge-identifier 500
                exit
                ccm-interval 1
                remote-mepid 84
            exit
        exit
```

```
# on MTU-4:
configure
    service
        vpls "VPLS 500"
            sap 1/1/2:8 split-horizon-group "site-1" create
                eth-cfm
                    mep 48 domain 1 association 1 direction down
                        fault-propagation-enable use-if-tlv
                        ccm-enable
                        no shutdown
                    exit
                exit
            exit
```

If CE-8 is a service router, upon receiving a CCM with isDown, an alarm will be triggered and the SAP will be brought down:

```
# on CE-8:
67 2021/01/19 09:13:19.447 UTC WARNING: OSPF #2047 vprn8 VR:  2 OSPFv2 (0)
```

```
"LCL_RTR_ID 10.50.50.8: Interface int-CE-8-MTU-4 state changed to down
(event IF_DOWN)"

66 2021/01/19 09:13:19.447 UTC WARNING: SNMP #2004 vprn8 int-CE-8-MTU-4
"Interface int-CE-8-MTU-4 is not operational"

65 2021/01/19 09:13:19.447 UTC MINOR: SVCMGR #2203 vprn8
"Status of SAP 1/1/2:8 in service 8 (customer 1) changed to admin=up oper=down
 flags=OamDownMEPFault "

64 2021/01/19 09:13:19.447 UTC MINOR: SVCMGR #2108 vprn8
"Status of interface int-CE-8-MTU-4 in service 8 (customer 1) changed to admin=up
 oper=down"

63 2021/01/19 09:13:19.447 UTC MINOR: ETH_CFM #2001 Base
"MEP 1/1/84 highest defect is now defRemoteCCM"
```

On CE-8, the status of the SAP can be verified as follows:

```
*A:CE-8# show service id 8 sap 1/1/2:8

===============================================================================
Service Access Points(SAP)
===============================================================================
Service Id         : 8
SAP                : 1/1/2:8                  Encap             : q-tag
Description        : (Not Specified)
Admin State        : Up                       Oper State        : Down
Flags              : OamDownMEPFault
Multi Svc Site     : None
Last Status Change : 01/19/2021 09:13:19
Last Mgmt Change   : 01/19/2021 09:00:42
===============================================================================
```

As also depicted in Figure 23: Access PE/CE signaling, PE-1 will signal pseudowire status standby (code 0x20) when PE-1 goes to non-DF state for MH-site-2. MTU-5 will receive that signaling and, based on the **ignore-standby-signaling** parameter, will decide whether to send the broadcast, unknown unicast, and multicast (BUM) traffic to PE-1. In case MTU-5 uses in its configuration **ignore-standby-signaling**, it will be sending BUM traffic on both pseudowires at the same time (which is not normally desired), ignoring the pseudowire status bits. The following output shows the MTU-5 spoke-SDP receiving the pseudowire status signaling. Although the spoke SDP stays operationally up, the Peer Pw Bits field shows *pwFwdingStandby* and MTU-5 will not send any traffic if the **ignore-standby-signaling** parameter is disabled.

```
*A:MTU-5# show service id 500 sdp 51:500 detail

===============================================================================
Service Destination Point (Sdp Id : 51:500) Details
===============================================================================
-------------------------------------------------------------------------------
 Sdp Id 51:500  -(192.0.2.1)
-------------------------------------------------------------------------------
Description     : (Not Specified)
SDP Id             : 51:500                   Type              : Spoke
Spoke Descr     : (Not Specified)
Split Horiz Grp    : (Not Specified)
Etree Root Leaf Tag: Disabled                 Etree Leaf AC     : Disabled
VC Type            : Ether                    VC Tag            : n/a
Admin Path MTU     : 8000                     Oper Path MTU     : 8000
Delivery           : MPLS
Far End            : 192.0.2.1                Tunnel Far End    : n/a
Oper Tunnel Far End: 192.0.2.1
```

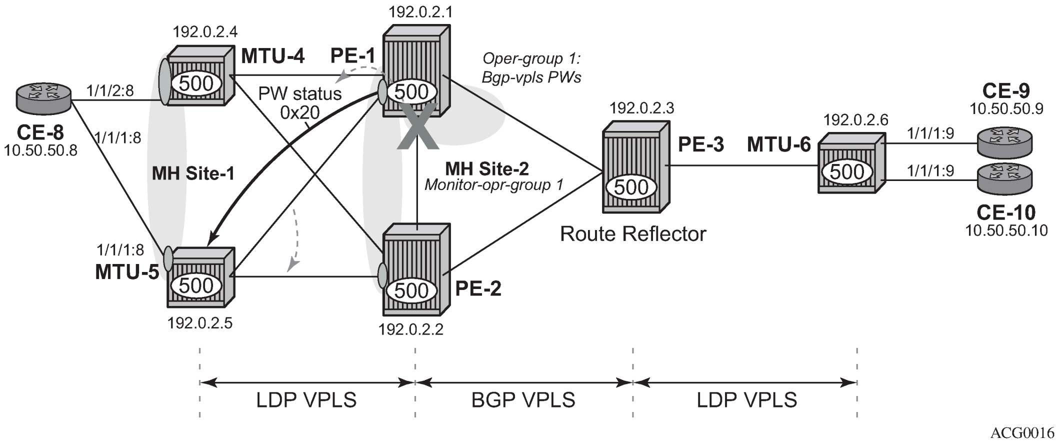
```
LSP Types          : RSVP
---snip---

Admin State        : Up                          Oper State        : Up
---snip---

Endpoint           : N/A                         Precedence        : 4
PW Status Sig      : Enabled
Force Vlan-Vc      : Disabled                    Force Qinq-Vc     : none
Class Fwding State : Down
Flags              : None
Time to RetryReset : never                       Retries Left      : 3
Mac Move           : Blockable                   Blockable Level   : Tertiary
Local Pw Bits      : None
Peer Pw Bits       : pwFwdingStandby
---snip---
```

Operational groups for BGP-MH

Operational groups (**oper-group**) introduce the capability of grouping objects into a generic group object and associating its status to other service endpoints (pseudowires, SAPs, IP interfaces) located in the same or in different service instances. The operational group status is derived from the status of the individual components using certain rules specific to the application using the concept. A number of other service entities—the monitoring objects—can be configured to monitor the operational group status and to drive their own status based on the **oper-group** status. In other words, if the operational group goes down, the monitoring objects will be brought down. When one of the objects included in the operational group comes up, the entire group will also come up, and therefore so will the monitoring objects.

This concept can be used to enhance the BGP-MH solution for avoiding black-holes on the PE selected as the DF if the rest of the VPLS endpoints fail (pseudowire spoke(s)/pseudowire mesh and/or SAP(s)). Figure 24: Oper-groups and BGP-MH illustrates the use of operational groups together with BGP-MH. On PE-1 (and PE-2) all of the BGP-VPLS pseudowires in the core are configured under the same **oper-group group-1**. MH-site-2 is configured as a monitoring object. When the two BGP-VPLS pseudowires go down, **oper-group group-1** will be brought down, therefore MH-site-2 on PE-1 will go down as well (PE-2 will become DF and PE-1 will signal standby to MTU-5).

Figure 24: Oper-groups and BGP-MH



In the preceding example, this feature provides a solution to avoid a black-hole when PE-1 loses its connectivity to the core.

Operational groups are configured in two steps:

1. Identify a set of objects whose forwarding state should be considered as a whole group, then group them under an operational group (in this case **oper-group group-1**, which is configured in the **bgp pw-template-binding** context).
2. Associate other existing objects (clients) with the oper-group using the **monitor-group** command (configured, in this case, in the **site MH-site-2**).

The following CLI excerpt shows the commands required (**oper-group**, **monitor-oper-group**).

```
# on PE-1:
configure
    service
        oper-group "group-1" create
        exit
        vpls 500
            bgp
                pw-template-binding 500 split-horizon-group "CORE"
                    oper-group "group-1"
                exit
            exit
            site "MH-site-2"
                monitor-oper-group "group-1"
            exit
```

When all the BGP-VPLS pseudowires go down, **oper-group group-1** will go down and therefore the monitoring object, **site MH-site-2**, will also go down and PE-2 will then be elected as DF. The log 99 gives information about this sequence of events:

```
# on PE-1:
configure
    service
        sdp 12
            shutdown
        exit
        sdp 13
            shutdown
        exit
```

```
*A:PE-1# show log log-id 99
---snip---

147 2021/01/19 09:20:08.753 UTC WARNING: SVCMGR #2531 Base BGP-MH
"Service-id 500 site MH-site-2 is not the designated-forwarder"

146 2021/01/19 09:20:08.753 UTC MAJOR: SVCMGR #2316 Base
"Processing of a SDP state change event is finished and the status of all affected SDP Bindings
 on SDP 13 has been updated."

145 2021/01/19 09:20:08.752 UTC MINOR: SVCMGR #2306 Base
"Status of SDP Bind 15:500 in service 500 (customer 1) changed to admin=up oper=down flags="

144 2021/01/19 09:20:08.752 UTC MINOR: SVCMGR #2326 Base
"Status of SDP Bind 15:500 in service 500 (customer 1) local PW status bits changed to pwFwding
Standby "

143 2021/01/19 09:20:08.752 UTC MINOR: SVCMGR #2542 Base
```

```
"Oper-group group-1 changed status to down"
```

PE-1 is no longer the DF, as follows:

```
*A:PE-1# show service id 500 site

===============================================================================
VPLS Sites
===============================================================================
Site                Site-Id   Dest                Mesh-SDP  Admin   Oper  Fwdr
-------------------------------------------------------------------------------
MH-site-2           2         sdp:15:500          no        Enabled down  No
-------------------------------------------------------------------------------
Number of Sites : 1
-------------------------------------------------------------------------------
===============================================================================
```

PE-2 becomes the DF:

```
*A:PE-2# show service id 500 site

===============================================================================
VPLS Sites
===============================================================================
Site                Site-Id   Dest                Mesh-SDP  Admin   Oper  Fwdr
-------------------------------------------------------------------------------
MH-site-2           2         sdp:25:500          no        Enabled up    Yes
-------------------------------------------------------------------------------
Number of Sites : 1
-------------------------------------------------------------------------------
===============================================================================
```

The process reverts when at least one BGP-VPLS pseudowire comes back up.

Show commands and debugging options

The main command to find out the status of a site is the **show service id x site** command.

```
*A:MTU-5# show service id 500 site

===============================================================================
VPLS Sites
===============================================================================
Site                Site-Id   Dest                Mesh-SDP  Admin   Oper  Fwdr
-------------------------------------------------------------------------------
MH-site-1           1         sap:1/1/1:8         no        Enabled up    No
-------------------------------------------------------------------------------
Number of Sites : 1
-------------------------------------------------------------------------------
===============================================================================
```

A **detail** modifier is available:

```
*A:MTU-5# show service id 500 site detail

===============================================================================
Site Information
===============================================================================
Site Name            : MH-site-1
```

```
-------------------------------------------------------------------------------
Site Id              : 1
Dest                 : sap:1/1/1:8          Mesh-SDP Bind    : no
Admin Status         : Enabled              Oper Status      : up
Designated Fwdr      : No
DF UpTime            : 0d 00:00:00          DF Chg Cnt       : 1
Boot Timer           : default              Timer Remaining  : 0d 00:00:00
Site Activation Timer: default              Timer Remaining  : 0d 00:00:00
Min Down Timer       : default              Timer Remaining  : 0d 00:00:00
Failed Threshold     : default(all)
Monitor Oper Grp     : (none)
-------------------------------------------------------------------------------
Number of Sites : 1
===============================================================================
```

The **detail** view of the command displays information about the BGP MH timers. The values are only shown if the global values are overridden by specific ones at service level (and will be tagged with *Ovr* if they have been configured at service level). The **Timer Remaining** field reflects the count down from the boot/site activation timers down to the moment when this router tries to become DF again. Again, this is only shown when the global timers have been overridden by the ones at service level.

The objects on the non-DF site will be brought down operationally and flagged with *StandByForMHProtocol*, for example, for SAP 1/1/1:8 on non-DF MTU-5:

```
*A:MTU-5# show service id 500 sap 1/1/1:8

===============================================================================
Service Access Points(SAP)
===============================================================================
Service Id         : 500
SAP                : 1/1/1:8                  Encap             : q-tag
Description        : (Not Specified)
Admin State        : Up                       Oper State        : Down
Flags              : StandByForMHProtocol
Multi Svc Site     : None
Last Status Change : 01/19/2021 08:30:37
Last Mgmt Change   : 01/19/2021 08:47:52
===============================================================================
```

For spoke SDP 25:500 on non-DF PE-2:

```
*A:PE-2# show service id 500 sdp 25:500 detail

===============================================================================
Service Destination Point (Sdp Id : 25:500) Details
===============================================================================
-------------------------------------------------------------------------------
 Sdp Id 25:500  -(192.0.2.5)
-------------------------------------------------------------------------------
Description     : (Not Specified)
SDP Id             : 25:500                   Type              : Spoke
---snip---

Admin State        : Up                       Oper State        : Down
---snip---

Flags              : StandbyForMHProtocol
---snip---
```

The BGP MH routes in the RIB, RIB-In and RIB-Out can be shown by using the corresponding **show router bgp routes** and **show router bgp neighbor x.x.x.x received-routes|advertised-routes** commands. The BGP MH routes are only shown when the operator uses the **l2-vpn** family modifier. Should the operator want to filter only the BGP MH routes out of the l2-vpn routes, the **multi-homing** filter has to be added to the **show router bgp routes** commands.

```
*A:PE-3# show router bgp routes l2-vpn
===============================================================================
 BGP Router ID:192.0.2.3        AS:65000       Local AS:65000
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP L2VPN Routes
===============================================================================
Flag  RouteType                   Prefix                             MED
      RD                          SiteId                             Label
      Nexthop                     VeId                   BlockSize   LocalPref
      As-Path                     BaseOffset             vplsLabelBa
                                                         se
-------------------------------------------------------------------------------
u*>i  VPLS                        -                      -           0
      65000:501                   -                                  -
      192.0.2.1                   501                    8           100
      No As-Path                  497                    524271
u*>i  MultiHome                   -                      -           0
      65000:501                   2                                  -
      192.0.2.1                   -                      -           100
      No As-Path                  -                      -
u*>i  VPLS                        -                      -           0
      65000:502                   -                                  -
      192.0.2.2                   502                    8           100
      No As-Path                  497                    524271
u*>i  MultiHome                   -                      -           0
      65000:502                   2                                  -
      192.0.2.2                   -                      -           100
      No As-Path                  -                      -
-------------------------------------------------------------------------------
Routes : 4
===============================================================================
```

For the L2 VPN BGP routes toward site 2 (PE-1 and PE-2) in detail:

```
*A:PE-3# show router bgp routes l2-vpn multi-homing siteid 2 detail
===============================================================================
 BGP Router ID:192.0.2.3        AS:65000       Local AS:65000
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP L2VPN-MULTIHOME Routes
===============================================================================
Original Attributes

Route Type     : MultiHome
Route Dist.    : 65000:501
Site Id        : 2
Nexthop        : 192.0.2.1
```

```
From           : 192.0.2.1
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None                   IGP Cost       : n/a
Connector      : None
Community      : target:65000:500
                 l2-vpn/vrf-imp:Encap=19: Flags=-DF: MTU=0: PREF=0
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.1
Flags          : Used Valid Best IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : n/a
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h05m40s

Modified Attributes
 ---snip---

-------------------------------------------------------------------------------Original
 Attributes

Route Type     : MultiHome
Route Dist.    : 65000:502
Site Id        : 2
Nexthop        : 192.0.2.2
From           : 192.0.2.2
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None                   IGP Cost       : n/a
Connector      : None
Community      : target:65000:500
                 l2-vpn/vrf-imp:Encap=19: Flags=none: MTU=0: PREF=0
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.2
Flags          : Used Valid Best IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : n/a
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h05m40s

Modified Attributes
 ---snip---

-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

The following shows the Layer 2 BGP routes on PE-1:

```
*A:PE-1# show service l2-route-table ?
```

```
  - l2-route-table [detail] [bgp-ad] [multi-homing] [bgp-vpls] [bgp-vpws] [all-routes]

 <detail>             : keyword - display detailed information
```

```
*A:PE-1# show service l2-route-table multi-homing

===============================================================================
Services: L2 Multi-Homing Route Information - Summary
===============================================================================
Svc Id     L2-Routes (RD-Prefix)        Next Hop        SiteId     State  DF
-------------------------------------------------------------------------------
500        65000:502                    192.0.2.2       2          up(0)  clear
-------------------------------------------------------------------------------
No. of L2 Multi-Homing Route Entries: 1
===============================================================================
```

In case PE-3 were the RR for MTU-4 and MTU-5 as well as for PE-1 and PE-2, PE-1 would have two more L2-routes for multi-homing in this table, as follows:

```
*A:PE-1# show service l2-route-table multi-homing

===============================================================================
Services: L2 Multi-Homing Route Information - Summary
===============================================================================
Svc Id     L2-Routes (RD-Prefix)        Next Hop        SiteId     State  DF
-------------------------------------------------------------------------------
500        65000:504                    192.0.2.4       1          up(0)  set
500        65000:505                    192.0.2.5       1          up(0)  clear
500        65000:502                    192.0.2.2       2          up(0)  clear
-------------------------------------------------------------------------------
No. of L2 Multi-Homing Route Entries: 3
===============================================================================
```

When operational groups are configured (as previously shown), the following **show** command helps to find the operational dependencies between monitoring objects and group objects.

```
*A:PE-1# show service oper-group "group-1" detail

===============================================================================
Service Oper Group Information
===============================================================================
Oper Group       : group-1
Creation Origin  : manual                          Oper Status: up
Hold DownTime    : 0 secs                          Hold UpTime: 4 secs
Members          : 2                               Monitoring : 1
===============================================================================

===================================================================
Member SDP-Binds for OperGroup: group-1
===================================================================
SdpId            SvcId      Type     IP address      Adm     Opr
-------------------------------------------------------------------
12:4294967292    500        BgpVpls  192.0.2.2       Up      Up
13:4294967293    500        BgpVpls  192.0.2.3       Up      Up
-------------------------------------------------------------------
SDP Entries found: 2
===================================================================

===============================================================================
Monitoring Sites for OperGroup: group-1
===============================================================================
SvcId     Site                Site-Id   Dest                Admin   Oper  Fwdr
-------------------------------------------------------------------------------
500       MH-site-2           2         sdp:15:500          Enabled up    Yes
-------------------------------------------------------------------------------
Site Entries found: 1
===============================================================================
```

For debugging, the following CLI sources can be used:

- **log-id 99** — Provides information about the site object changes and DF changes.
- **debug router bgp update** command — Shows the BGP updates for BGP MH, including the sent and received BGP MH NLRIs and flags.

```
# on MTU-4:
debug
    router
        bgp
            update
```

- **debug router ldp** command — Provides information about the pseudowire status bits being signaled as well as the MAC flush messages.

```
# on MTU-4:
debug
    router
        ldp
            peer 192.0.2.1
                packet
                    init detail
                    label detail
```

As an example, log-id 99 shows the following debug output after disabling MH-site-1 on MTU-4:

```
# on MTU-4:
configure
    service
        vpls "VPLS 500"
            sap 1/1/1:7
                shutdown
            exit
            sap 1/1/2:8
                shutdown
            exit
```

```
*A:MTU-4# show log log-id 99
===============================================================================
Event Log 99
===============================================================================
---snip---

122 2021/01/19 09:38:17.885 UTC WARNING: SVCMGR #2531 Base BGP-MH
"Service-id 500 site MH-site-1 is not the designated-forwarder"

121 2021/01/19 09:38:17.884 UTC MINOR: SVCMGR #2203 Base
"Status of SAP 1/1/2:8 in service 500 (customer 1) changed to admin=down oper=down flags=Sap
AdminDown MhStandby"

---snip---
```

Log 2 has been configured to log BGP updates and LDP commands.

```
*A:MTU-4# show log log-id 2
===============================================================================
Event Log 2
===============================================================================
---snip---

4 2021/01/19 09:38:17.893 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 86
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family L2VPN
        NextHop len 4 NextHop 192.0.2.5
        [MH] site-id: 1, RD 65000:505
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.5
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        1.1.1.1
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:65000:500
        l2-vpn/vrf-imp:Encap=19: Flags=-DF: MTU=0: PREF=0
"

2 2021/01/19 09:38:17.885 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 72
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family L2VPN
        NextHop len 4 NextHop 192.0.2.4
        [MH] site-id: 1, RD 65000:504
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:65000:500
        l2-vpn/vrf-imp:Encap=19: Flags=D: MTU=0: PREF=0
"

1 2021/01/19 09:38:17.885 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Address Withdraw packet (msgId 348) to 192.0.2.1:0
Protocol version = 1
MAC Flush (All MACs learned from me)
Service FEC PWE3: ENET(5)/500 Group ID = 0 cBit = 0
"
```

Assuming all the recommended tools are enabled, a DF to non-DF transition can be shown as well as the corresponding MAC flush messages and related BGP processing.

If MH-site-2 is torn down on PE-1, the **debug router bgp update** command would allow us to see two BGP updates from PE-1:

- A BGP MH update for site-id 2 with flag D set (because the site is down).

- A BGP VPLS update for veid=501 and flag D set. This is due to the fact that there are no more active objects on the VPLS, besides the BGP pseudowires.

```
# on PE-1:
configure
    service
        vpls "VPLS 500"
            spoke-sdp 14:500
                shutdown
            exit
            spoke-sdp 15:500
                shutdown
            exit
```

```
*A:PE-1# show log log-id 2

===============================================================================
Event Log 2
===============================================================================
---snip---

5 2021/01/19 09:42:39.897 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 72
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family L2VPN
        NextHop len 4 NextHop 192.0.2.1
        [VPLS/VPWS] preflen 17, veid: 501, vbo: 497, vbs: 8, label-base: 524271,
                    RD 65000:501
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:65000:500
        l2-vpn/vrf-imp:Encap=19: Flags=D: MTU=1514: PREF=0
"

4 2021/01/19 09:42:39.897 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 72
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family L2VPN
        NextHop len 4 NextHop 192.0.2.1
        [MH] site-id: 2, RD 65000:501
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:65000:500
        l2-vpn/vrf-imp:Encap=19: Flags=D: MTU=0: PREF=0
"
```

The D flag, sent along with the BGP VPLS update for veid 501, would be seen on the remote core PEs as though it was a pseudowire status fault (although there is no TLDP running in the core).

```
*A:PE-2# show service id 500 all | match Flag
```

```
Flags                : PWPeerFaultStatusBits
Flags                : None
Flags                : None
Flags                : None
```

Conclusion

SR OS supports a wide range of service resiliency options as well as the best-of-breed system level HA and MPLS mechanisms for the access and the core. BGP MH for VPLS completes the service resiliency tool set by adding a mechanism that has some good advantages over the alternative solutions:

- BGP MH provides a common resiliency mechanism for attachment circuits (SAPs), pseudowires (spoke SDPs), split horizon groups and mesh bindings
- BGP MH is a network-based technique which does not need interaction to the CE or MTU to which it is providing redundancy to.

The examples used in this chapter illustrate the configuration of BGP MH for access CEs and MTUs. Show and debug commands have also been suggested so that the operator can verify and troubleshoot the BGP MH procedures.

BGP Virtual Private Wire Services

This chapter describes BGP Virtual Private Wire Service (VPWS) configurations.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter is applicable to SR OS and was initially written for SR OS Release 11.0.R4. The CLI in the current edition is based on SR OS Release 21.2.R1. There are no prerequisites for this configuration.

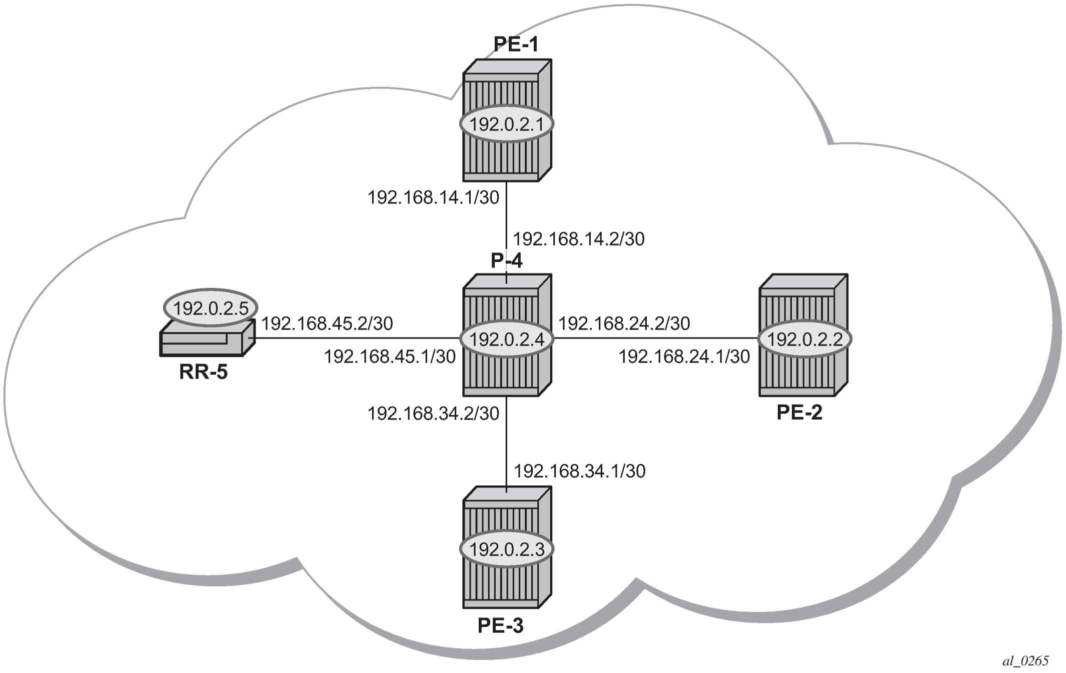
Overview

The following two IETF standards describe the provisioning of Virtual Private Wire Services (VPWS):

- RFC 4447, *Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)*, describes Label Distribution Protocol (LDP) VPWS, where VPWS pseudowires are signaled using LDP between Provider Edge (PE) Routers.
- RFC 6624, *Layer 2 Virtual Private Networks Using BGP for Auto-Discovery and Signaling*, describes the use of Border Gateway Protocol (BGP) for signaling of pseudowires between such PEs.

Figure 25: Example topology shows the example topology with five SR OS routers located in the same Autonomous System (AS). There are three PE routers connected to a single P router and a route reflector (RR) for the AS. The PE routers are all BGP VPWS-aware. The Provider (P) router is BGP VPWS-unaware and does not take part in the BGP process.

Figure 25: Example topology



The following configuration tasks are completed as a prerequisite:

- IS-IS or OSPF is configured on each of the network interfaces between the PE/P routers and route reflector.
- MPLS is configured on all interfaces between PE routers and P routers. It is not required between P-4 and RR-5.
- LDP is configured on interfaces between PE and P routers. It is not required between P-4 and the RR-5.
- RSVP is configured on interfaces between PE and P routers. It is not required between P-4 and the RR-5.

BGP VPWS

In this architecture, a VPWS is a collection of two (or three in case of redundancy) BGP VPWS service instances present on different PEs in a provider network.

The PEs communicate with each other at the control plane level by means of BGP updates containing BGP VPWS Network Layer Reachability Information (NLRI). Each update contains enough information for a PE to determine the presence of other BGP VPWS instances on peering PEs and to set up pseudowire connectivity for data flow between peers containing the same BGP VPWS service. Therefore, auto-discovery and pseudowire signaling is achieved using a single BGP update message.

Each PE with a BGP VPWS instance is identified by a VPWS edge identifier (VE-ID) and the presence of other BGP VPWS instances is determined using the exchange of standard BGP extended community route targets (RTs) between PEs.

Each PE will advertise, via the RR, the presence of its BGP VPWS instance to all other PEs, along with a block of multiplexer labels (for BGP VPWS, one label per block) that can be used to communicate between each instance, plus a BGP next-hop that determines a labeled transport tunnel to be used between PEs.

Each BGP VPWS instance is configured with import and export route target extended communities for topology control, along with VE identification.

Configuration

The following examples show the configuration of four BGP VPWS scenarios:

- Single homed BGP VPWS
  - using auto-provisioned SDPs
  - using pre-provisioned SDPs
- Dual homed BGP VPWS
  - with single pseudowire
  - with active/standby pseudowire

Configure MP-iBGP

The first step is to configure an MP-iBGP session between each of the PEs and the RR. The configuration for all PEs is as follows:

```
# on PE-1, PE-2, and PE-3:
configure
    router Base
        autonomous-system 65536
        bgp
            group "INTERNAL"
                family l2-vpn
                peer-as 65536
                neighbor 192.0.2.5
                exit
            exit
        exit
```

The IP addresses can be derived from Figure 25: Example topology.

On RR-5, the BGP configuration is as follows:

```
# on RR-5:
configure
    router Base
        autonomous-system 65536
        bgp
            group "INTERNAL"
                family l2-vpn
                peer-as 65536
                cluster 1.1.1.1
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
        exit
```

The following command on RR-5 shows that BGP sessions with each PE are established and have a negotiated L2 VPN address family capability.

```
*A:RR-5# show router bgp summary all

===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
Description
ServiceId          AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
192.0.2.1
Def. Instance  65536        5    0 00h01m24s 0/0/0 (L2VPN)
                            5    0
192.0.2.2
Def. Instance  65536        5    0 00h01m24s 0/0/0 (L2VPN)
                            5    0
192.0.2.3
Def. Instance  65536        5    0 00h01m24s 0/0/0 (L2VPN)
                            5    0

-------------------------------------------------------------------------------
```

Pseudowire templates

BGP VPWS utilizes pseudowire (PW) templates to dynamically instantiate SDP bindings for a service to signal the egress service de-multiplexer labels used by remote PEs to reach the local PE.

The template determines the signaling parameters of the pseudowire, such as vc-type, vlan-vc-tag, hash-label, filters, and so on. The following parameters are recognized by BGP VPWS; the remainder is ignored.

```
configure
  service
  - pw-template <policy-id> [create] [prefer-provisioned-sdp] [name <name>]
                                                              [auto-gre-sdp]
  - no pw-template <policy-id>
  - pw-template <policy-id> use-provisioned-sdp [create] [name <name>]
       accounting-policy acct-policy-id
       no accounting-policy
       [no] collect-stats
       [no] controlword
       egress
         filter ipv6 <filter-id>
         filter ip <filter-id>
         filter mac <filter-id>
         no filter [ip <filter-id>] [mac <filter-id>] [ipv6 <filter-id>]
         [no] filter-name [ip name] [mac name] [ipv6 name]
         no qos [<network-policy-id>]
         qos <network-policy-id> port-redirect-group <queue-group-name>
                                                     instance <instance-id>
         qos name <network-policy-name> port-redirect-group <queue-group-name>
```

```
                                                     instance <instance-id>
       [no] force-vlan-vc-forwarding
       hash-label [signal-capability]
       no hash-label
       entropy-label
       ingress
         filter ipv6 <filter-id>
         filter ip <filter-id>
         filter mac <filter-id>
         no filter [ip <filter-id>] [mac <filter-id>] [ipv6 <filter-id>]
         [no] filter-name [ip name] [mac name] [ipv6 name]
         qos <network-policy-id> fp-redirect-group <queue-group-name>
                                                   instance <instance-id>
         qos name <network-policy-name> fp-redirect-group <queue-group-name>
                                                   instance <instance-id>
         no qos [<network-policy-id>]
       [no] sdp-exclude group-name
       [no] sdp-include group-name
       vc-type {ether | vlan}
       vlan-vc-tag 0..4094
       [no] vlan-vc-tag
```

Note:

- The encapsulation type in the Layer-2 extended community is either 4 (Ethernet VLAN tagged mode) or 5 (Ethernet raw mode), depending on the **vc-type** parameter.
- The **force-vlan-vc-forwarding** function will add a tag (equivalent to vc-type vlan) and will allow for customer QoS transparency (dot1p + Drop Eligibility (DE) bits).

The MPLS transport tunnel between PEs can be signaled using LDP or RSVP-TE.

LDP-based SDPs can be automatically instantiated or pre-provisioned. RSVP-TE-based SDPs have to be pre-provisioned. If pre-provisioned pseudowires are used, the PW template must be created with the **use-provisioned-sdp** parameter. Alternatively, the **prefer-provisioned-sdp** parameter can be used, in which case a pre-provisioned SDP will be used if available; if not, LDP-based SDPs can be automatically instantiated, see chapter LDP VPLS Using BGP Auto-Discovery — Prefer Provisioned SDP.

Pseudowire templates for auto-SDP creation using LDP

In order to use an LDP transport tunnel for data flow between PEs, link layer LDP needs to be configured between all PEs/Ps so that a transport label for each PE system interface is available. For example, on PE-1:

```
# on PE-1:
configure
    router Base
        ldp
            interface-parameters
                interface "int-PE-1-P-4" dual-stack
                    ipv4
                        no shutdown
                    exit
```

Using this mechanism, SDPs can be auto-instantiated with SDP-IDs starting at the higher end of the SDP numbering range, such as 32767. Any subsequent SDPs created use SDP-IDs decrementing from this value.

A pseudowire template is required. The following example is created using the default values:

```
# on PE-1, PE-2, and PE-3:
configure
    service
        pw-template 3 name "PW3" create
        exit
```

Pseudowire templates for provisioned SDPs using RSVP-TE

RSVP-TE LSPs need to be created between the PE routers on which provisioned SDPs will be used as prerequisite.

The MPLS interface and LSP configuration for PE-1 are:

```
# on PE-1:
configure
    router Base
        mpls
            interface "int-PE-1-P-4‟
            exit
            path "dyn"
                no shutdown
            exit
            lsp "LSP-PE-1-PE-2"
                to 192.0.2.2
                primary "dyn"
                exit
                no shutdown
            exit
            lsp "LSP-PE-1-PE-3"
                to 192.0.2.3
                primary "dyn"
                exit
                no shutdown
            exit
            no shutdown
```

The MPLS and LSP configuration for PE-2 are similar to that of PE-1 with the appropriate interfaces and LSP names configured.

To use an RSVP-TE tunnel as transport between PEs, it is necessary to bind the RSVP-TE LSP between PEs to an SDP.

On PE-1, the SDP toward PE-2 is configured as follows. Similar SDPs are required on each PE to the remote PEs in the service where provisioned SDPs are to be used.

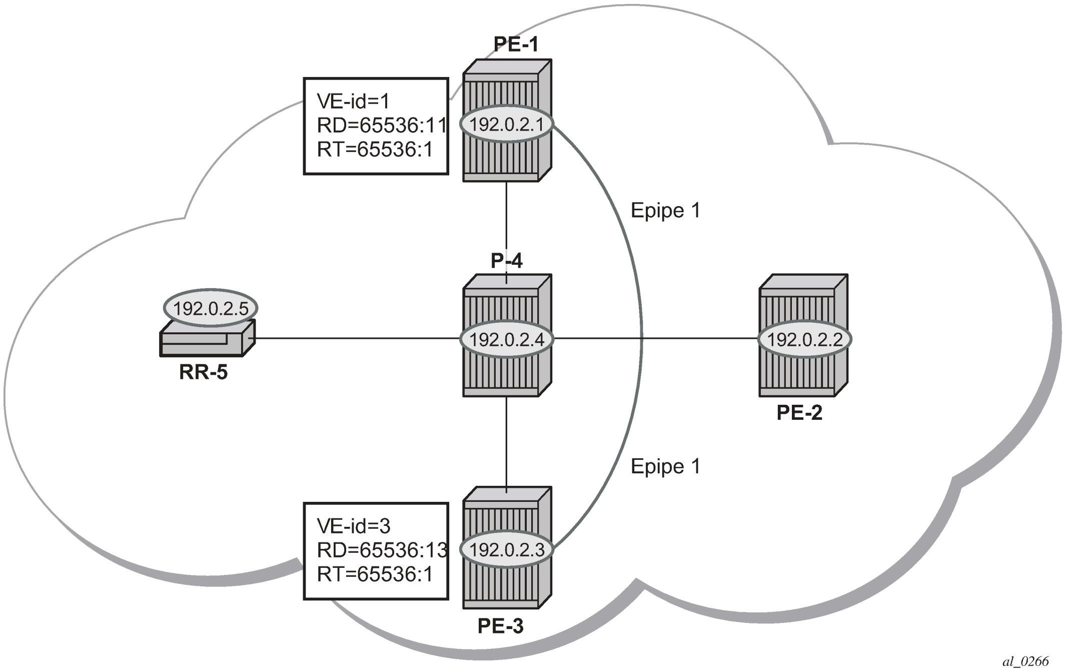
```
# on PE-1:
configure
    service
        sdp 12 mpls create
            description "SDP-PE-1-PE-2_RSVP_BGP"
            signaling bgp
            far-end 192.0.2.2
            lsp "LSP-PE-1-PE-2‟
            no shutdown
        exit
```

The **signaling bgp** parameter is required. BGP VPWS instances using BGP VPWS signaling can use BGP-signaled SDPs. However, TLDP-signaled (default) SDPs that are bound to RSVP-based LSPs will not be used as SDPs within BGP VPWS.

Single-homed BGP VPWS using auto-provisioned SDPs

Figure 26: Single-homed BGP VPWS using auto-provisioned SDPs shows a schematic of a single homed BGP VPWS between PE-1 and PE-3 where SDPs are auto-provisioned. In this case, the transport tunnels are LDP-signaled.

Figure 26: Single-homed BGP VPWS using auto-provisioned SDPs



The following shows the configuration required on PE-1 for a BGP VPWS service using a pseudowire template configured for auto-provisioning of SDPs.

```
# on PE-1:
configure
    service
        pw-template 1 name "PW1" create
            vc-type vlan
        exit
        epipe 1 name "Epipe1" customer 1 create
            bgp
                route-distinguisher 65536:11
                route-target export target:65536:1 import target:65536:1
                pw-template-binding 1
                exit
            exit
            bgp-vpws
                ve-name "PE-1"
                    ve-id 1
                exit
                remote-ve-name "PE-3"
                    ve-id 3
                exit
                no shutdown
            exit
            sap 1/1/4:1 create
            exit
            no shutdown
        exit
```

The **bgp** context specifies parameters that are required for BGP VPWS.

Within the **bgp** context, parameters are configured that are used by the neighboring PEs to determine the membership of a BGP VPWS, in other words, the auto-discovery of PEs in the same BGP VPWS. Within the **bgp** context, the RD is configured, along with the route target extended communities. Route target communities are used to determine membership of a BGP VPWS. The import and export route targets at the BGP level are mandatory. The PW template binding is then applied and its parameters are used for both the routes sent by this PE and the received routes matching the route target value.

Within the **bgp-vpws** context, the signaling parameters are configured. These determine the service labels required for the data plane of the VPWS instance.

The VPWS Edge ID (VE-ID) is a numerical value assigned to each PE within a BGP VPWS. This value must be unique for a BGP VPWS, with the exception of multi-homed scenarios, where two dual-homed PEs can have the same VE-ID and are distinguishable by the site preference (or by the tie breaking rules from the *draft-ietf-bess-vpls-multihoming-03*).

Changes to the pseudowire template are not taken into account once the pseudowire has been set up (changes of RT are refreshed though). PW-templates can be re-evaluated with the **tools perform service eval-pw-template** command. The **eval-pw-template** checks if all of the bindings using this PW template policy are still meant to be using this policy. If the template has changed and **allow-service-impact** is true, then the old binding is removed and it is re-added using the new template.

```
*A:PE-1# tools perform service eval-pw-template 1
eval-pw-template succeeded for Svc 1 Tx L2 ExtComm, Policy 1
eval-pw-template succeeded for Svc 1 32767:4294967295 Policy 1
```

VE-ID and BGP label allocations

For a point-to-point VPWS, there are only two members within the BGP VPWS service, so only one label entry is required by each remote service. For dual-homed scenarios, there are two labels for the redundant site, one from each dual-homed PE.

Each PE allocates a label per BGP VPWS instance for the remote PEs, so it signals blocks with one label. It achieves this by advertising three parameters in a BGP update message. For more information about these parameters, see chapter BGP VPLS.

- A Label Base (LB) which is the lowest label in the block.
- A VE Block size (VBS) which is always 1 and cannot be changed.
- A VE Base Offset (VBO) corresponding to the first label in the label block.

PE-3 service creation

On PE-3, Epipe 1 is configured using PW template 1, as follows. PE-3 has been allocated a VE-ID of 3. For completeness, the PW template is also shown.

```
# on PE-3:
configure
    service
        pw-template 1 name "PW1" create
            vc-type vlan
        exit
        epipe 1 name "Epipe1" customer 1 create
            bgp
                route-distinguisher 65536:13
                route-target export target:65536:1 import target:65536:1
                pw-template-binding 1
                exit
            exit
            bgp-vpws
                ve-name "PE-3"
                    ve-id 3
                exit
                remote-ve-name "PE-1"
                    ve-id 1
                exit
                no shutdown
            exit
            sap 1/1/4:1 create
            exit
            no shutdown
        exit
```

PE-1 service operation verification

The following command shows that the BGP VPWS service is enabled on PE-1:

```
*A:PE-1# show service id 1 bgp-vpws

===============================================================================
BGP VPWS Information
===============================================================================
Admin State          : Enabled
VE Name              : PE-1                 VE Id            : 1
PW Tmpl used         : 1

Remote-Ve Information
-------------------------------------------------------------------------------
Remote VE Name       : PE-3                 Remote VE Id     : 3
===============================================================================
```

The following shows the BGP information used by the BGP VPWS service on PE-1:

```
*A:PE-1# show service id 1 bgp

===============================================================================
BGP Information
===============================================================================
Vsi-Import           : None
```

```
Vsi-Export           : None
Route Dist           : 65536:11
Oper Route Dist      : 65536:11
Oper RD Type         : configured
Rte-Target Import    : 65536:1              Rte-Target Export: 65536:1
Oper RT Imp Origin   : configured           Oper RT Import   : 65536:1
Oper RT Exp Origin   : configured           Oper RT Export   : 65536:1

PW-Template Id       : 1
Endpoint             : <none>
BFD Template         : None
BFD-Enabled          : no                   BFD-Encap        : ipv4
Import Rte-Tgt       : None
-------------------------------------------------------------------------------
===============================================================================
```

Epipe 1 is operationally up on PE-1, as follows:

```
*A:PE-1# show service id 1 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : Epipe
MACSec enabled    : no
Name              : Epipe1
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 03/04/2021 15:25:11
Last Mgmt Change  : 03/04/2021 15:25:11
Test Service      : No
Admin State       : Up                  Oper State        : Up
MTU               : 1514
Vc Switching      : False
SAP Count         : 1                   SDP Bind Count    : 1
Per Svc Hashing   : Disabled
Vxlan Src Tep Ip  : N/A
Force QTag Fwd    : Disabled
Oper Group        : <none>

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:1                              q-tag        1578    1578    Up   Up
sdp:32767:4294967295 SB(192.0.2.3)       BgpVpws      0       1552    Up   Up
===============================================================================
```

The SAP and SDP are all operationally up. The indication *SB* next to the SDP-ID signifies “Spoke” and “BGP”.

The following output shows the ingress and egress labels for PE-1.

```
*A:PE-1# show service id 1 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId             Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
32767:4294967295 BgpVpws  192.0.2.3       Up      Up        524281    524281
```

```
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

The following debug output from PE-1 shows the BGP VPWS NLRI update for Epipe 1 sent by PE-1 to RR-5. This update will then be received by the other PEs.

```
# on PE-1:
debug
    router "Base"
        bgp
            update
```

```
3 2021/03/04 15:25:41.024 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 76
    Flag: 0x90 Type: 14 Len: 32 Multiprotocol Reachable NLRI:
        Address Family L2VPN
        NextHop len 4 NextHop 192.0.2.1
        [VPLS/VPWS] preflen 21, veid: 1, vbo: 3, vbs: 1, label-base: 524281,
                    RD 65536:11, csv: 0x00000000, type 1, len 1,
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:65536:1
        l2-vpn/vrf-imp:Encap=4: Flags=none: MTU=1514: PREF=0
"
```

The control flags within the extended community indicate the status of the BGP VPWS instance.

The control flags are the following:

```
0 1 2 3 4 5 6 7
+-+-+-+-+-+-+-+-+
|D|A|F|Z|Z|Z|C|S| (Z = MUST Be Zero)
+-+-+-+-+-+-+-+-+
```

- D: access circuit down indicator. D is 1 if all access circuits are down, otherwise D is 0.
- A: automatic site ID allocation, which is not supported. This is ignored on receipt and set to 0 on sending.
- F: MAC flush indicator, this relates to VPLS. This is set to 0 and ignored on receipt.
- C: presence of a control word. Control word usage is not supported. This is set to 0 on sending (control word not present) and if a non-zero value is received (indicating a control word is required), the pseudowire will not be created.
- S: sequenced delivery. Sequenced delivery is not supported. This is set to 0 on sending (no sequenced delivery) and if a non-zero value is received (indicating sequenced delivery required), the pseudowire will not be created.

The BGP VPWS NLRI is based on the BGP VPLS NLRI, but is extended with a Circuit Status Vector (CSV). The circuit status vector is used to indicate the status of both the SAP and the spoke-SDP within the local service. Because the VE block size used is 1, the most significant bit in the circuit status vector TLV value will be set to 1 if either the SAP or spoke-SDP is down; otherwise, it will be set to 0.

```
# on PE-1:
configure
    service
        epipe "Epipe1"
            sap 1/1/4:1
                shutdown
```

```
6 2021/03/04 15:31:59.024 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 76
    Flag: 0x90 Type: 14 Len: 32 Multiprotocol Reachable NLRI:
        Address Family L2VPN
        NextHop len 4 NextHop 192.0.2.1
        [VPLS/VPWS] preflen 21, veid: 1, vbo: 3, vbs: 1, label-base: 524281,
                    RD 65536:11, csv: 0x00000080, type 1, len 1,
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:65536:1
        l2-vpn/vrf-imp:Encap=4: Flags=D: MTU=1514: PREF=0
"
```

After disabling the local SAP, the CSV has the most significant bit set to 1 (0x80). The following command shows the BGP VPWS update received on PE-3:

```
*A:PE-3# show service l2-route-table bgp-vpws detail

===============================================================================
Services: L2 Bgp-Vpws Route Information - Summary
===============================================================================

Svc Id         : 1
VeId           : 1
PW Temp Id     : 1
RD             : *65536:11
Next Hop       : 192.0.2.1
State (D-Bit)  : down(1)
Path MTU       : 1514
Control Word   : 0
Seq Delivery   : 0
Status         : active
Tx Status      : active
CSV            : 80
Preference     : 0
Sdp Bind Id    : 32767:4294967295
===============================================================================
```

On PE-1, SAP 1/1/4:1 is re-enabled as follows:

```
# on PE-1:
configure
    service
        epipe "Epipe1"
            sap 1/1/4:1
```

```
                no shutdown
            exit
```

PE-3 service operation verification

Similar to PE-1, the service operation should be validated on PE-3.

Single-homed BGP VPWS using pre-provisioned SDP

It is possible to configure BGP VPWS instances that use RSVP-TE transport tunnels. In this case, the SDPs must be created with the MPLS LSPs mapped and with the signaling set to BGP, because the service labels are signaled using BGP. The PW template configured within the BGP VPWS instance must use the keyword **use-provisioned-sdp** (or **prefer-provisioned-sdp**).

Figure 27: Single-homed BGP VPWS using pre-provisioned SDP

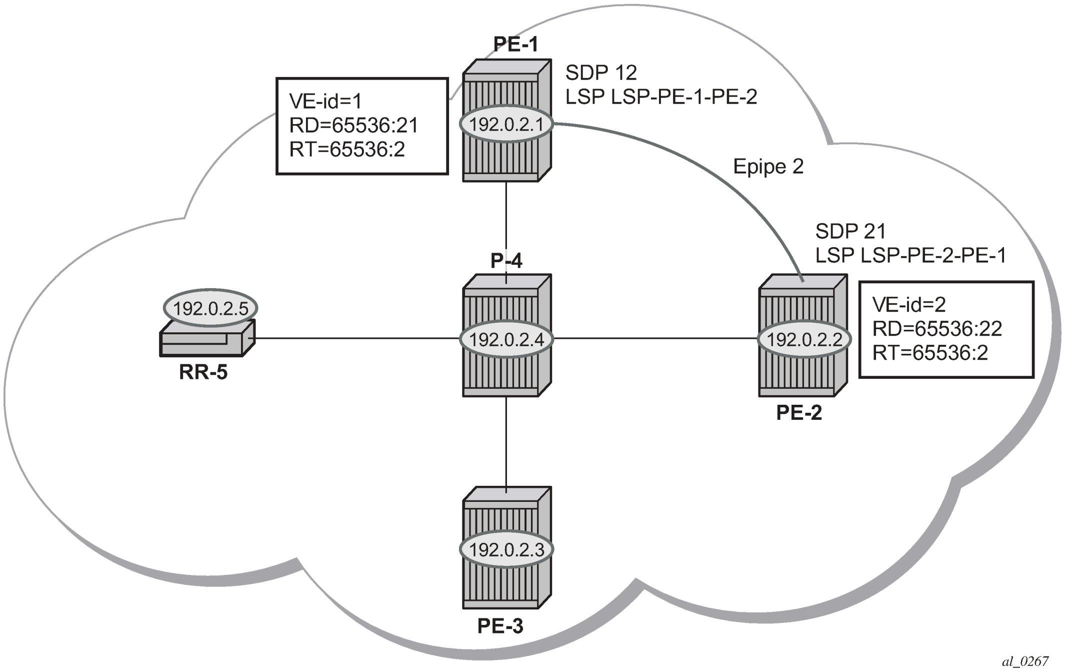


Figure 27: Single-homed BGP VPWS using pre-provisioned SDP shows a schematic of a BGP VPWS where SDPs are pre-provisioned with RSVP-TE signaled transport tunnels.

On PE-1, SDP 12 toward PE-2 is configured as follows:

```
# on PE-1:
configure
    service
        sdp 12 mpls create
            description "SDP-PE-1-PE-2_RSVP_BGP"
            signaling bgp
            far-end 192.0.2.2
            lsp "LSP-PE-1-PE-2"
            no shutdown
        exit
```

On PE-2, SDP 21 toward PE-1 is configured as follows:

```
# on PE-2:
configure
    service
        sdp 21 mpls create
            description "SDP-PE-2-PE-1_RSVP_BGP"
            signaling bgp
            far-end 192.0.2.1
            lsp "LSP-PE-2-PE-1"
            no shutdown
        exit
```

To create a spoke SDP within a service that uses the RSVP-TE transport tunnel, a pseudowire template is required that has the **use-provisioned-sdp** parameter set.

The PW template is provisioned on both PEs as follows:

```
# on PE-1 and PE-2:
configure
    service
        pw-template 2 name "PW2" use-provisioned-sdp create
        exit
```

The following output shows the configuration required for a BGP VPWS service using a PW template configured for using pre-provisioned RSVP-TE SDPs.

```
# on PE-1:
configure
    service
        epipe 2 name "Epipe2" customer 1 create
            bgp
                route-distinguisher 65536:21
                route-target export target:65536:2 import target:65536:2
                pw-template-binding 2
                exit
            exit
            bgp-vpws
                ve-name "PE-1"
                    ve-id 1
                exit
                remote-ve-name "PE-2"
                    ve-id 2
                exit
                no shutdown
            exit
            sap 1/1/4:2 create
            exit
            no shutdown
```

The route distinguisher and route target extended community values for Epipe 2 are different from that in Epipe 1. This is to differentiate between the two as their visibility is global within the BGP domain. The VE-ID values can be reused in each Epipe instance, as long as they are unique within the instance.

Similarly, the configuration is as follows on PE-2, where the VE-ID is 2:

```
# on PE-2:
```

```
configure
    service
         epipe 2 name "Epipe2" customer 1 create
            bgp
                route-distinguisher 65536:22
                route-target export target:65536:2 import target:65536:2
                pw-template-binding 2
                exit
            exit
            bgp-vpws
                ve-name "PE-2"
                    ve-id 2
                exit
                remote-ve-name "PE-1"
                    ve-id 1
                exit
                no shutdown
            exit
            sap 1/1/4:2 create
            exit
            no shutdown
```

The service Epipe 2 is operationally up on PE-1, as follows:

```
*A:PE-1# show service id 2 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 2                   Vpn Id            : 0
Service Type      : Epipe
---snip---

Admin State       : Up                  Oper State        : Up
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:2                              q-tag        1578    1578    Up   Up
sdp:12:4294967294 S(192.0.2.2)           BgpVpws      0       1552    Up   Up
===============================================================================
```

The SDP-ID is the pre-provisioned SDP 12.

For completeness, the following command shows that the service is operationally up on PE-2.

```
*A:PE-2# show service id 2 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 2                   Vpn Id            : 0
Service Type      : Epipe
---snip---

Admin State       : Up                  Oper State        : Up
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
```

```
-------------------------------------------------------------------------------
Identifier                              Type        AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:2                             q-tag       1578    1578    Up   Up
sdp:21:4294967295 S(192.0.2.1)          BgpVpws     0       1552    Up   Up
===============================================================================
```

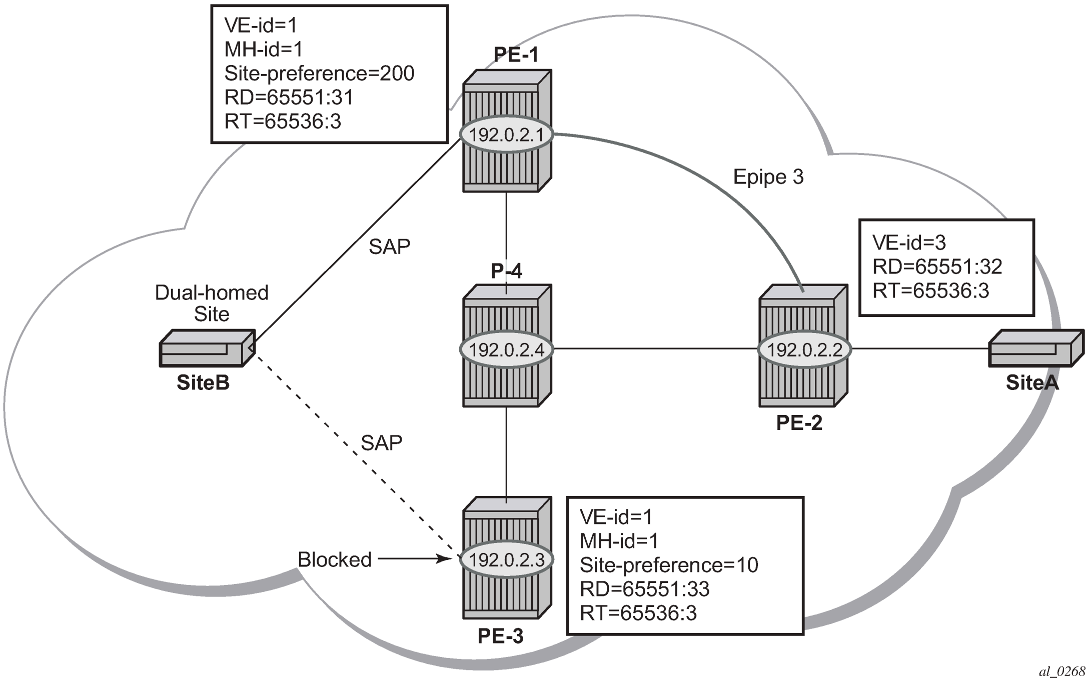
The SDP-ID used is the pre-provisioned SDP 21.

Dual-homed BGP VPWS with single pseudowire

For access redundancy, an Epipe using a BGP VPWS service can be configured as dual-homed, as described in *draft-ietf-bess-vpls-multihoming-03*. It can be configured with a single pseudowire setup, where the redundant pseudowire is not created until the initially active pseudowire is removed.

The following diagram shows a setup where an Epipe is configured on each PE. Site B is dual-homed to PE-1 and PE-3 with the remote PE-2 connected to site A; each site connection uses a SAP. A single pseudowire using Ethernet Raw Mode encapsulation connects PE-2 to PE-1 or PE-3 (but not both at the same time). The pseudowire is signaled using BGP VPWS over a tunnel LSP between the PEs.

Figure 28: Dual-homed BGP VPWS with single pseudowire



BGP multi-homing is configured for the dual-homed site B using a site-ID=1. The site-preference on PE-1 is set to 200 and to 10 on PE-3, this ensures that PE-1 will be the site’s Designated Forwarder (DF) and the pseudowire from PE-2 will be created to PE-1 when PE-1 is fully operational (no pseudowire is created on PE-2 to PE-3). If PE-1 fails, or the multi-homing site fails over to PE-3, then the pseudowire from PE-2 to PE-1 will be removed and a new pseudowire will be created from PE-2 to PE-3.

On PE-1, Epipe 3 is configured as follows:

```
# on PE-1:
configure
    service
         pw-template 3 name "PW3" create
         exit
         epipe 3 name "Epipe3" customer 1 create
            bgp
                route-distinguisher 65536:31
                route-target export target:65536:3 import target:65536:3
                pw-template-binding 3
                exit
            exit
            bgp-vpws
                ve-name "PE-1"
                    ve-id 1
                exit
                remote-ve-name "PE-2"
                    ve-id 2
                exit
                no shutdown
            exit
            site "SITEB" create
                site-id 1
                sap 1/1/4:3
                site-preference 200
                no shutdown
            exit
            sap 1/1/4:3 create
            exit
            no shutdown
        exit
```

Epipe 3 is configured on PE-3 with the same VE-ID as on PE-1, as follows:

```
# on PE-3:
configure
    service
        pw-template 3 name "PW3" create
        exit
        epipe 3 name "Epipe3" customer 1 create
            bgp
                route-distinguisher 65536:33
                route-target export target:65536:3 import target:65536:3
                pw-template-binding 3
                exit
            exit
            bgp-vpws
                ve-name "PE-3"
                    ve-id 1
                exit
                remote-ve-name "PE-2"
                    ve-id 2
                exit
                no shutdown
            exit
            site "SITEB" create
                site-id 1
                sap 1/1/4:3
                site-preference 10
                no shutdown
            exit
            sap 1/1/4:3 create
            exit
            no shutdown
        exit
```

In the preceding configurations, the **remote-ve-name** for PE-2 uses VE-ID 2 on both PE-1 and PE-3.

Epipe 3 is configured on PE-2 as follows:

```
# on PE-2:
configure
    service
        pw-template 3 name "PW3" create
        exit
        epipe 3 name "Epipe3" customer 1 create
            bgp
                route-distinguisher 65536:32
                route-target export target:65536:3 import target:65536:3
                pw-template-binding 3
                exit
            exit
            bgp-vpws
                ve-name "PE-2"
                    ve-id 2
                exit
                remote-ve-name "PE-1 or PE-3"
                    ve-id 1
                exit
                no shutdown
            exit
            sap 1/1/4:3 create
            exit
            no shutdown
        exit
```

On PE-2, the **remote-ve-name** is configured as “PE-1 or PE-3”; this is because both of these PEs are configured with VE-ID 1.

As a result of this configuration, there are multiple route entries for RD 65536:31 on PE-2. In the BGP routing table, there are two entries per partner PE, one for the BGP-MH update (with site-ID=1) and the other for the BGP-VPWS update (with VE-ID=1).

```
*A:PE-2# show router bgp routes l2-vpn rd 65536:31
===============================================================================
 BGP Router ID:192.0.2.2        AS:65536       Local AS:65536
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP L2VPN Routes
===============================================================================
Flag  RouteType                   Prefix                             MED
      RD                          SiteId                             Label
      Nexthop                     VeId                   BlockSize   LocalPref
      As-Path                     BaseOffset             vplsLabelBa
                                                         se
-------------------------------------------------------------------------------
u*>i  MultiHome                   -                      -           0
      65536:31                    1                                  -
      192.0.2.1                    -                      -           200
      No As-Path                   -                      -
u*>i  VPWS                         -                      -           0
      65536:31                     -                                  -
      192.0.2.1                    1                      1           200
      No As-Path                   2                      524279
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

```
*A:PE-2# show router bgp routes l2-vpn rd 65536:33
===============================================================================
 BGP Router ID:192.0.2.2        AS:65536       Local AS:65536
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP L2VPN Routes
===============================================================================
Flag  RouteType                    Prefix                             MED
      RD                           SiteId                             Label
      Nexthop                      VeId                   BlockSize   LocalPref
      As-Path                      BaseOffset             vplsLabelBa
                                                          se
-------------------------------------------------------------------------------
u*>i  MultiHome                    -                      -           0
      65536:33                     1                                  -
      192.0.2.3                    -                      -           10
      No As-Path                   -                      -
u*>i  VPWS                         -                      -           0
      65536:33                     -                                  -
      192.0.2.3                    1                      1           10
      No As-Path                   2                      524280
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

The route to PE-1 has the higher site preference, so it is selected as the target for the pseudowire.

```
*A:PE-2# show service l2-route-table bgp-vpws detail

===============================================================================
Services: L2 Bgp-Vpws Route Information - Summary
===============================================================================

---snip---

Svc Id         : 3
VeId           : 1
PW Temp Id     : 3
RD             : *65536:31
Next Hop       : 192.0.2.1
State (D-Bit)  : up(0)
Path MTU       : 1514
Control Word   : 0
Seq Delivery   : 0
Status         : active
Tx Status      : active
CSV            : 0
Preference     : 200
```

```
Sdp Bind Id    : 32767:4294967292
===============================================================================
```

After disabling the SAP in the service on PE-1, BGP update messages are received. The VPLS/VPWS message received on PE-2 from PE-1 shows in the CSV that the access circuit is down (the CSV has the most-significant bit set to 1 (0x80)), so PE-2 selects the update from PE-3 to create the pseudowire. The BGP-MH update received by PE-2 from PE-1 also shows that the local site is down as indicated by the flags=D.

Note in the following debug output:

- BGP MH (multi-homing) entry uses encap-type=19.
- BGP VPWS entry uses encap-type=5 (Ethernet raw mode).

```
# Disable SAP in Epipe 3 on PE-1:
configure
    service
        epipe "Epipe3"
            sap 1/1/4:3
                shutdown
```

```
34 2021/03/04 15:56:35.904 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 90
    Flag: 0x90 Type: 14 Len: 32 Multiprotocol Reachable NLRI:
        Address Family L2VPN
        NextHop len 4 NextHop 192.0.2.1
        [VPLS/VPWS] preflen 21, veid: 1, vbo: 2, vbs: 1, label-base: 524279,
                    RD 65536:31, csv: 0x00000080, type 1, len 1,
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 0
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.1
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        1.1.1.1
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:65536:3
        l2-vpn/vrf-imp:Encap=5: Flags=D: MTU=1514: PREF=200
"
```

```
35 2021/03/04 15:56:35.904 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 86
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family L2VPN
        NextHop len 4 NextHop 192.0.2.1
        [MH] site-id: 1, RD 65536:31
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 0
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.1
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        1.1.1.1
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:65536:3
        l2-vpn/vrf-imp:Encap=19: Flags=D: MTU=0: PREF=200
"
```

The result can be shown on PE-2 because the spoke SDP to PE-3 is now up (active).

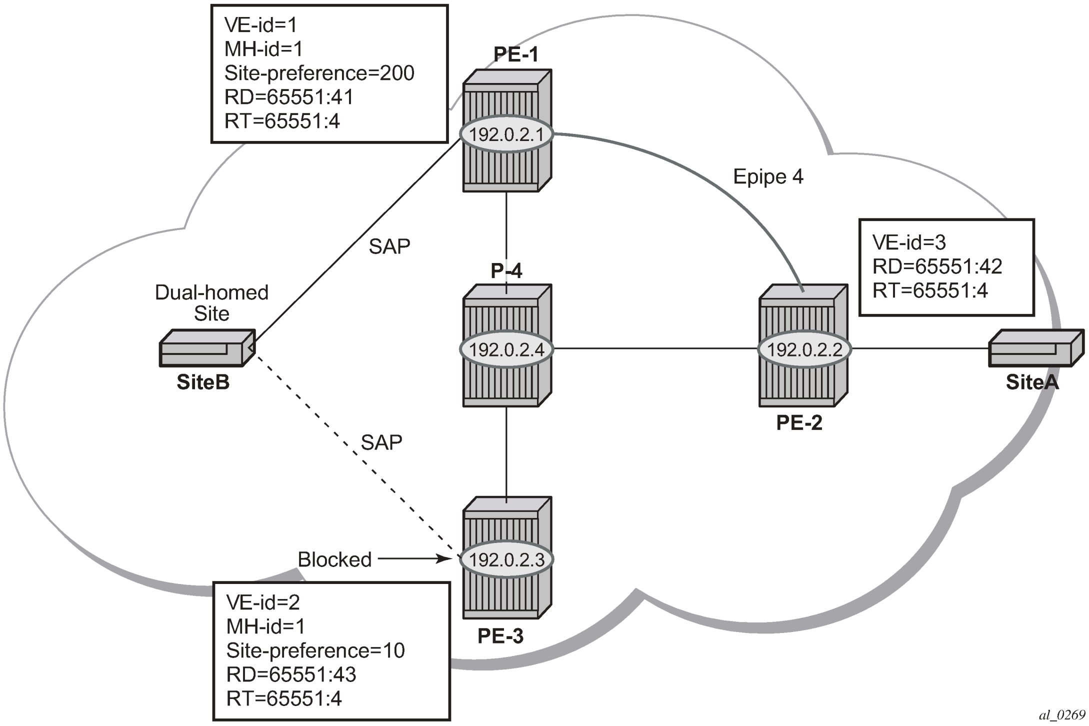
```
*A:PE-2# show service l2-route-table bgp-vpws detail

===============================================================================
Services: L2 Bgp-Vpws Route Information - Summary
===============================================================================

---snip---

Svc Id         : 3
VeId           : 1
PW Temp Id     : 3
RD             : *65536:33
Next Hop       : 192.0.2.3
State (D-Bit)  : up(0)
Path MTU       : 1514
Control Word   : 0
Seq Delivery   : 0
Status         : active
Tx Status      : active
CSV            : 0
Preference     : 10
Sdp Bind Id    : 32767:4294967291
===============================================================================
```

Dual-homed BGP VPWS with active/standby pseudowire

The second method for BGP VPWS pseudowire redundancy is an active/standby configuration. Whereas in the solution with one pseudowire, the redundant nodes use the same VE-ID for the remote PE and different preferences; in the active/standby solution, the redundant nodes use different VE-IDs for the remote PE and different preferences. The node connecting to both pseudowires (PE-2 in this example) has both remote VE-IDs configured. This allows for faster failover because the standby pseudowire is instantiated in addition to the active pseudowire. If more than two applicable BGP updates are received, at most one standby pseudowire is created (based on the BGP VPWS tie breaking rules).

Figure 29: Dual-homed BGP VPWS with active/standby pseudowire shows a setup where an Epipe is configured on each PE. Site B is dual-homed to PE-1 and PE-3 with the remote PE-2 connected to site A; each site connection uses a SAP. The active/standby pseudowires using Ethernet raw mode encapsulation connect PE-2 to PE-1 and PE-3. The pseudowires are signaled using BGP VPWS over tunnel LSPs between the PEs.

Figure 29: Dual-homed BGP VPWS with active/standby pseudowire



BGP Multi-Homing (MH) is configured for the dual-homed site B using a site-ID=1. The site preference on PE-1 is set to 200 and to 10 on PE-3; this ensures that PE-1 will be the site's DF for the MH site. The active pseudowire from PE-2 will be created to PE-1 with the standby pseudowire being created to PE-3. If PE-1 fails, or the multi-homing site fails over to PE-3, then the pseudowire from PE-2 to PE-3 will become active (used as the data path between site A and B).

Epipe 4 is configured on PE-1 as follows:

```
# on PE-1:
configure
    service
         pw-template 3 name "PW3" create
         exit
         epipe 4 name "Epipe4" customer 1 create
            bgp
                route-distinguisher 65536:41
                route-target export target:65536:4 import target:65536:4
                pw-template-binding 3
                exit
            exit
            bgp-vpws
                ve-name "PE-1"
                    ve-id 1
                exit
                remote-ve-name "PE-2"
                    ve-id 2
                exit
                no shutdown
            exit
            site "SITEB" create
                site-id 1
                sap 1/1/4:4
                site-preference 200
                no shutdown
            exit
            sap 1/1/4:4 create
            exit
            no shutdown
        exit
```

Epipe 4 is configured on PE-3 with local VE-ID is 3 (different from previous example), as follows:

```
# on PE-3:
configure
    service
         pw-template 3 name "PW3" create
         exit
         epipe 4 name "Epipe4" customer 1 create
            bgp
                route-distinguisher 65536:43
                route-target export target:65536:4 import target:65536:4
                pw-template-binding 3
                exit
            exit
            bgp-vpws
                ve-name "PE-3"
                    ve-id 3
                exit
                remote-ve-name "PE-2"
                    ve-id 2
                exit
                no shutdown
            exit
            site "SITEB" create
                site-id 1
                sap 1/1/4:4
                site-preference 10
                no shutdown
            exit
            sap 1/1/4:4 create
            exit
            no shutdown
        exit
```

Epipe 4 is configured on PE-2 as follows. Two remote VE names are configured, PE-1 and PE-3 (this is the maximum number allowed).

```
# on PE-2:
configure
    service
         pw-template 3 name "PW3" create
         exit
         epipe 4 name "Epipe4" customer 1 create
            bgp
                route-distinguisher 65536:42
                route-target export target:65536:4 import target:65536:4
                pw-template-binding 3
                exit
            exit
            bgp-vpws
                ve-name "PE-2"
                    ve-id 2
                exit
                remote-ve-name "PE-1"
                    ve-id 1
                exit
                remote-ve-name "PE-3"
                    ve-id 3
                exit
                no shutdown
            exit
            sap 1/1/4:4 create
            exit
            no shutdown
```

Compared with the single pseudowire solution, both pseudowires are signaled and up on all PEs. The pseudowire with the higher preference is forwarding traffic (to PE-1), while the Tx status to the standby PE-3 is set to inactive, as follows:

```
*A:PE-2# show service l2-route-table bgp-vpws detail

===============================================================================
Services: L2 Bgp-Vpws Route Information - Summary
===============================================================================

---snip---

Svc Id         : 4
VeId           : 1
PW Temp Id     : 3
RD             : *65536:41
Next Hop       : 192.0.2.1
State (D-Bit)  : up(0)
Path MTU       : 1514
Control Word   : 0
Seq Delivery   : 0
Status         : active
Tx Status      : active
CSV            : 0
Preference     : 200
Sdp Bind Id    : 32767:4294967289

Svc Id         : 4
VeId           : 3
PW Temp Id     : 3
RD             : *65536:43
Next Hop       : 192.0.2.3
State (D-Bit)  : up(0)
Path MTU       : 1514
Control Word   : 0
Seq Delivery   : 0
Status         : active
Tx Status      : inactive
CSV            : 0
Preference     : 10
Sdp Bind Id    : 32766:4294967288
===============================================================================
```

The choice of pseudowire to be used to transmit traffic from PE-2 to PE-1 can also be seen in the endpoint created in the BGP VPWS service. Endpoints are automatically created for the pseudowires within a BGP VPWS service, regardless of whether active/standby pseudowires are used; these endpoints are created with a system generated name that ends with the BGP VPWS service id.

```
*A:PE-2# show service id 4 endpoint
```

```
===============================================================================
Service 4 endpoints
===============================================================================
Endpoint name                : _tmnx_BgpVpws-4
Description                  : Automatically created BGP-VPWS endpoint
Creation Origin              : bgpVpws
Revert time                  : 0
Act Hold Delay               : 0
Standby Signaling Master     : false
Standby Signaling Slave      : false
Tx Active (SDP)              : 32767:4294967289
Tx Active Up Time            : 0d 00:02:07
Revert Time Count Down       : never
Tx Active Change Count       : 3
Last Tx Active Change        : 03/04/2021 16:04:40
-------------------------------------------------------------------------------
Members
-------------------------------------------------------------------------------
Spoke-sdp: 32766:4294967288 Prec:4                 Oper Status: Up
Spoke-sdp: 32767:4294967289 Prec:4                 Oper Status: Up
===============================================================================
===============================================================================
```

The following command has no effect on an automatically created VPWS endpoint.

```
tools perform service id <service-id> endpoint <endpoint-name> force-switchover <..>
```

Conclusion

BGP VPWS allows the delivery of Layer 2 virtual private wire services to customers where BGP is commonly used. This chapter shows the configuration of single and dual-homed BGP VPWS services together with the associated show output, which can be used to verify and troubleshoot them.

BGP VPLS

This chapter describes advanced BGP VPLS configurations.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS Release 9.0.R3. The CLI in the current edition corresponds to SR OS Release 20.10.R2. There are no prerequisites for this configuration.

Overview

The following two IETF standards describe the provisioning of Virtual Private LAN Services (VPLS).

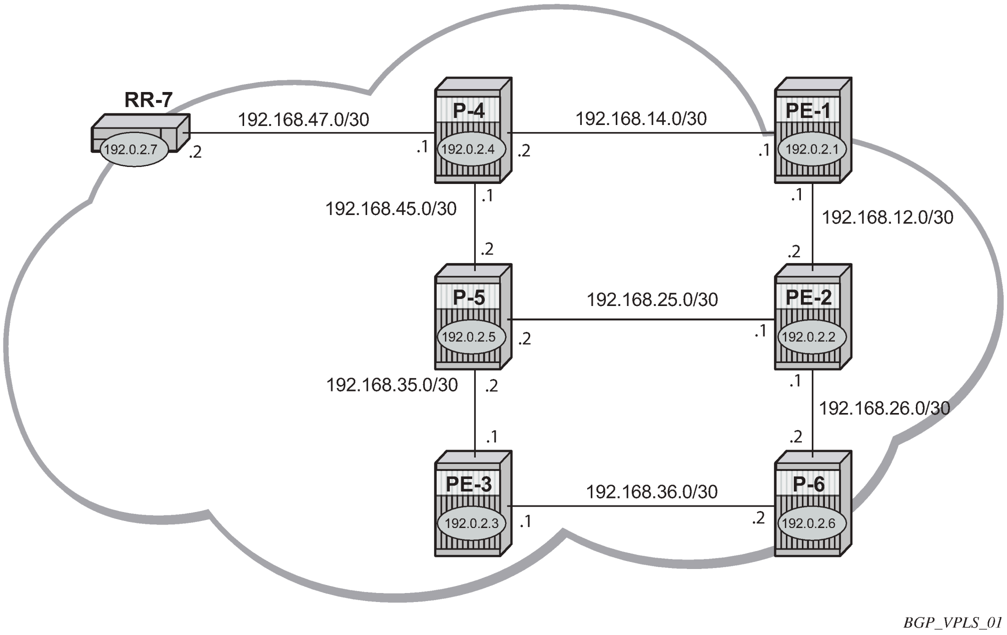
- RFC 4762, *Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling*, describes Label Distribution Protocol (LDP) VPLS, where VPLS pseudowires are signaled using LDP between VPLS Provider Edge (PE) routers, either configured manually or auto-discovered using BGP.
- RFC 4761, *Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling*, describes the use of Border Gateway Protocol (BGP) for both the auto-discovery of VPLS PEs and signaling of pseudowires between such PEs.

The purpose of this chapter is to describe the configuration and troubleshooting for BGP-VPLS.

Knowledge of BGP-VPLS RFC 4761 architecture and functionality is assumed throughout this chapter, as well as knowledge of Multi-Protocol BGP (MP-BGP).

Figure 30: Example topology shows the example topology with seven SR OS nodes located in the same Autonomous System (AS).

Figure 30: Example topology



There are three Provider Edge (PE) routers, and RR-7 acts as a Route Reflector (RR) for the AS. The PE routers are all VPLS-aware, the Provider (P) routers are VPLS-unaware and do not take part in the BGP process.

The following configuration tasks are completed as a prerequisite:

- IS-IS or OSPF on each of the network interfaces between the PE/P routers and RR.
- MPLS is configured on all interfaces between PE routers and P routers. MPLS is not required between P-4 and RR-7.
- LDP is configured on interfaces between PE and P routers. It is not required between P-4 and the RR-7.
- The RSVP protocol is enabled.

BGP VPLS

In this architecture, a VPLS instance is a collection of local VPLS instances present on a number of PEs in a provider network. In this context, any VPLS-aware PE is also known as a VPLS Edge (VE) device.

The PEs communicate with each other at the control plane level by means of BGP updates containing BGP-VPLS Network Layer Reachability Information (NLRI). Each update contains enough information for a PE to determine the presence of other local VPLS instances on peering PEs and to set up pseudowire connectivity for data flow between peers containing a local VPLS within the same VPLS instance. Therefore, auto-discovery and pseudowire signaling are achieved using a single BGP update message.

Each PE within a VPLS instance is identified by a VPLS Edge identifier (VE-ID) and the presence of a VPLS instance is determined using the exchange of standard BGP extended community RTs between PEs.

Each PE will advertise, via the route reflectors, the presence of each VPLS instance to all other PEs, along with a block of multiplexer labels that can be used to communicate between such instances plus a BGP next hop that determines a labeled transport tunnel between PEs.

Each VPLS instance is configured with import and export RT extended communities for topology control, along with VE identification.

Configuration

The first step is to configure an MP-iBGP session between each of the PEs and the RR for the L2-VPN address family, as follows:

```
# on PE-1, PE-2, and PE-3:
configure
    router Base
        autonomous-system 65536
        bgp
            group "INTERNAL"
                family l2-vpn
                peer-as 65536
                neighbor 192.0.2.7
                exit
            exit
            no shutdown
        exit
```

The IP addresses can be derived from Figure 30: Example topology.

The configuration for RR-7 is as follows:

```
# on RR-7:
configure
    router Base
        autonomous-system 65536
        bgp
            cluster 1.1.1.1
            group "RR-INTERNAL"
                family l2-vpn
                peer-as 65536
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
            no shutdown
        exit
```

On PE-1, the BGP session with RR-7 is established with the L2-VPN address family capability negotiated, as follows:

```
*A:PE-1# show router bgp neighbor 192.0.2.7

===============================================================================
BGP Neighbor
===============================================================================
-------------------------------------------------------------------------------
```

```
Peer                 : 192.0.2.7
Description          : (Not Specified)
Group                : INTERNAL
-------------------------------------------------------------------------------
Peer AS              : 65536            Peer Port            : 51108
Peer Address         : 192.0.2.7
Local AS             : 65536            Local Port           : 179
Local Address        : 192.0.2.1
Peer Type            : Internal         Dynamic Peer         : No
State                : Established      Last State           : Established
Last Event           : recvOpen
Last Error           : Cease (Connection Collision Resolution)
Local Family         : L2-VPN
Remote Family        : L2-VPN
Hold Time            : 90               Keep Alive           : 30
Min Hold Time        : 0
Active Hold Time     : 90               Active Keep Alive    : 30
Cluster Id           : None
Preference           : 170              Num of Update Flaps  : 0
---snip---

Local Capability     : RtRefresh MPBGP 4byte ASN
Remote Capability    : RtRefresh MPBGP 4byte ASN
---snip---
```

On RR-7, the BGP sessions with each PE are established, and have negotiated the L2-VPN address family capability, as follows:

```
*A:RR-7# show router bgp summary all

===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
Description
ServiceId          AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
192.0.2.1
Def. Instance  65536        7    0 00h02m19s 0/0/0 (L2VPN)
                            7    0
192.0.2.2
Def. Instance  65536        7    0 00h02m19s 0/0/0 (L2VPN)
                            7    0
192.0.2.3
Def. Instance  65536        7    0 00h02m19s 0/0/0 (L2VPN)
                            7    0

-------------------------------------------------------------------------------
```

A full mesh of RSVP-TE LSPs is configured between the PE routers. On PE-1, the MPLS interface and LSP configuration are as follows:

```
# on PE-1:
configure
    router Base
        mpls
            interface "int-PE-1-PE-2"
            exit
            interface "int-PE-1-P-4"
            exit
            path "loose"
                no shutdown
            exit
            lsp "LSP-PE-1-PE-2"
                to 192.0.2.2
                primary "loose"
                exit
                no shutdown
            exit
            lsp "LSP-PE-1-PE-3"
                to 192.0.2.3
                primary "loose"
                exit
                no shutdown
            exit
            no shutdown
        exit
```

The MPLS and LSP configuration for PE-2 and PE-3 are similar to that of PE-1 with the appropriate interfaces and LSP names configured.

BGP VPLS PE configuration

Pseudowire templates

Pseudowire templates are used by BGP to dynamically instantiate SDP bindings for a service to signal the egress service de-multiplexer labels used by remote PEs to reach the local PE.

The template determines the signaling parameters of the pseudowire, control word presence, MAC-pinning, filters, and so on, plus other usage characteristics such as split horizon groups (SHGs).

The MPLS transport tunnel between PEs can be signaled using LDP or RSVP-TE.

LDP based pseudowires can be automatically instantiated. RSVP-TE based SDPs have to be pre-provisioned.

Pseudowire templates for auto-SDP creation using LDP

In order to use an LDP transport tunnel for data flow between PEs, link layer LDP must be configured between all PEs/Ps, so that a transport label for each PE’s system interface is available.

```
# on PE-1:
configure
    router Base
        ldp
            interface-parameters
                interface "int-PE-1-PE-2" dual-stack
                    ipv4
                        no shutdown
                    exit
                    no shutdown
                exit
                interface "int-PE-1-P-4" dual-stack
                    ipv4
                        no shutdown
                    exit
                    no shutdown
                exit
            exit
        exit
```

Using this mechanism, SDPs can be auto-instantiated with SDP-IDs starting at the higher end of the SDP numbering range, such as 32767. Any subsequent SDPs created use SDP-IDs decrementing from this value.

A pseudowire template is required containing an SHG. Each SDP created with this template is contained within an SHG so that traffic cannot be forwarded between them.

```
# on PE-1:
configure
    service
        pw-template 1 name "PW1" create
            split-horizon-group "VPLS-SHG"
            exit
        exit
```

The pseudowire template also has the following options available when used for BGP-VPLS:

```
*A:PE-1>config>service# pw-template ?

  ---snip---
  [no] controlword
  ---snip---
  [no] force-vlan-vc-forwarding
  ---snip---
   vc-type {ether | vlan}
  ---snip---
```

- The control word will determine whether the C flag is set in the Layer 2 extended community and, therefore, if a control word is used in the pseudowire.
- The encap type in the Layer 2 extended community is always 19 (VPLS encap), therefore, the vc-type will always be **ether** regardless of the configured value on the vc-type.
- The **force-vlan-vc-forwarding** command will add a tag (equivalent to **vc-type vlan**) and will allow for customer QoS transparency (dot1p + Drop Eligibility (DE) bits).

Pseudowire templates for provisioned SDPs using RSVP-TE

To use an RSVP-TE tunnel as transport between PEs, it is necessary to bind the RSVP-TE LSP between PEs to an SDP.

The following SDP is created from PE-1 to PE-2:

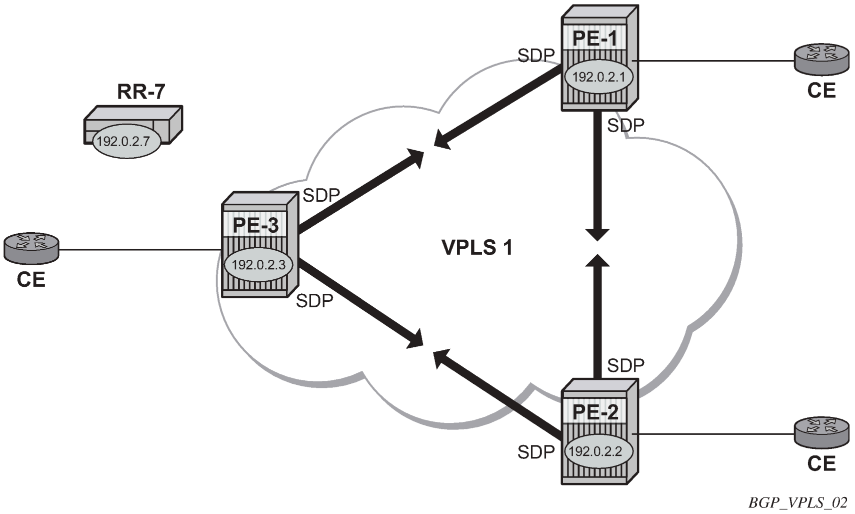
```
# on PE-1:
configure
    service
        sdp 12 mpls create
            description "SDP-PE-1-PE-2_RSVP_BGP"
            signaling bgp
            far-end 192.0.2.2
            lsp "LSP-PE-1-PE-2"
            no shutdown
        exit
```

The **signaling bgp** parameter is required for BGP-VPLS to be able to use this SDP. Conversely, SDPs that are bound to RSVP-based LSPs with signaling set to the default value of **tldp** will not be used as SDPs within BGP-VPLS.

BGP VPLS using auto-provisioned SDPs

Figure 31: BGP VPLS using auto-provisioned SDPs shows a VPLS instance where SDPs are auto-provisioned. In this case, the transport tunnels are LDP-signaled.

Figure 31: BGP VPLS using auto-provisioned SDPs



The following shows the configuration required on PE-1 for a BGP-VPLS service using a pseudowire template configured for auto-provisioning of SDPs.

```
# on PE-1:
configure
    service
        pw-template 1 name "PW1" create
            split-horizon-group "VPLS-SHG"
            exit
        exit
        vpls 1 name "VPLS1_PE-1" customer 1 create
            bgp
                route-distinguisher 65536:1
                route-target export target:65536:1 import target:65536:1
                pw-template-binding 1
                exit
            exit
            bgp-vpls
                max-ve-id 10
                ve-name "PE-1"
                    ve-id 1
                exit
                no shutdown
            exit
            sap 1/1/4:1.0 create
            exit
            no shutdown
        exit
```

The **bgp** context specifies parameters which are valid for all of the VPLS BGP applications, such as BGP multi-homing (BGP-MH), BGP auto-discovery (BGP-AD), and BGP-VPLS.

Within the **bgp** context, parameters are configured that are used by neighboring PEs to determine membership of a VPLS instance, such as the auto-discovery of PEs containing the same VPLS instance; the route distinguisher (RD) is configured, along with the route target (RT) extended communities.

RT communities are used to determine membership of a VPLS instance. The import RT at the BGP level is mandatory. The pseudowire template bind is then applied by the service manager on the received routes matching the RT value.

Within the **bgp-vpls** context, the signaling parameters are configured. These determine the service labels required for the data plane of the VPLS instance.

The VPLS edge ID (VE-ID) is a numerical value assigned to each PE within a VPLS instance. This value should be unique for a VPLS instance; no two PEs within the same instance should have the same VE-ID values.

A more specific RT can be applied to a pseudowire template in order to define a specific pseudowire topology, rather than only a full mesh, using the command within the **bgp** context:

pw-template *template-id* **[split-horizon-group** *groupname***] [import-rt** *import-rt-value (up to 5 max)***]**

Changes to the import policies are not taken once the pseudowire has been set up (changes on RT are refreshed though). Pseudowire templates can be re-evaluated with the command **tools perform service eval-pw-template**. The **eval-pw-template** command checks whether all the bindings using this pseudowire template policy are still meant to use this policy.

If the policy has changed and **allow-service-impact** is true, then the old binding is removed and it is re-added with the new template.

VE-ID and BGP label allocations

The choice of VE-ID is crucial in ensuring efficient allocation of de-multiplexer labels. The most efficient choice is for VE-IDs to be allocated starting at 1 and incrementing for each PE as the following section explains.

The **max-ve-id** *value* determines the range of the VE-ID value that can be configured. If a PE receives a BGP-VPLS update containing a VE-ID with a greater value than the configured **max-ve-id**, then the update is dropped and no service labels are installed for this VE-ID.

The **max-ve-id** command also checks the locally-configured VE-ID, and prevents a higher value from being used.

Each PE allocates blocks of labels per VPLS instance to remote PEs, in increments of eight labels. It achieves this by advertising three parameters in a BGP update message,

- A label base (LB) which is the lowest label in the block
- A VE Block Size (VBS) which is always eight labels, and cannot be changed
- A VE Base Offset (VBO).

This defines a block of labels in the range (LB, LB+1, ..., LB+VBS-1).

As an example, if the label base (LB) = 524272, then the range for the block is 524272 to 524279, which is exactly eight labels, as per the block size. (The last label in the block is calculated as 524272+8-1 = 524279)

The label allocated by the PE to each remote PE within the VPLS is chosen from this block and is determined by its VE-ID. In this way, each remote PE has a unique de-multiplexer label for that VPLS.

To reduce label wastage, contiguous VE-IDs in the range (N..N+7) per VPLS should be chosen, where N>0.

Assuming a collection of PEs with contiguous VE-IDs, the following labels will be chosen by PEs from the label block allocated by PE-1 which has a VE-ID =1.

Table 1: VE-IDs and Labels

VE-ID	Label
2	524273
3	524274
4	524275
5	524276
6	524277
7	524278
8	524279

This shows that the label allocated to a PE is (LB+VEID-1). The “1” is the VE block offset (VBO).

This means that the label allocated to a PE router within the VPLS can now be written as (LB + VEID - VBO), which means that (VEID - VBO) calculation must always be at least zero and be less than the block size, which is always 8.

For VE-ID < 8, a label will be allocated from this block.

For the next block of 8 VE-IDs (VE-ID 9 to VE-ID 16) a new block of 8 labels must be allocated, so a new BGP update is sent, with a new label base, and a block offset of 9.

Table 2: VE-IDs and Number of Labels shows how the choice of VE-IDs can affect the number of label blocks allocated, and therefore the number of labels:

Table 2: VE-IDs and Number of Labels

VE-ID	Block Offset	Labels Allocated
1-8	1	8
9-16	9	8
17-24	17	8
25-32	25	8

VE-ID	Block Offset	Labels Allocated
33-40	33	8
41-48	41	8
49-56	49	8

This shows that the most efficient use of labels occurs when the VE-IDs for a set of PEs are chosen from the same block offset.

If VE-IDs are chosen that map to different block offsets, then each PE will have to send multiple BGP updates to signal service labels. Each PE sends label blocks in BGP updates to each of its BGP neighbors for all label blocks in which at least one VE-ID has been seen by this PE (it does not advertise label blocks which do not contain an active VE-ID, where active VE-ID means the VE-ID of this PE or any other PE in this VPLS).

The **max-ve-id** must be configured first, and determines the maximum value of the VE-ID that can be configured within the PE. The VE-ID value cannot be higher than this within the PE configuration, VE-ID <= max-VE-ID. Similarly, if the VE-ID within a received NLRI is higher than the **max-ve-id** *value*, it will not be accepted as valid consequently the max-ve-id configured on all PEs must be greater than or equal to any VE-ID used in the VPLS.

Only one VE-ID value can be configured. If the VE-ID value is changed, BGP withdraws the NLRI and sends a route-refresh.

If the same VE-ID is used in different PEs for the same VPLS, a Designated Forwarder (DF) election takes place.

Executing the **shutdown** command triggers an MP-UNREACH-NLRI from the PE to all BGP peers.

The **no shutdown** command triggers an MP-REACH-NLRI to the same peers.

PE-2 service creation

On PE-2, a VPLS service using pseudowire template 1 is created. In order to make the label allocation more efficient, PE-2 has been allocated a VE-ID value of 2. For completeness, the pseudowire template is also shown.

```
# on PE-2:
configure
    service
        pw-template 1 name "PW1" create
            split-horizon-group "VPLS-SHG"
            exit
        exit
        vpls 1 name "VPLS1_PE-2" customer 1 create
            bgp
                route-distinguisher 65536:1
                route-target export target:65536:1 import target:65536:1
                pw-template-binding 1
                exit
            exit
            bgp-vpls
                max-ve-id 10
                ve-name "PE-2"
                    ve-id 2
                exit
                no shutdown
            exit
            sap 1/1/4:1.0 create
            exit
            no shutdown
        exit
```

The **max-ve-id** *value* is set to 10 to allow an increase in the number of PEs that could be a part of this VPLS instance.

PE-3 service creation

The following configuration creates a VPLS instance on PE-3, using a VE-ID value of 3.

```
# on PE-3:
configure
    service
        pw-template 1 name "PW1" create
            split-horizon-group "VPLS-SHG"
            exit
        exit
        vpls 1 name "VPLS1_PE-3" customer 1 create
            bgp
                route-distinguisher 65536:1
                route-target export target:65536:1 import target:65536:1
                pw-template-binding 1
                exit
            exit
            bgp-vpls
                max-ve-id 10
                ve-name "PE-3"
                    ve-id 3
                exit
                no shutdown
            exit
            sap 1/1/4:1.0 create
            exit
            no shutdown
        exit
```

PE-1 service operation verification

The following command shows that the BGP-VPLS site is enabled on PE-1.

```
*A:PE-1# show service id 1 bgp-vpls

===============================================================================
BGP VPLS Information
===============================================================================
Max Ve Id            : 10                   Admin State      : Enabled
VE Name              : PE-1                 VE Id            : 1
PW Tmpl used         : 1
===============================================================================
```

The following command shows that the service is operationally up on PE-1:

```
*A:PE-1# show service id 1 base
```

```
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : VPLS
MACSec enabled    : no
Name              : VPLS1_PE-1
---snip---

Admin State       : Up                  Oper State        : Up
MTU               : 1514
SAP Count         : 1                   SDP Bind Count    : 2
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:1.0                            qinq         1522    1522    Up   Up
sdp:32766:4294967294 SB(192.0.2.3)       BgpVpls      0       1556    Up   Up
sdp:32767:4294967295 SB(192.0.2.2)       BgpVpls      0       1556    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

The SAP and SDPs are all operationally up. The *SB* flags for the SDPs signify Spoke and BGP.

The ingress labels for PE-2 and PE-3—the labels allocated by PE-1—can be seen as follows:

```
*A:PE-1# show service id 1 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
32766:4294967294 BgpVpls  192.0.2.3       Up      Up        524274    524272
32767:4294967295 BgpVpls  192.0.2.2       Up      Up        524273    524270
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
```

As can be seen from the following output, a BGP-VPLS NLRI update is sent to the route reflector (192.0.2.7) and is received by each PE.

On PE-1, debugging is enabled for BGP updates:

```
# on PE-1:
debug
    router "Base"
        bgp
            update
```

PE-1 has sent the following BGP NLRI update for VPLS 1 to RR-7:

```
1 2021/01/25 15:27:27.169 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.7
"Peer 1: 192.0.2.7: UPDATE
Peer 1: 192.0.2.7 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 72
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family L2VPN
        NextHop len 4 NextHop 192.0.2.1
        [VPLS/VPWS] preflen 17, veid: 1, vbo: 1, vbs: 8, label-base: 524272, RD 65536:1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:65536:1
        l2-vpn/vrf-imp:Encap=19: Flags=none: MTU=1514: PREF=0
"
```

The control flags within the extended community indicate the status of the VPLS instance.

The control flag D indicates that all attachment circuits are Down, or the VPLS is disabled. The flags are used in BGP-MH when determining which PEs are DF, see chapter BGP Multi-Homing for VPLS Networks.

When flags=none, then all attachment circuits are up. In the preceding example, no flags are present, but should all SAPs become operationally down, then the control flag D would be seen in the debug message. To simulate this, the SAP 1/1/4:1 is disabled on PE-1:

```
# on PE-1:
configure
    service
        vpls "VPLS1_PE-1"
            sap 1/1/4:1.0
                shutdown
```

All SAPs in VPLS 1 on PE-1 are operationally down, so PE-1 sends a BGP update message with control flag D set, as follows:

```
5 2021/01/25 15:40:08.169 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.7
"Peer 1: 192.0.2.7: UPDATE
Peer 1: 192.0.2.7 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 72
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family L2VPN
        NextHop len 4 NextHop 192.0.2.1
        [VPLS/VPWS] preflen 17, veid: 1, vbo: 1, vbs: 8, label-base: 524272, RD 65536:1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:65536:1
        l2-vpn/vrf-imp:Encap=19: Flags=D: MTU=1514: PREF=0
"
```

The SAP is re-enabled with the following command on PE-1:

```
# on PE-1:
configure
    service
        vpls "VPLS1_PE-1"
            sap 1/1/4:1.0
                no shutdown
```

The BGP VPLS signaling parameters are also present in the BGP update message, namely the VE-ID of the PE within the VPLS instance, the VBO and VBS, and the label base. The target indicates the VPLS instance, which must be matched against the import RTs of the receiving PEs.

The signaling parameters can be seen within the BGP update with following command:

```
*A:PE-1# show router bgp routes l2-vpn rd 65536:1 hunt
===============================================================================
 BGP Router ID:192.0.2.1        AS:65536       Local AS:65536
===============================================================================
---snip---

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
Route Type     : VPLS
Route Dist.    : 65536:1
VeId           : 1                      Block Size     : 8
Base Offset    : 1                      Label Base     : 524272
Nexthop        : 192.0.2.1
To             : 192.0.2.7
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None                   IGP Cost       : n/a
Connector      : None
Community      : target:65536:1
                 l2-vpn/vrf-imp:Encap=19: Flags=none: MTU=1514: PREF=0
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.7
Origin         : IGP
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : n/a
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0

-------------------------------------------------------------------------------
Routes : 4
===============================================================================
```

In this configuration example, PE-1 (192.0.2.1) with VE-ID =1 has sent an update with base offset (VBO) =1, block size (VBS) = 8, and label base 524272. This means that labels 524272 (LB) to 524279 (LB +VBS-1) are available as de-multiplexer labels, egress labels to be used to reach PE-1 for VPLS 1.

PE-2 receives this update from PE-1. This is seen as a valid VPLS BGP route from PE-1 through the route reflector with next-hop 192.0.2.1.

```
*A:PE-2# show router bgp routes l2-vpn rd 65536:1
===============================================================================
 BGP Router ID:192.0.2.2        AS:65536       Local AS:65536
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP L2VPN Routes
===============================================================================
Flag  RouteType                   Prefix                             MED
```

```
      RD                            SiteId                            Label
      Nexthop                       VeId                   BlockSize   LocalPref
      As-Path                       BaseOffset             vplsLabelBa
                                                           se
-------------------------------------------------------------------------------
u*>i  VPLS                          -                      -           0
      65536:1                       -                                  -
      192.0.2.1                     1                      8           100
      No As-Path                    1                      524272
i     VPLS                          -                      -           0
      65536:1                       -                                  -
      192.0.2.2                     2                      8           100
      No As-Path                    1                      524270
u*>i  VPLS                          -                      -           0
      65536:1                       -                                  -
      192.0.2.3                     3                      8           100
      No As-Path                    1                      524272
-------------------------------------------------------------------------------
Routes : 3
===============================================================================
```

PE-2 uses this information in conjunction with its own VE-ID to calculate the egress label toward PE-1, using the condition VBO < VE-ID < (VBO+VBS).

The VE-ID of PE-2 is in the Label Block covered by VBO =1, thus,

Label calculation = label base + local VE-ID - Base offset

= 524272 + 2 - 1

Egress label used = 524273

This is verified using the following command on PE-2 where the egress label toward PE-1 (192.0.2.1) is 524273.

```
*A:PE-2# show service id 1 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
32766:4294967294 BgpVpls  192.0.2.3       Up      Up        524272    524273
32767:4294967295 BgpVpls  192.0.2.1       Up      Up        524270    524273
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
```

PE-3 also receives this update from PE-1 by the RR. This is seen as a valid VPLS BGP route from PE-1 with next-hop 192.0.2.1.

```
*A:PE-3# show router bgp routes l2-vpn rd 65536:1
===============================================================================
 BGP Router ID:192.0.2.3        AS:65536       Local AS:65536
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP L2VPN Routes
```

```
===============================================================================
Flag  RouteType                   Prefix                             MED
      RD                          SiteId                             Label
      Nexthop                     VeId                   BlockSize   LocalPref
      As-Path                     BaseOffset             vplsLabelBa
                                                         se
-------------------------------------------------------------------------------
u*>i  VPLS                        -                      -           0
      65536:1                     -                                  -
      192.0.2.1                   1                      8           100
      No As-Path                  1                      524272
u*>i  VPLS                        -                      -           0
      65536:1                     -                                  -
      192.0.2.2                   2                      8           100
      No As-Path                  1                      524270
i     VPLS                        -                      -           0
      65536:1                     -                                  -
      192.0.2.3                   3                      8           100
      No As-Path                  1                      524272
-------------------------------------------------------------------------------
Routes : 3
===============================================================================
```

The VE-ID of PE-3 is also in the label block covered by block offset VBO =1.

Label calculation = label base + local VE-ID - VBO

= 524272 + 3 - 1

Egress label used = 524274

This is verified using the following command on PE-3 where egress label toward 192.0.2.1 is 524274.

```
*A:PE-3# show service id 1 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
32766:4294967293 BgpVpls  192.0.2.2       Up      Up        524273    524272
32767:4294967294 BgpVpls  192.0.2.1       Up      Up        524272    524274
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
```

PE-2 service operation verification

The service is operationally up on PE-2, as follows.

```
*A:PE-2# show service id 1 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : VPLS
MACSec enabled    : no
Name              : VPLS1_PE-2
---snip---
```

```
Admin State       : Up                  Oper State       : Up
MTU               : 1514
SAP Count         : 1                   SDP Bind Count   : 2
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:1.0                            qinq         1522    1522    Up   Up
sdp:32766:4294967294 SB(192.0.2.3)       BgpVpls      0       1556    Up   Up
sdp:32767:4294967295 SB(192.0.2.1)       BgpVpls      0       1556    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

PE-2 de-multiplexer label calculation

In the same way that PE-1 allocates a label base (LB), block size (VBS), and base offset (VBO), PE-2 also allocates the same parameters for PE-1 and PE-3 to calculate the egress service label required to reach PE-2.

```
*A:PE-2# show router bgp routes l2-vpn rd 65536:1 hunt
===============================================================================
 BGP Router ID:192.0.2.2        AS:65536       Local AS:65536
===============================================================================
---snip---

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
Route Type     : VPLS
Route Dist.    : 65536:1
VeId           : 2                      Block Size     : 8
Base Offset    : 1                      Label Base     : 524270
Nexthop        : 192.0.2.2
To             : 192.0.2.7
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None                   IGP Cost       : n/a
Connector      : None
Community      : target:65536:1
                 l2-vpn/vrf-imp:Encap=19: Flags=none: MTU=1514: PREF=0
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.7
Origin         : IGP
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : n/a
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0

-------------------------------------------------------------------------------
Routes : 4
===============================================================================
```

This is verified using the following command on PE-1 to show the egress label toward PE-2 (192.0.2.2) where the egress label toward PE-2 = 524270 + 1 – 1 = 524270.

```
*A:PE-1# show service id 1 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
32766:4294967294 BgpVpls  192.0.2.3       Up      Up        524274    524272
32767:4294967295 BgpVpls  192.0.2.2       Up      Up        524273    524270
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
```

This is also verified using the following command on PE-3 to show the egress label toward PE-2 (192.0.2.2) where the egress label toward PE-2 = 524270 + 3 – 1 = 524272.

```
*A:PE-3# show service id 1 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
32766:4294967294 BgpVpls  192.0.2.2       Up      Up        524273    524272
32767:4294967295 BgpVpls  192.0.2.1       Up      Up        524272    524274
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
```

PE-3 service operation verification

The following command shows that the service is operationally up on PE-3:

```
*A:PE-3# show service id 1 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : VPLS
MACSec enabled    : no
Name              : VPLS1_PE-3
---snip---

Admin State       : Up                  Oper State        : Up
MTU               : 1514
SAP Count         : 1                   SDP Bind Count    : 2
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
```

```
sap:1/1/4:1.0                            qinq         1522    1522    Up   Up
sdp:32766:4294967293 SB(192.0.2.2)       BgpVpls      0       1556    Up   Up
sdp:32767:4294967294 SB(192.0.2.1)       BgpVpls      0       1556    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

```
*A:PE-3# show service id 1 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
32766:4294967293 BgpVpls  192.0.2.2       Up      Up        524273    524272
32767:4294967294 BgpVpls  192.0.2.1       Up      Up        524272    524274
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
```

PE-3 de-multiplexer label verification

PE-3 also allocates the required parameters for PE-1 and PE-2 to calculate the egress service label required to reach PE-3.

This is verified using the following command on PE-1 to show the egress label toward PE-3 (192.0.2.3) (524272) where egress label toward PE-2 = 524270. The Label Base equals 524272 on PE-3 and 524270 on PE-2.

```
*A:PE-1# show service id 1 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
32766:4294967294 BgpVpls  192.0.2.3       Up      Up        524274    524272
32767:4294967295 BgpVpls  192.0.2.2       Up      Up        524273    524270
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
```

This is also verified using the following command on PE-2 to show the egress label toward PE-3 (192.0.2.3) which is using auto-provisioned SDP 32766.

```
*A:PE-2# show service id 1 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
32766:4294967294 BgpVpls  192.0.2.3       Up      Up        524272    524273
32767:4294967295 BgpVpls  192.0.2.1       Up      Up        524270    524273
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
```

```
===============================================================================
```

This example has shown that for VPLS instance with 3 PEs, not all labels allocated by a PE will be used by remote PEs as de-multiplexer service labels. There will be some wastage of label space, so there is a necessity to choose VE-IDs that keep this waste to a minimum.

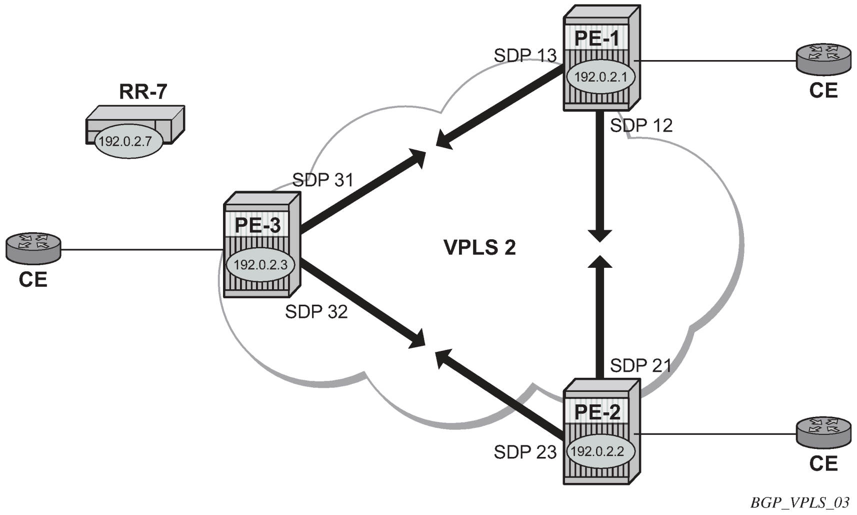
The next example will show an even more wasteful use of labels by using a random choice of VE-IDs.

BGP VPLS using pre-provisioned SDP

It is possible to configure BGP-VPLS instances that use RSVP-TE transport tunnels. In this case, the SDP must be created with the MPLS LSPs mapped and with signaling set to BGP, as the service labels are signaled using BGP. The pseudowire template configured within the BGP-VPLS instance must use the **use-provisioned-sdp** keyword. This example also examines the effect of using VE-IDs that are not all within the same contiguous block.

Figure 32: BGP VPLS using pre-provisioned SDP shows an example of a VPLS instance where SDPs are pre-provisioned with RSVP-TE signaled transport tunnels.

Figure 32: BGP VPLS using pre-provisioned SDP



On the PEs, the following SDPs are configured with RSVP transport tunnels.

```
# on PE-1:
configure
    service
        sdp 12 mpls create
            description "SDP-PE-1-PE-2_RSVP_BGP"
            signaling bgp
            far-end 192.0.2.2
            lsp "LSP-PE-1-PE-2"
            no shutdown
        exit
        sdp 13 mpls create
            description "SDP-PE-1-PE-3_RSVP_BGP"
            signaling bgp
            far-end 192.0.2.3
            lsp "LSP-PE-1-PE-3"
            no shutdown
        exit
```

```
# on PE-2:
configure
    service
        sdp 21 mpls create
            description "SDP-PE-2-PE-1_RSVP_BGP"
            signaling bgp
            far-end 192.0.2.1
            lsp "LSP-PE-2-PE-1"
            no shutdown
        exit
        sdp 23 mpls create
            description "SDP-PE-2-PE-3_RSVP_BGP"
            signaling bgp
            far-end 192.0.2.3
            lsp "LSP-PE-2-PE-3"
            no shutdown
        exit
```

```
# on PE-3:
configure
    service
        sdp 31 mpls create
            description "SDP-PE-3-PE-1_RSVP_BGP"
            signaling bgp
            far-end 192.0.2.1
            lsp "LSP-PE-3-PE-1"
            no shutdown
        exit
        sdp 32 mpls create
            description "SDP-PE-3-PE-2_RSVP_BGP"
            signaling bgp
            far-end 192.0.2.2
            lsp "LSP-PE-3-PE-2"
            no shutdown
        exit
```

Pre-provisioned BGP-SDPs can also be used with BGP-VPLS. For reference, they are configured as follows:

```
# on PE-3:
configure
    service
        sdp 332 mpls create
            signaling bgp
            far-end 192.0.2.2
            no shutdown
        exit
```

To create an SDP within a service that uses the RSVP transport tunnel, a pseudowire template is required that has the **use-provisioned-sdp** parameter set. It is also possible to configure the **prefer-provisioned-sdp** parameter, see chapter *LDP VPLS Using BGP Auto-Discovery – Prefer Provisioned SDP*.

Once again, an SHG is included to prevent forwarding between pseudowires.

The following pseudowire template is provisioned on all PEs:

```
# on PE-1, PE-2, and PE-3:
configure
    service
        pw-template 2 name "PW2" use-provisioned-sdp create
            split-horizon-group "VPLS-SHG"
            exit
        exit
```

The following output shows the configuration required for a BGP-VPLS service using a pseudowire template configured for using pre-provisioned RSVP-TE SDPs.

```
# on PE-1:
configure
    service
        vpls 2 name "VPLS2_PE-1" customer 1 create
            bgp
                route-distinguisher 65536:2
                route-target export target:65536:2 import target:65536:2
                pw-template-binding 2
                exit
            exit
            bgp-vpls
                max-ve-id 100
                ve-name "PE-1"
                    ve-id 1
                exit
                no shutdown
            exit
            sap 1/1/4:2.0 create
            exit
            no shutdown
        exit
```

The RD and RT extended community values for VPLS 2 are different from the ones in VPLS 1. The VE-ID value for PE-1 can be the same as the one in VPLS 1, but these must be different within the same VPLS instance on the other PEs — PE-2 should not have VE-ID = 1.

On PE-2, the configuration is as follows with the VE-ID value equal to 20, which will result in a label from a different block:

```
# on PE-2:
configure
    service
        vpls 2 name "VPLS2_PE-2" customer 1 create
            bgp
                route-distinguisher 65536:2
                route-target export target:65536:2 import target:65536:2
                pw-template-binding 2
                exit
            exit
            bgp-vpls
                max-ve-id 100
                ve-name "PE-2"
                    ve-id 20
                exit
                no shutdown
            exit
            sap 1/1/4:2.0 create
            exit
            no shutdown
        exit
```

On PE-3, the configuration is as follows with the VE-ID value equal to 3:

```
# on PE-3:
configure
    service
        vpls 2 name "VPLS2_PE-3" customer 1 create
            bgp
                route-distinguisher 65536:2
                route-target export target:65536:2 import target:65536:2
                pw-template-binding 2
                exit
            exit
            bgp-vpls
                max-ve-id 100
                ve-name "PE-3"
                    ve-id 3
                exit
                no shutdown
            exit
            sap 1/1/4:2.0 create
            exit
            no shutdown
        exit
```

The service is operationally up on PE-1, as follows:

```
*A:PE-1# show service id 2 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 2                   Vpn Id            : 0
Service Type      : VPLS
MACSec enabled    : no
Name              : VPLS2_PE-1
---snip---

Admin State       : Up                  Oper State        : Up
MTU               : 1514
SAP Count         : 1                   SDP Bind Count    : 2
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:2.0                            qinq         1522    1522    Up   Up
sdp:12:4294967292 S(192.0.2.2)           BgpVpls      0       1556    Up   Up
sdp:13:4294967293 S(192.0.2.3)           BgpVpls      0       1556    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

The SDP 12 and 13 are the pre-provisioned SDPs.

The service is operationally up on PE-2, as follows:

```
*A:PE-2# show service id 2 base
```

```
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 2                   Vpn Id            : 0
Service Type      : VPLS
MACSec enabled    : no
Name              : VPLS2_PE-2
---snip---

Admin State       : Up                  Oper State        : Up
MTU               : 1514
SAP Count         : 1                   SDP Bind Count    : 2
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:2.0                            qinq         1522    1522    Up   Up
sdp:21:4294967292 S(192.0.2.1)           BgpVpls      0       1556    Up   Up
sdp:23:4294967293 S(192.0.2.3)           BgpVpls      0       1556    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

The service is operationally up on PE-3, as follows:

```
*A:PE-3# show service id 2 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 2                   Vpn Id            : 0
Service Type      : VPLS
MACSec enabled    : no
Name              : VPLS2_PE-3
---snip---

Admin State       : Up                  Oper State        : Up
MTU               : 1514
SAP Count         : 1                   SDP Bind Count    : 2
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:2.0                            qinq         1522    1522    Up   Up
sdp:31:4294967292 S(192.0.2.1)           BgpVpls      0       1556    Up   Up
sdp:32:4294967291 S(192.0.2.2)           BgpVpls      0       1556    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

PE-1 de-multiplexer label calculation

In the case of VPLS 1, all VE-IDs are in the range of a single label block. In the case of VPLS 2, the VE-IDs are in different blocks, for example, the VE-ID 20 is in a different block to VE-IDs 1 and 3.

As the label allocation is block-dependent, multiple label blocks must be advertised by each PE to encompass this.

Consider PE-1’s BGP update NLRIs.

```
*A:PE-1# show router bgp routes l2-vpn rd 65536:2 hunt
===============================================================================
 BGP Router ID:192.0.2.1        AS:65536       Local AS:65536
===============================================================================
---snip---

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
Route Type     : VPLS
Route Dist.    : 65536:2
VeId           : 1                       Block Size     : 8
Base Offset    : 1                       Label Base     : 524264
Nexthop        : 192.0.2.1
To             : 192.0.2.7
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None                   IGP Cost       : n/a
Connector      : None
Community      : target:65536:2
                 l2-vpn/vrf-imp:Encap=19: Flags=none: MTU=1514: PREF=0
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.7
Origin         : IGP
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : n/a
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0

Route Type     : VPLS
Route Dist.    : 65536:2
VeId           : 1                       Block Size     : 8
Base Offset    : 17                      Label Base     : 524256
Nexthop        : 192.0.2.1
To             : 192.0.2.7
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None                   IGP Cost       : n/a
Connector      : None
Community      : target:65536:2
                 l2-vpn/vrf-imp:Encap=19: Flags=none: MTU=1514: PREF=0
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.7
Origin         : IGP
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : n/a
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0

-------------------------------------------------------------------------------
Routes : 8
===============================================================================
```

Two NLRIs updates are sent to the route reflector, with the following label parameters:

1. LB = 524264, VBS = 8, VBO = 1
2. LB = 524256, VBS = 8, VBO = 17

PE-2 has a VE-ID of 20. Applying the condition VBO < VE-ID < (VBO+VBS)

- Update 1: LB = 524264, VBS = 8, VBO = 1
- VBO < VE-ID for VE-ID = 20 is true
- VE-ID < (VBO+VBS) for VE-ID = 20 is false.
- PE-2 cannot choose a label from this block.
- Update 2: LB = 524256, VBS = 8, VBO = 17
- VBO < VE-ID for VE-ID = 20 is true
- VE-ID < (VBO+VBS) for VE-ID = 20 is true.
- PE-2 chooses label 524256 + 20 - 17 = 524259 (LB + VEID - VBO)

The egress label chosen is verified by examining the egress label toward PE-1 (192.0.2.1) on PE-2.

```
*A:PE-2# show service id 2 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
21:4294967292    BgpVpls  192.0.2.1       Up      Up        524254    524259
23:4294967293    BgpVpls  192.0.2.3       Up      Up        524256    524259
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
```

PE-3 has a VE-ID of 3. Applying the condition VBO < VE-ID < (VBO+VBS)

- Update 1: LB = 524264, VBS = 8, VBO = 1
- VBO < VE-ID for VE-ID = 3 is true
- VE-ID < (VBO+VBS) for VE-ID = 3 is true.
- PE-3 chooses label 524264 + 3 - 1 = 524266 (LB + VEID - VBO)
- Update 2: LB = 524256, VBS = 8, VBO = 17
- VBO < VE-ID for VE-ID = 3 is false
- VE-ID < (VBO+VBS) for VE-ID = 3 is true.
- PE-3 cannot choose a label from this block.

The egress label chosen is verified by examining the egress label toward PE-1 (192.0.2.1) on PE-3.

```
*A:PE-3# show service id 2 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
31:4294967292    BgpVpls  192.0.2.1       Up      Up        524264    524266
```

```
32:4294967291    BgpVpls  192.0.2.2       Up      Up        524259    524256
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
```

To illustrate the allocation of label blocks by a PE, against the actual use of the same labels, consider the following. When BGP updates from each PE signal the multiplexer labels in blocks of eight, the allocated label values are added to the in-use pool. First check what label range can be allocated dynamically.

```
*A:PE-1# show router mpls-labels label-range

===============================================================================
Label Ranges
===============================================================================
Label Type      Start Label End Label   Aging       Available   Total
-------------------------------------------------------------------------------
Static          32          18431       -           18400       18400
Dynamic         18432       524287      0           505824      505856
    Seg-Route   0           0           -           0           0
===============================================================================
```

Verify which labels in the dynamic range are in use. The label pool of PE-1 can be verified as per the following output which shows labels used along with the associated protocol:

```
*A:PE-1# show router mpls-labels label 18432 524287 in-use

=================================================================
MPLS Labels from 18432 to 524287 (In-use)
=================================================================
Label               Label Type          Label Owner
-----------------------------------------------------------------
524256              dynamic             BGP
524257              dynamic             BGP
524258              dynamic             BGP
524259              dynamic             BGP
524260              dynamic             BGP
524261              dynamic             BGP
524262              dynamic             BGP
524263              dynamic             BGP
524264              dynamic             BGP
524265              dynamic             BGP
524266              dynamic             BGP
524267              dynamic             BGP
524268              dynamic             BGP
524269              dynamic             BGP
524270              dynamic             BGP
524271              dynamic             BGP
524272              dynamic             BGP
524273              dynamic             BGP
524274              dynamic             BGP
524275              dynamic             BGP
524276              dynamic             BGP
524277              dynamic             BGP
524278              dynamic             BGP
524279              dynamic             BGP
524280              dynamic             RSVP
524281              dynamic             RSVP
524282              dynamic             ILDP
524283              dynamic             ILDP
524284              dynamic             ILDP
524285              dynamic             ILDP
```

```
524286              dynamic             ILDP
524287              dynamic             ILDP
-----------------------------------------------------------------
In-use labels (Owner: All) in specified range   : 32
In-use labels in entire range                   : 32
=================================================================
```

This shows that 24 labels have been allocated for use by BGP. Of this number, 16 labels have been allocated for use by PEs within VPLS 2 to communicate with PE-1, the blocks with label base 524256 and with label base 524264.

There are only two neighboring PEs within this VPLS instance, so only two labels will ever be used in the data plane for traffic destined to PE-1. These are 524259 and 524266. The remaining labels have no PE with the associated VE-ID that can use them.

Once again, this case emphasizes that to reduce label wastage, contiguous VE-IDs in the range (N..N+7) per VPLS should be chosen, where N>0.

Conclusion

BGP-VPLS allows the delivery of Layer 2 VPN services to customers where BGP is commonly used. The examples presented in this chapter show the configuration of BGP-VPLS together with the associated show outputs which can be used for verification and troubleshooting.

Black-hole MAC for EVPN Loop Protection

This chapter provides information about Black-hole MAC for EVPN Loop Protection.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written based on SR OS Release 15.0.R4, but the CLI in the current edition corresponds to SR OS Release 21.2.R2. Black-hole MAC for EVPN loop protection is supported in SR OS Release 15.0.R1, and later.

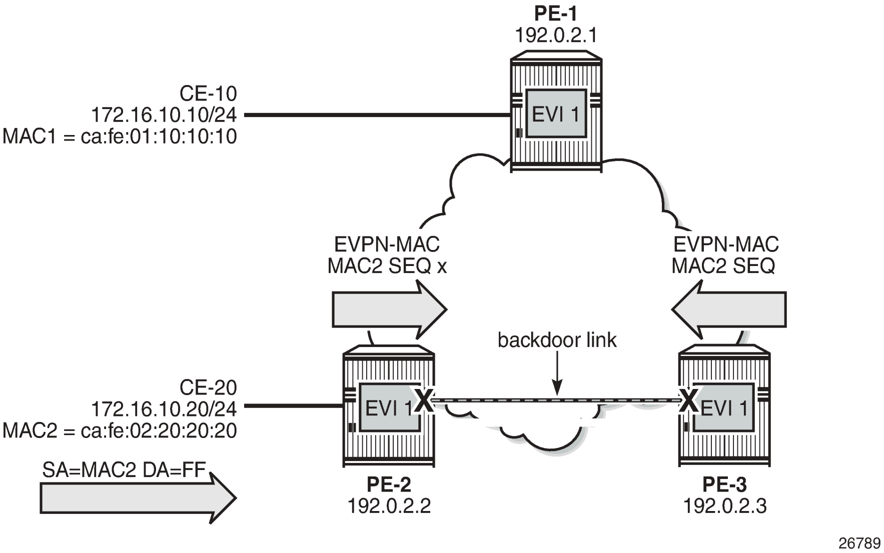
Chapters Auto-Learn MAC Protect in EVPN and Conditional Static Black-Hole MAC in EVPN are prerequisite reading.

Overview

Service providers are migrating VPLS networks to EVPN and require the same or better loop protection mechanisms, such as **mac-move** or **auto-learn-mac-protect** (ALMP). Chapter Auto-Learn MAC Protect in EVPN describes how traffic is protected in "static" networks, where the CEs do not move to a different port or PE, and MAC addresses are always learned first on the correct SAP/SDP-bindings. However, ALMP does not provide a loop protection solution in EVPN networks that require mobility and ALMP has issues with all-active multi-homing. Since mobility and all-active multi-homing are two of the key advantages of EVPN compared to VPLS, an alternate loop protection mechanism is required. This chapter describes an example for the black-hole based loop protection solution, based on *draft-snr-bess-evpn-loop-protect*.

Figure 33: Black-hole MAC for EVPN loop protection shows a topology using black-hole MAC for EVPN loop protection.

Figure 33: Black-hole MAC for EVPN loop protection



VPLS 1 with EVI 1 is configured on all PEs. A backdoor link exists between PE-2 and PE-3 (in this case, caused by misconfiguration: additional SAPs are configured in VPLS 1). When CE-20 sends Broadcast, Unknown unicast, or Multicast (BUM) traffic, its source address MAC2 is learned by PE-2, which sends an EVPN-MAC route for MAC2 to its BGP peers. PE-2 floods the frame to its EVPN-MPLS destinations (PE-1 and PE-3) as well as its local SAPs (including the backdoor link to PE-3).

PE-3 receives the EVPN-MAC route from PE-2, but due to the backdoor link, it also learns MAC2 on its local SAP. Following the MAC mobility procedures, PE-3 advertises MAC2 with a higher sequence number to its BGP peers. PE-3 floods the frame to its EVPN-MPLS destinations and to its local SAPs.

Note:
The preceding simplified description assumes that PE-3 receives the EVPN-MAC route prior to learning MAC2 from the backdoor link, which may or may not be the case. Regardless of how MAC2 is learned, the MAC duplication procedures are invoked.

PE-2 and PE-3 keep learning and advertising MAC2 until the configured number of MAC moves (**num-moves**) has been reached. Then, MAC2 is detected as duplicate and will not be advertised again until the **retry** interval has expired.

If the **mac-duplication black-hole-dup-mac** option is configured, MAC2 will be added to the FDB as black-hole MAC, so traffic with MAC DA = MAC2 will be discarded. Also, MAC addresses assigned to a black-hole destination are considered as protected, so traffic with MAC SA = MAC2 will not be forwarded due to one of the following reasons:

- When the SAPs/SDP-bindings or BGP-EVPN MPLS/VXLAN destinations are configured with **restrict-protected-source discard-frame**, the frames are discarded before any MAC SA is learned or the MAC DA is looked up.
- When the SAP/SDP-binding is configured with **restrict-protected-source**, an incoming frame with MAC SA = black-hole MAC causes the system to bring down the corresponding SAP/SDP-binding.

Assuming PE-3 detects MAC2 as duplicate and installs it as black-hole MAC, PE-3 will discard the broadcast frames with MAC SA = MAC2, so the loop is broken, whereas the legitimate traffic between CE-10 and CE-20 is allowed (assuming PE-2 does not black-hole MAC2).

Black-hole MAC duplication is enabled with the **black-hole-dup-mac** keyword in the **mac-duplication** context, as follows:

```
*A:PE-3>config>service>vpls>bgp-evpn# mac-duplication ?
  - mac-duplication

 [no] black-hole-dup* - Enable/disable BGP-EVPN black-hole duplicate MAC traffic
      detect          - Configure BGP EVPN Mac Duplication Detection
 [no] retry           - Configure BGP EVPN Mac Duplication Retry
```

```
# on PE-3:
configure
    service
        vpls "VPLS 1"
            bgp-evpn
                mac-duplication
                    black-hole-dup-mac
```

When enabled, the operation is as follows:

- Each node that learns a MAC address that has been advertised by a BGP peer will send an EVPN-MAC route for that MAC address with a higher sequence number. When the number of MAC moves exceeds the configured threshold (by default, five MAC moves in three minutes), the MAC address is detected as duplicate and no EVPN-MAC routes will be sent for that MAC address until the retry interval (default nine minutes) has elapsed.
- When MAC2 is detected as duplicate, the system will:
  - Add MAC2 to the duplicate MAC list
  - Add MAC2 in the FDB as protected MAC associated with a black-hole endpoint (type *EvpnD:P* and source identifier *black-hole*)
    - Incoming frames with MAC DA = MAC2 will be discarded based on a MAC lookup in the FDB.
    - MAC addresses assigned to a black-hole destination are protected and incoming frames with MAC SA = MAC2 will be discarded or the system will bring down the SAP/SDP-binding, depending on the **restrict-protected-src** setting on the SAP/SDP/EVPN endpoint.

The following output shows the FDB with black-hole MAC address ca:fe:02:20:20:20 (type EvpnD:P):

```
*A:PE-3# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          ca:fe:01:10:10:10 mpls:                   Evpn     04/28/21 09:59:12
                             192.0.2.1:524284
           ldp:65537
1          ca:fe:02:20:20:20 black-hole              EvpnD:P  04/28/21 09:59:12
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

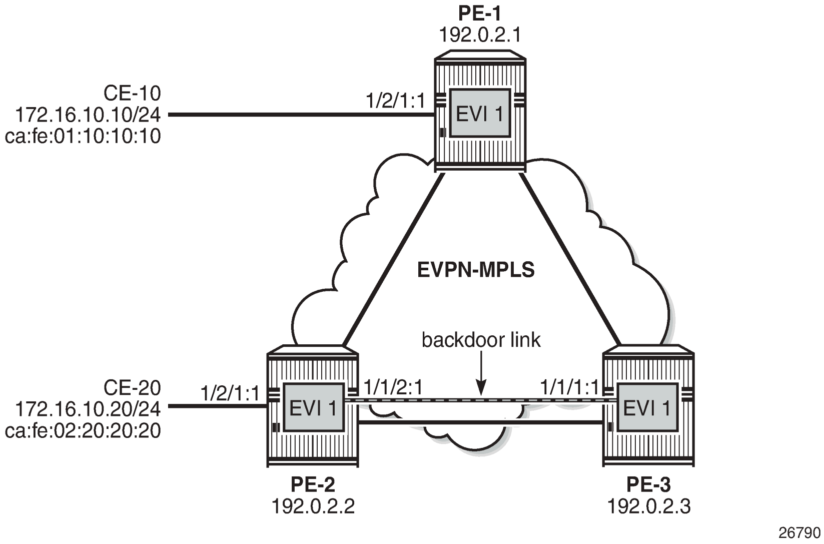
The duplicate MAC address will be removed from the FDB and the process will be restarted in the following cases:

- Retry interval events:
  - When the retry interval expires.
  - When the user configures **no retry** on the service that detected the duplicate MAC address.
- MAC relearning events:
  - When the remote PE withdraws the MAC address (due to aging or **clear service fdb**). Local attempts to clear a black-hole MAC (via **clear service fdb**) will fail because the type of the MAC entry is not "learned", but "EvpnD:P".
  - When configuring a local conditional static MAC address (CStatic:P) prevents the EvpnD:P entry for the same MAC address from being installed in the FDB as black-hole, if the SAP/SDP-binding where the MAC is configured is operationally up.
- CPM switchover event

Configuration

Figure 34: Example topology shows the example topology with three PEs and two CEs. A loop will occur when CE-20 sends Broadcast, Unknown unicast, or Multicast (BUM) traffic. Traffic between PE-2 and PE-3 will be sent over the regular router interfaces between the PEs, but also over the backdoor link (SAP 1/1/2:1 in VPLS 1 on PE-2 and SAP 1/1/1:1 in VPLS 1 on PE-3).

Figure 34: Example topology



The initial configuration includes:

- Cards, MDAs, ports
- Router interfaces
- IS-IS on all router interfaces (alternatively, OSPF can be used)
- LDP on all router interfaces

Enable black-hole MAC duplication detection in EVPN

BGP is configured for address family EVPN on all PEs with PE-3 as route reflector. The following is the BGP configuration on PE-3:

```
# on PE-3:
configure
    router Base
        autonomous-system 64500
        bgp
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family evpn
                cluster 192.0.2.3
                peer-as 64500
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.2
                exit
            exit
        exit
```

VPLS 1 is configured on all PEs with BGP-EVPN and MAC duplication enabled; on PE-2, as follows:

```
# on PE-2:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 1
                mac-duplication
                    detect num-moves 3 window 1
                    retry 2
                    black-hole-dup-mac
                exit
                mpls bgp 1
                    restrict-protected-src discard-frame
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            sap 1/1/2:1 create                          # backdoor link to PE-3
            exit
            sap 1/2/1:1 create                          # to CE-20
            exit
            no shutdown
        exit
```

To speed up MAC duplication detection, MAC duplication is detected after three MAC moves (default: five MAC moves). To shorten the retry interval, the time window is reduced to one minute (default: three minutes). When a MAC address has been detected as duplicated, the system removes the duplicate MAC entry after a retry interval of two minutes (default: nine minutes). The retry interval must be at least twice the time window for MAC duplication detection.

On the EVPN-MPLS endpoints, **restrict-protected-src discard-frame** must be configured. When MAC address ca:fe:02:20:20:20 is detected on PE-3 as a duplicate MAC address that is black-holed, the EVPN-MPLS endpoints on PE-3 should discard all frames with MAC SA ca:fe:02:20:20:20.

The configuration on the other PEs is similar; only the SAPs are different. VPLS 1 on PE-1 has SAP 1/2/1:1 to CE-10, but no SAP to a backdoor link; VPLS 1 on PE-3 has SAP 1/1/1:1 to the backdoor link to PE-2, but no SAP to a CE.

When CE-20 sends BUM traffic, its MAC SA ca:fe:02:20:20:20 is learned by PE-2 and advertised in EVPN-MAC routes. Because of the backdoor link to PE-3, PE-3 also learns MAC SA ca:fe:02:20:20:20 and advertises it to its BGP peers. The MAC-mobility sequence number is increased until the threshold of three MAC moves is reached. The following BGP EVPN-MAC route with sequence number 2 is sent by PE-2 to PE-3:

```
# on PE-2:
17 2021/04/28 09:59:11.599 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 89
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-MAC Len: 33 RD: 192.0.2.2:1 ESI: ESI-0, tag: 0, mac len: 48
                       mac: ca:fe:02:20:20:20, IP len: 0, IP: NULL, label1: 8388544
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:1
        bgp-tunnel-encap:MPLS
        mac-mobility:Seq:2
"
```

The FDB on PE-2 shows that MAC ca:fe:02:20:20:20 has been learned on the SAP toward CE-20 (but it could also have been learned on the backdoor SAP or even be black-holed), as follows:

```
*A:PE-2# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          ca:fe:01:10:10:10 mpls:                   Evpn     04/28/21 09:59:12
                             192.0.2.1:524284
           ldp:65537
1          ca:fe:02:20:20:20 sap:1/2/1:1             L/0      04/28/21 09:59:12
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The following FDB on PE-3 shows that MAC ca:fe:02:20:20:20 has been detected as a duplicate and protected MAC (type EvpnD:P) associated with a black-hole endpoint:

```
*A:PE-3# show service id 1 fdb mac ca:fe:02:20:20:20
```

```
===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          ca:fe:02:20:20:20 black-hole              EvpnD:P  04/28/21 09:59:12
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The following BGP-EVPN information for VPLS 1 on PE-3 shows the settings for MAC duplication detection, and the number of and list of detected duplicate MAC addresses:

```
*A:PE-3# show service id 1 bgp-evpn

===============================================================================
BGP EVPN Table
===============================================================================
MAC Advertisement  : Enabled           Unknown MAC Route  : Disabled
CFM MAC Advertise  : Disabled
Creation Origin    : manual
MAC Dup Detn Moves : 3                 MAC Dup Detn Window: 1
MAC Dup Detn Retry : 2                 Number of Dup MACs : 1
MAC Dup Detn BH    : Enabled
IP Route Advert    : Disabled
Sel Mcast Advert   : Disabled

EVI                : 1
Ing Rep Inc McastAd: Enabled
Accept IVPLS Flush : Disabled

-------------------------------------------------------------------------------
Detected Duplicate MAC Addresses             Time Detected
-------------------------------------------------------------------------------
ca:fe:02:20:20:20                            04/28/2021 09:59:12
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

The following message is logged in log "99" on PE-3 when VPLS 1 has detected duplicate MACs:

```
# on PE-3:
69 2021/04/28 10:04:40.266 UTC MINOR: SVCMGR #2331 Base
"VPLS Service 1 has MAC(s) detected as duplicates by EVPN mac-duplication detection."
```

MAC address ca:fe:02:20:20:20 remains in the FDB as duplicate and black-holed until the retry interval expires, as follows:

```
*A:PE-3>config>service>vpls>bgp-evpn>mac-duplication# retry ?
  - no retry
  - retry <minutes>

 <minutes>            : [2..60]
```

By default, the retry interval is nine minutes, but in this example, it is set to two minutes, which is the minimum value. The retry interval must be at least twice the time window for MAC duplication detection, which is by default three minutes, but reduced to one minute in this example. The following error is raised when attempting to configure a retry interval of two minutes for a detection time window of three minutes:

```
*A:PE-3>config>service>vpls>bgp-evpn>mac-duplication# retry 2
MINOR: SVCMGR #1003 Inconsistent value - mac-duplication detection retry time should be atleast
 twice that of detect window
```

After the retry interval expires, the MAC duplication is released.

Log "99" shows the following message when VPLS 1 no longer has duplicate MAC addresses:

```
# on PE-3:
70 2021/04/28 10:06:43.398 UTC MINOR: SVCMGR #2332 Base
"VPLS Service 1 no longer has MAC(s) detected as duplicates by EVPN mac-duplication detection."
```

MAC address ca:fe:02:20:20:20 remains in the FDB with type Evpn instead of EvpnD:P. BGP routes only disappear after a withdraw message has been received, whereas locally learned MAC addresses are flushed.

```
*A:PE-3# show service id 1 fdb mac ca:fe:02:20:20:20

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          ca:fe:02:20:20:20 mpls:                   Evpn     04/28/21 10:06:43
                             192.0.2.2:524284
           ldp:65538
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

Clear commands

The following FDB entry on PE-3 of type EvpnD:P cannot be cleared with a normal FDB **clear** command:

```
*A:PE-3# show service id 1 fdb mac ca:fe:02:20:20:20

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          ca:fe:02:20:20:20 black-hole              EvpnD:P  04/28/21 10:07:52
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The following error is raised when attempting to clear this FDB entry:

```
*A:PE-3# clear service id 1 fdb mac ca:fe:02:20:20:20
MAJOR: LOG #1202 Cannot perform clear operation - Entry is not of learned type
```

Log "99" shows the following message:

```
72 2021/04/28 10:08:17.960 UTC INDETERMINATE: LOGGER #2010 Base Clear SVCMGR
"Clear function clearSvcIdFdbMac has been run with parameters: svc-id="1" mac=
"ca:fe:02:20:20:20".  The completion result is: failure.  Additional error text, if any, is:
 Entry is not of learned type"
```

The following **clear** command releases the MAC duplication from the entry in the FDB, but it does not remove the entry from the FDB if it was learned from EVPN. The type is changed from EvpnD:P to Evpn.

```
*A:PE-3# clear service id 1 evpn mac-dup-detect ca:fe:02:20:20:20
```

```
*A:PE-3# show service id 1 fdb mac ca:fe:02:20:20:20

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          ca:fe:02:20:20:20 mpls:                   Evpn     04/28/21 10:09:50
                             192.0.2.2:524284
           ldp:65538
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

Instead of clearing the MAC duplication state for one specific MAC address, all duplicate MAC addresses can be cleared by the following command:

```
*A:PE-3# clear service id 1 evpn mac-dup-detect all
```

When the MAC duplication is released, VPLS 1 no longer has duplicate MAC addresses detected, as follows:

```
*A:PE-3# show service id 1 bgp-evpn | match "Detected" pre-lines 2 post-lines 5

-------------------------------------------------------------------------------
Detected Duplicate MAC Addresses             Time Detected
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
```

Log "99" shows the following messages related to the **clear** commands:

```
76 2021/04/28 10:10:13.078 UTC INDETERMINATE: LOGGER #2010 Base Clear SVCMGR
"Clear function cliClearSvcIdEvpnDupDetMacAll has been run with parameters: svc-id="1".  The
 completion result is: success.  Additional error text, if any, is: "

75 2021/04/28 10:09:49.947 UTC INDETERMINATE: LOGGER #2010 Base Clear SVCMGR
"Clear function cliClearSvcIdEvpnDupDetMac has been run with parameters: svc-id="1"mac=
"ca:fe:02:20:20:20".  The completion result is: success.  Additional error text, if any, is: "
```

Restrict Protected Source option

By default, the frames with MAC SA or DA equal to the duplicate MAC address are discarded, but the SAP/SDP-binding where the frame enters the VPLS remains operationally up. With the **restrict-protected-src** option, the system will bring the SAP/SDP-binding down where the frame with duplicate source MAC enters. The configuration on PE-2 and PE-3 is modified with **restrict-protected-src** on the SAP to the backdoor link, as follows:

```
# on PE-2:
configure
    service
        vpls "VPLS 1"
            sap 1/1/2:1
                restrict-protected-src

# on PE-3:
configure
    service
        vpls "VPLS 1"
            sap 1/1/1:1
                restrict-protected-src
```

When CE-20 sends BUM traffic, PE-3 detects MAC ca:fe:02:20:20:20 as duplicate. Log "99" shows that a duplicate MAC address has been detected, that protected MAC address ca:fe:02:20:20:20 has been received on SAP 1/1/1:1 in VPLS 1, and that the status of SAP 1/1/1:1 in VPLS 1 is changed to operationally down, with flag *RxProtSrcMac* indicating that a protected source MAC has been received.

```
80 2021/04/28 10:11:40.885 UTC MINOR: SVCMGR #2203 Base
"Status of SAP 1/1/1:1 in service 1 (customer 1) changed to admin=up oper=down flags=RxProtSrc
Mac "

79 2021/04/28 10:11:40.885 UTC MINOR: SVCMGR #2208 Base
"Protected MAC ca:fe:02:20:20:20 received on SAP 1/1/1:1 in service 1. The SAP will be
 disabled."

78 2021/04/28 10:11:39.886 UTC MINOR: SVCMGR #2331 Base
"VPLS Service 1 has MAC(s) detected as duplicates by EVPN mac-duplication detection."
```

The following shows that SAP 1/1/1:1 in VPLS 1 on PE-3 is operationally down with flag RxProtSrcMac:

```
*A:PE-3# show service id 1 sap 1/1/1:1

===============================================================================
Service Access Points(SAP)
===============================================================================
Service Id         : 1
SAP                : 1/1/1:1                  Encap             : q-tag
Description        : (Not Specified)
Admin State        : Up                       Oper State        : Down
Flags              : RxProtSrcMac
Multi Svc Site     : None
Last Status Change : 04/28/2021 10:11:41
Last Mgmt Change   : 04/28/2021 10:11:19
===============================================================================
```

The only way to re-enable the SAP is to disable and enable the SAP, as follows:

```
*A:PE-3# configure service vpls 1 sap 1/1/1:1 shutdown
```

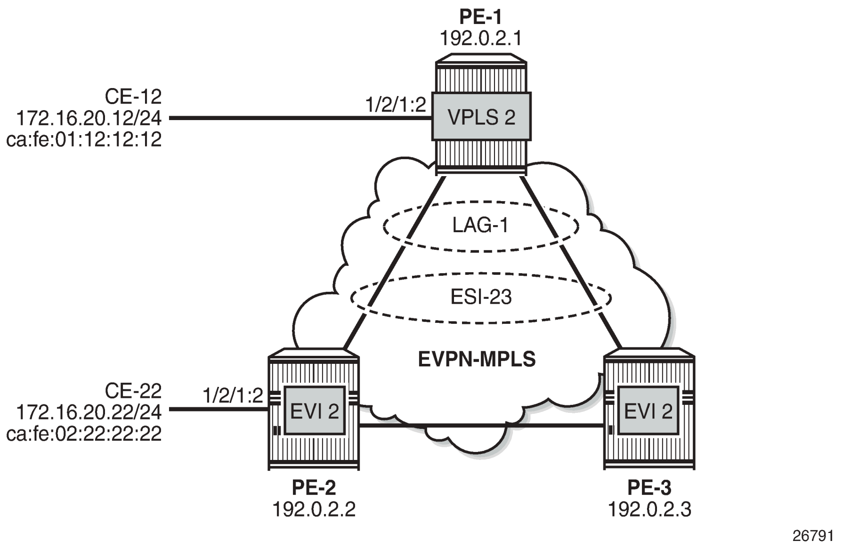
```
*A:PE-3# configure service vpls 1 sap 1/1/1:1 no shutdown
```

```
*A:PE-3# show service id 1 sap

===============================================================================
SAP(Summary), Service 1
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
1/1/1:1                         1          1     none    1     none   Up   Up
-------------------------------------------------------------------------------
Number of SAPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

Black-hole MAC duplication in all-active multi-homing

Figure 35: Example topology with all-active multi-homing shows the example topology with all-active multi-homing.

Figure 35: Example topology with all-active multi-homing



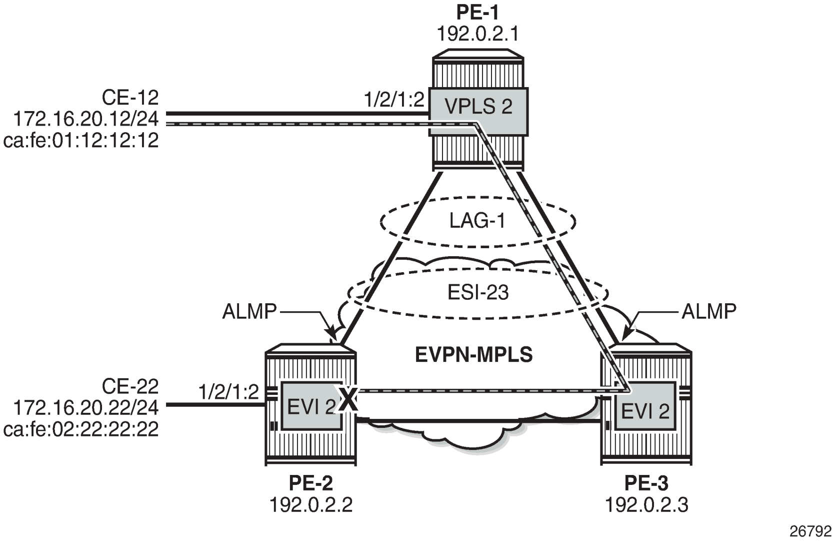
In this topology, the backdoor link is removed. On PE-1, VPLS 2 is configured without EVPN; on PE-2 and PE-3, VPLS 2 is configured with EVPN-MPLS. LAG 1 is configured on the PEs and Ethernet Segment (ES) ESI-23 is created on PE-2 and PE-3, as follows:

```
# on PE-2, PE-3:
configure
    service
```

```
        system
            bgp-evpn
                ethernet-segment "ESI-23" create
                    esi 01:00:00:00:00:23:00:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 1
                    no shutdown
                exit
```

The reason why black-hole MAC duplication should be configured instead of ALMP is the following. When ALMP is configured on SAP lag-1:2 on PE-2 and PE-3, MAC address ca:fe:01:12:12:12 of CE-12 is learned and protected on the SAP on both PEs. Traffic sent from CE-12 to CE-22 that is hashed over the direct link between PE-1 and PE-2 will reach its destination. Traffic that is hashed over the link between PE-1 and PE-3 will be forwarded by PE-3 to PE-2, but PE-2 will drop the traffic because it contains a MAC SA that is protected locally, as shown in Figure 36: Traffic dropped when ALMP is configured in all-active multi-homing.

Figure 36: Traffic dropped when ALMP is configured in all-active multi-homing



When black-hole MAC duplication is configured instead of ALMP, traffic hashed on the link to PE-3 is forwarded to PE-2 and to CE-22. This is because MAC duplication is ES-aware and the same MAC seen on the same ES in two different PEs will never be detected as duplicate.

The configuration of VPLS 2 in PE-2 is as follows:

```
# on PE-2:
configure
    service
        vpls 2 name "VPLS 2" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 2
                mac-duplication
```

```
                    black-hole-dup-mac
                exit
                mpls bgp 1
                    restrict-protected-src discard-frame
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/2/1:2 create
                no shutdown
            exit
            sap lag-1:2 create
                no shutdown
            exit
            no shutdown
```

The configuration of VPLS 2 on PE-3 is similar.

Conclusion

Black-hole MAC for EVPN MAC duplication protects EVPN services against customer-created backdoors or loops, while supporting MAC mobility and all-active multi-homing.

Conditional Static Black-Hole MAC in EVPN

This chapter provides information about Conditional Static Black-Hole MAC in EVPN.

Topics in this chapter include:

Applicability

This chapter was initially written for SR OS Release 14.0.R6, but the CLI in the current edition is based on SR OS Release 21.2.R1. Conditional static black-hole MAC is supported on EVPN services only, including EVPN-VXLAN and EVPN-MPLS, in SR OS Release 14.0.R1, and later.

Overview

A static black-hole MAC address is a local FDB record associated with a black-hole instead of a SAP or SDP-binding. Black-hole MAC addresses offer a scalable way to filter frames in the data plane based on MAC DA or SA, regardless of how the frame is arriving in the system. Black-hole MAC addresses can be configured in EVPN in the following ways:

- Static configured black-hole MAC address
- Anti-spoof MAC address in proxy Address Resolution Protocol/Neighbor Discovery (proxy-ARP/ND)
- MAC-duplication black-hole (supported in SR OS Release 15.0.R1, and later), see chapter Black-hole MAC for EVPN Loop Protection

When a specific MAC address is configured as a static black-hole MAC address, all frames with MAC DA equal to this black-hole MAC address will be dropped. Also, black-hole MAC addresses are treated as protected MAC addresses, which allows filtering on MAC SA; see chapter Auto-Learn MAC Protect in EVPN.

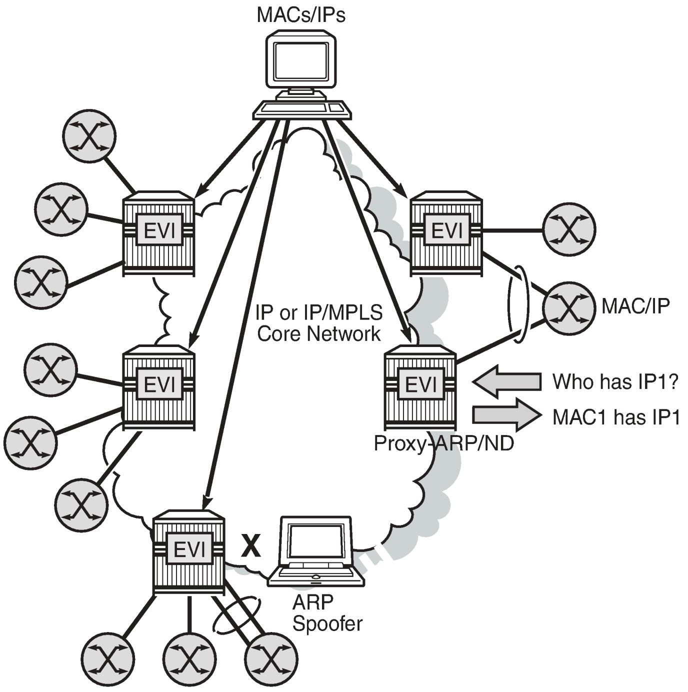
The default behavior on the SAP/SDP-bindings is Restricted Protected Source Discard Frame (RPS-DF). Therefore, all frames with MAC SA equal to the black-hole MAC address will, by default, be dropped on the SAP/SDP-binding where the frames enter the service. Instead of dropping the frames, the entire SAP/SDP-binding can be brought operationally down, if the SAP/SDP-binding is explicitly configured with Restricted Protected Source (RPS) without any parameter. The SAP/SDP-binding can only be brought up manually by disabling (shutdown) and re-enabling (no shutdown) the SAP/SDP-binding. On the EVPN endpoints between PEs, it is possible to configure RPS-DF, not RPS. When configured, the EVPN endpoint will drop frames with MAC SA equal to the black-hole MAC address.

Black-hole MAC addresses can be used as an alternative to MAC filters, which simplifies the deployment of proxy-ARP/ND with anti-spoof MAC addresses. ARP/ND spoofing is a technique whereby an attacker sends fake ARP/ND messages to a broadcast domain. Generally, the aim is to get the routers in the broadcast domain to associate the attacker's MAC address with the IP address of another host, causing any traffic destined to that IP address to be sent to the attacker instead. To prevent this from happening, a proxy-ARP/ND with duplicate IP detection monitors the number of times the MAC changes for an offending IP address. When a certain number of MAC moves are detected in a defined period, the system flags the proxy-ARP entry as duplicate for a defined hold time and an alarm is sent to log 99.

Chapter EVPN for MPLS Tunnels describes the proxy-ARP/ND configuration with the option to define an anti-spoof MAC (AS-MAC) address for EVPN-MPLS networks using MAC filters, including some recommended settings. The AS-MAC address will be advertised with the duplicate IP address in gratuitous ARP (GARP) and ARP replies to all CEs in the EVPN (in the case of proxy-ND, unsolicited Neighbor Advertisement messages are sent instead of GARP messages).

ARP/ND broadcast traffic is a security issue for Internet eXchange Providers (IXPs) and service providers with large Layer 2 domains. In such networks, administrators try to avoid ARP/ND flooding. Figure 37: Proxy-ARP/ND and ARP spoofing shows the proxy-ARP/ND feature where local ARP/ND requests are responded by the system on behalf of the IP interface owners.

Figure 37: Proxy-ARP/ND and ARP spoofing



26244

EVPN can suppress ARP/ND flooding within an EVPN service if all the attached hosts advertise their presence. Therefore, EVPN is preferred in IXPs to mitigate and even eliminate the ARP/ND flooding issue. The proxy-ARP/ND agent responds to local ARP/ND requests using a proxy-ARP/ND table per service. This table is populated by EVPN entries (MAC-IP pairs), static entries configured in the service, and dynamic entries snooped from ARP/GARP/ND messages sent by the ISP routers. The static entries and snooped dynamic entries are also advertised in EVPN-MAC routes.

As well as the proxy-ARP/ND, SR OS supports an anti-spoofing mechanism that can detect and block an ARP spoofing attack or a misconfigured duplicated IP address. When using MAC filters, the same anti-spoof-mac option must be configured in all the PEs and this filter may be configured on all the PE SAPs/

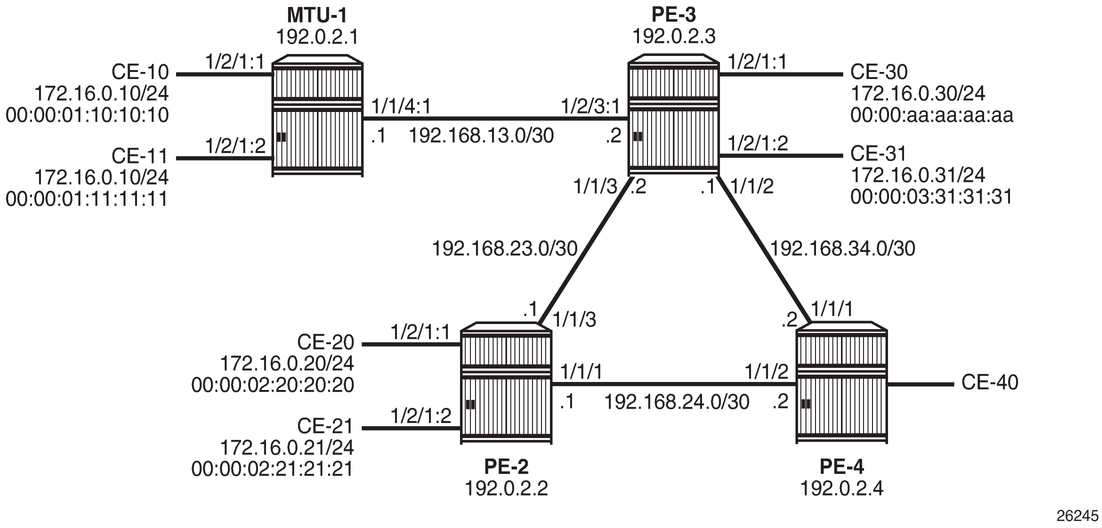
SDP-bindings to discard all the frames with MAC DA equal to the anti-spoof MAC address. This requires a lot of configuration and is prone to configuration errors.

Conditional static black-hole MAC addresses can be configured for the anti-spoof MAC address so that frames with MAC DA equal to the anti-spoof MAC address can be discarded based on a MAC address lookup in the FDB, as opposed to a MAC filter entry. Less configuration is required and this simplifies the deployment of proxy-ARP/ND with AS-MAC. The configuration example in this chapter includes proxy-ARP, but the behavior is similar for proxy-ND.

Configuration

Figure 38: Example topology shows the example topology. Traffic will be sent between the CEs and may be dropped in the PEs if the MAC DA or MAC SA matches a black-hole MAC address. IP address 172.16.0.10/24 is duplicate (CE-10 and CE-11).

Figure 38: Example topology



The initial configuration on the nodes includes:

- Cards, MDAs, ports
- Router interfaces
- IS-IS between PEs
- LDP between PEs

BGP is configured between the PEs for address family EVPN with PE-2 as route reflector (RR). Instead of an RR, a full mesh can also be configured between the PEs. The BGP configuration on PE-2 is as follows:

```
# on RR PE-2:
configure
    router Base
        autonomous-system 64500
        bgp
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family evpn
                cluster 1.1.1.1
                peer-as 64500
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
            exit
        exit
```

VPLS 1 is configured on all PEs and on MTU-1 (MTU-1's VPLS 1 is connected to PE-3 by a SAP). The VPLS configuration on the PEs includes EVPN-MPLS, as follows:

```
# on PE-3:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    ingress-replication-bum-label
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/2/1:1 create
                no shutdown
            exit
            sap 1/2/3:1 create
                no shutdown
            exit
            no shutdown
        exit
```

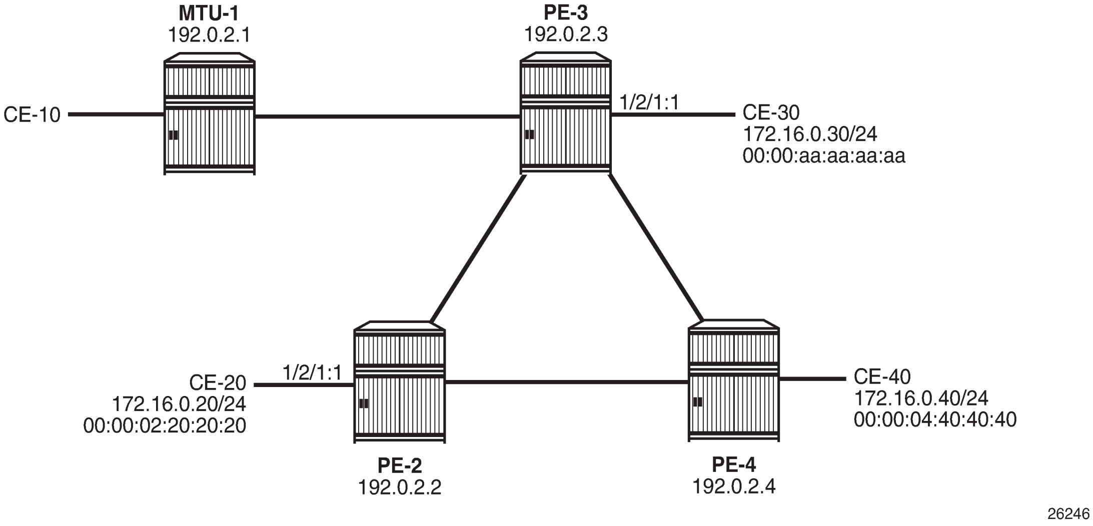
Conditional static black-hole MAC

Conditional static black-hole MAC address is an extension to the conditional static MAC address, but with the **black-hole** keyword. It is a scalable way to filter MAC DA or SA in the data plane, regardless of how the frame is arriving at the system (SAP/SDP-bindings or EVPN termination endpoints).

When the static black-hole MAC is added to the FDB, all Ethernet frames with MAC DA equal to the black-hole MAC are dropped. Filtering based on the MAC SA is explained in the next section: Conditional static black-hole MAC in combination with restrict protected source.

Figure 39: Conditional static black-hole MAC shows the example setup with conditional static black-hole MAC 00:00:aa:aa:aa:aa.

Figure 39: Conditional static black-hole MAC



When no conditional static black-hole MAC is configured, CE-30 can receive and send traffic from and to the other CEs; for instance, from and toward CE-20, as follows:

```
*A:PE-2# ping router 10 172.16.0.30
PING 172.16.0.30 56 data bytes
64 bytes from 172.16.0.30: icmp_seq=1 ttl=64 time=0.836ms.
64 bytes from 172.16.0.30: icmp_seq=2 ttl=64 time=0.841ms.
---snip---

*A:PE-3# ping router 10 172.16.0.20
PING 172.16.0.20 56 data bytes
64 bytes from 172.16.0.20: icmp_seq=1 ttl=64 time=3.69ms.
64 bytes from 172.16.0.20: icmp_seq=2 ttl=64 time=0.814ms.
---snip---
```

In this example, CE-20 and CE-30 correspond to VPRN 10 configured on PE-2 and PE-3 (using a hairpin to loop the traffic back to the PE).

Conditional static black-hole MAC 00:00:aa:aa:aa:aa (which corresponds to the MAC address of CE-30) is configured in VPLS 1 on PE-3 as follows:

```
# on PE-3:
configure
    service
        vpls "VPLS 1"
            static-mac
                mac 00:00:aa:aa:aa:aa create black-hole
```

The black-hole MAC is added as a conditional static (CStatic) MAC that is protected (P), as follows:

```
*A:PE-3# show service id 1 fdb mac 00:00:aa:aa:aa:aa

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
```

```
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:aa:aa:aa:aa black-hole              CStatic: 03/24/21 15:46:33
                                                     P
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The source identifier is black-hole and it is applicable to frames that enter the VPLS on this node, regardless of how they enter the VPLS (SAP, SDP-binding, or EVPN endpoint).

The conditional static black-hole MAC is advertised to the BGP peers in a BGP-EVPN MAC route with the sticky/static bit set, as follows:

```
# on PE-3:
9 2021/03/24 15:46:32.883 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 89
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-MAC Len: 33 RD: 192.0.2.3:1 ESI: ESI-0, tag: 0, mac len: 48
                       mac: 00:00:aa:aa:aa:aa, IP len: 0, IP: NULL, label1: 8388544
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:1
        bgp-tunnel-encap:MPLS
        mac-mobility:Seq:0/Static
"
```

The MAC route is added to the FDB on the other PEs as a static (S) and protected (P) MAC; for example, on PE-2, as follows:

```
*A:PE-2# show service id 1 fdb mac 00:00:aa:aa:aa:aa

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:aa:aa:aa:aa mpls:                   EvpnS:P  03/24/21 15:46:33
                             192.0.2.3:524284
           ldp:65537
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

When CE-20 sends an ICMP request to CE-30, the MAC DA 00:00:aa:aa:aa:aa is black-holed on PE-3, and no ICMP request succeeds, as follows:

```
*A:PE-2# ping router 10 172.16.0.30
PING 172.16.0.30 56 data bytes
Request timed out. icmp_seq=1.
Request timed out. icmp_seq=2.
Request timed out. icmp_seq=3.
Request timed out. icmp_seq=4.
Request timed out. icmp_seq=5.

---- 172.16.0.30 PING Statistics ----
5 packets transmitted, 0 packets received, 100% packet loss
```

The port statistics show that the traffic was sent from PE-2 to PE-3, where it entered on port 1/1/3, then got discarded. To verify this, the port statistics are cleared on PE-2 and PE-3, then 1000 ICMP packets are sent from CE-20, as follows:

```
*A:PE-2# clear port 1/[1..2]/[1..4] statistics
*A:PE-3# clear port 1/[1..2]/[1..4] statistics
*A:PE-2# ping router 10 172.16.0.30 rapid count 1000
---snip---
1000 packets transmitted, 0 packets received, 100% packet loss
```

The 1000 packets are received at SAP 1/2/1:1 on PE-2, as follows:

```
*A:PE-2# show port 1/2/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/2/1                                  1000                          106000
                                          0                               0
===============================================================================
```

These packets are forwarded to port 1/1/3 toward PE-3, as follows:

```
*A:PE-2# show port 1/1/3 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/3                                    13                            1306
                                       1013                          125254
===============================================================================
```

On the interfaces between the PEs, other packets are sent besides the ICMP requests, such as IS-IS messages; therefore, the number of packets is slightly greater than 1000.

On PE-3, these packets are received on port 1/1/3, as follows:

```
*A:PE-3# show port 1/1/3 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/3                                  1024                          126444
                                         24                            2351
===============================================================================
```

The FDB entry for this MAC DA is black-holed and no traffic is received on SAP 1/2/1:1 toward CE-30; therefore, the statistics for port 1/2/1 are empty and nothing is displayed, as follows:

```
*A:PE-3# show port 1/2/1 statistics
*A:PE-3#
```

It is possible to configure the black-hole MAC address on a different PE; for example, on PE-4 instead of PE-3. The conditional static black-hole MAC address configuration in VPLS 1 on PE-3 is removed, as follows:

```
# on PE-3:
configure
    service
        vpls "VPLS 1"
            static-mac
                no mac 00:00:aa:aa:aa:aa
```

The conditional static black-hole MAC is configured on PE-4 instead, as follows:

```
# on PE-4:
configure
    service
        vpls "VPLS 1"
            static-mac
                mac 00:00:aa:aa:aa:aa create black-hole
```

PE-4 sends EVPN-MAC updates to its peers. PE-2 learns that all traffic with MAC DA 00:00:aa:aa:aa:aa should be redirected to PE-4, as shown in the FDB on PE-2:

```
*A:PE-2# show service id 1 fdb mac 00:00:aa:aa:aa:aa

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:aa:aa:aa:aa mpls:                   EvpnS:P  03/24/21 15:51:14
                             192.0.2.4:524284
           ldp:65538
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The port statistics are cleared on all PEs and 1000 ICMP packets are sent from CE-20 to CE-30, as follows:

```
*A:PE-2# ping router 10 172.16.0.30 rapid count 1000
---snip---
1000 packets transmitted, 0 packets received, 100% packet loss
```

On PE-2, traffic is not forwarded on the direct link (port 1/1/3) toward PE-3, but redirected to PE-4 (port 1/1/1) instead, as follows:

```
*A:PE-2# show port 1/1/[1..3] statistics

===============================================================================
Port Statistics on Slot 1
```

```
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/1                                    15                            1464
                                       1014                          125360
===============================================================================

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/3                                    15                            1433
                                         16                            1589
===============================================================================
```

On PE-4, traffic is received on port 1/1/2, then discarded because the MAC DA equals the static black-hole MAC in the FDB, as follows. No traffic is forwarded to PE-3, where CE-30 is attached.

```
*A:PE-4# show port 1/[1..2]/[1..4] statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/1                                    22                            2192
                                         22                            2192
===============================================================================

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/2                                  1025                          126476
                                         24                            2351
===============================================================================
```

The configuration is restored with conditional static black-hole MAC in VPLS 1 on PE-3, not on PE-4, as follows:

```
# on PE-3:
configure
    service
        vpls "VPLS 1"
            static-mac
                mac 00:00:aa:aa:aa:aa create black-hole
```

```
# on PE-4:
configure
    service
        vpls "VPLS 1"
            static-mac
                no mac 00:00:aa:aa:aa:aa
```

Conditional static black-hole MAC in combination with restrict protected source

For Ethernet frames with MAC SA equal to the static black-hole MAC, the treatment is the same as for protected MACs (see chapter Auto-Learn MAC Protect in EVPN), but for conditional static black-hole MACs, ALMP need not be enabled on the SAP or SDP-binding:

- When a frame is received with MAC SA equal to the black-hole MAC, it is dropped, because RPS-DF is enabled on the SAP or SDP-binding, by default. RPS-DF need not be enabled explicitly. The default is **no restrict-protected-src**, which operates as RPS-DF. An error message is raised when the following command is entered:

```
*A:PE-3>config>service>vpls>sap# restrict-protected-src discard-frame
MINOR: SVCMGR #7888 Cannot be configured/enabled with EVPN
```

- When RPS is enabled instead of RPS-DF, the SAP or SDP-binding where the frame was received, with MAC SA equal to the black-hole MAC, is brought operationally down. The SAP or SDP-binding can be brought up manually by disabling (shutdown) and re-enabling (no shutdown) the SAP or SDP-binding. RPS is enabled on SAP 1/2/1:1 as follows:

```
# on PE-3:
configure
    service
        vpls "VPLS 1"
            sap 1/2/1:1
                restrict-protected-src
```

- Optionally, RPS-DF can be enabled on the EVPN-MPLS endpoint or EVPN-VXLAN endpoint. When enabled, the EVPN endpoint will discard frames with MAC SA equal to the black-hole MAC. RPS cannot be configured instead of RPS-DF on EVPN endpoints. It is not an option to bring the EVPN endpoint down when a frame is received with MAC SA equal to the static black-hole MAC. The commands to enable RPS-DF on the EVPN-MPLS endpoints and EVPN-VXLAN endpoints are as follows:

```
*A:PE-3>config>service>vpls>bgp-evpn>mpls# restrict-protected-src ?
  - no restrict-protected-src
  - restrict-protected-src discard-frame

 <discard-frame>      : keyword - discard frame and  trap on a protected MAC
```

```
*A:PE-3>config>service>vpls>vxlan$ restrict-protected-src ?
  - no restrict-protected-src
  - restrict-protected-src discard-frame
```

With the default configuration (RPS-DF on SAP/SDP-bindings), the behavior is as follows for conditional static black-hole MAC 00:00:aa:aa:aa:aa configured in VPLS 1 on PE-3. All traffic from CE-30 with MAC SA 00:00:aa:aa:aa:aa is black-holed on SAP 1/2/1:1 on PE-3, because the default behavior on SAP 1/2/1:1 is RPS-DF, and the frame is discarded. The packets are received on port 1/2/1 (SAP 1/2/1:1) and dropped. No packets are forwarded to port 1/1/3 toward PE-2 or any other port.

```
*A:PE-3# clear port 1/[1..2]/[1..4] statistics
*A:PE-3# ping router 10 172.16.0.20 rapid count 1000
---snip---
1000 packets transmitted, 0 packets received, 100% packet loss
*A:PE-3# show port 1/[1..2]/[1..4] statistics
```

```
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/2                                    17                            1732
                                         17                            1732
===============================================================================

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/3                                    20                            1995
                                         18                            1787
===============================================================================

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/2/1                                  1000                          106000
                                          0                               0
===============================================================================
```

If the static MAC is configured in VPLS 1 on PE-4 and not on PE-3, PE-3 will still discard the packets with MAC SA 00:00:aa:aa:aa:aa arriving on SAP 1/2/1:1, because it learned from the EVPN-MAC updates that MAC 00:00:aa:aa:aa:aa is a protected MAC on PE-4. Therefore, traffic with this MAC SA is not expected and not allowed on PE-3, as follows:

```
# on PE-3:
configure
    service
        vpls "VPLS 1"
            static-mac
                no mac 00:00:aa:aa:aa:aa
```

```
# on PE-3:
configure
    service
        vpls "VPLS 1"
            static-mac
                mac 00:00:aa:aa:aa:aa create black-hole
```

```
*A:PE-3# show service id 1 fdb mac 00:00:aa:aa:aa:aa

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:aa:aa:aa:aa mpls:                   EvpnS:P  03/24/21 15:54:54
                             192.0.2.4:524284
           ldp:65538
-------------------------------------------------------------------------------
```

```
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

```
*A:PE-3# ping router 10 172.16.0.20 rapid count 1000
---snip---
1000 packets transmitted, 0 packets received, 100% packet loss
```

```
*A:PE-3# show port 1/[1..2]/[1..4] statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/2                                    22                            2192
                                         22                            2192
===============================================================================

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/3                                    25                            2476
                                         24                            2351
===============================================================================

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/2/1                                  1000                          106000
                                          0                               0
===============================================================================
```

The configuration is restored as follows:

```
# on PE-3:
configure
    service
        vpls "VPLS 1"
            static-mac
                mac 00:00:aa:aa:aa:aa create black-hole
```

```
# on PE-4:
configure
    service
        vpls "VPLS 1"
            static-mac
                no mac 00:00:aa:aa:aa:aa
```

Optionally, RPS-DF can be configured on the EVPN-MPLS endpoints on the PEs, as follows:

```
# on PE-2, PE-3, PE-4:
configure
    service
```

```
        vpls "VPLS 1"
            bgp-evpn
                mpls bgp 1
                    restrict-protected-src discard-frame
```

When RPS-DF is configured on the EVPN-MPLS endpoints, frames with MAC SA 00:00:aa:aa:aa:aa can be discarded by the EVPN endpoints between the PEs. However, in this example this is not required, because any frame with MAC SA 00:00:aa:aa:aa:aa will be dropped by the local SAP before it can be forwarded to an EVPN endpoint.

It is possible to configure RPS without any parameters on SAP 1/2/1:1 on PE-3, as follows:

```
# on PE-3:
configure
    service
        vpls "VPLS 1"
            sap 1/2/1:1
                restrict-protected-src
```

When CE-30 sends traffic with MAC SA equal to a protected MAC address (black-hole or not), the entire SAP 1/2/1:1 will be brought operationally down, as follows:

```
*A:PE-3# ping router 10 172.16.0.20
PING 172.16.0.20 56 data bytes
Request timed out. icmp_seq=1.
Request timed out. icmp_seq=2.
---snip---
---- 172.16.0.20 PING Statistics ----
5 packets transmitted, 0 packets received, 100% packet loss
```

```
*A:PE-3# show service id 1 sap

===============================================================================
SAP(Summary), Service 1
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
1/2/1:1                         1          1     none    1     none   Up   Down
1/2/3:1                         1          1     none    1     none   Up   Up
-------------------------------------------------------------------------------
Number of SAPs : 2
-------------------------------------------------------------------------------
===============================================================================
```

The following information for SAP 1/2/1:1 in VPLS 1 shows that this SAP is operationally down because a protected source MAC address was received on this SAP (Flags: RxProtSrcMac), as follows:

```
*A:PE-3# show service id 1 sap 1/2/1:1

===============================================================================
Service Access Points(SAP)
===============================================================================
Service Id         : 1
SAP                : 1/2/1:1                  Encap             : q-tag
Description        : (Not Specified)
Admin State        : Up                       Oper State        : Down
Flags              : RxProtSrcMac
Multi Svc Site     : None
Last Status Change : 03/24/2021 15:54:06
Last Mgmt Change   : 03/24/2021 15:53:25
===============================================================================
```

Log 99 shows that a protected MAC was received on SAP 1/2/1:1 and the SAP went operationally down with flag RxProtSrcMac, as follows:

```
107 2021/03/24 15:54:05.661 UTC MINOR: SVCMGR #2208 Base
"Protected MAC 00:00:aa:aa:aa:aa received on SAP 1/2/1:1 in service 1. The SAP will be
 disabled."

108 2021/03/24 15:54:05.662 UTC MINOR: SVCMGR #2203 Base
"Status of SAP 1/2/1:1 in service 1 (customer 1) changed to admin=up oper=down flags=RxProtSrc
Mac "
```

The SAP can only be brought up manually by disabling and re-enabling the SAP, as follows:

```
*A:PE-3# configure service vpls "VPLS 1" sap 1/2/1:1 shutdown
*A:PE-3# configure service vpls "VPLS 1" sap 1/2/1:1 no shutdown
```

```
*A:PE-3# show service id 1 sap

===============================================================================
SAP(Summary), Service 1
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
1/2/1:1                         1          1     none    1     none   Up   Up
1/2/3:1                         1          1     none    1     none   Up   Up
-------------------------------------------------------------------------------
Number of SAPs : 2
-------------------------------------------------------------------------------
===============================================================================
```

The default behavior of SAP 1/2/1:1 is RPS-DF, which is configured by removing the RPS configuration, as follows:

```
# on PE-3:
configure
    service
        vpls "VPLS 1"
            sap 1/2/1:1
                no restrict-protected-src
```

The conditional static black-hole MAC configuration is removed as follows:

```
# on PE-3:
configure
    service
        vpls "VPLS 1"
            static-mac
                no mac 00:00:aa:aa:aa:aa
```

Black-hole MAC in services with proxy-ARP/ND

In this example, only proxy-ARP is shown, not proxy-ND. However, the configuration and procedures for proxy-ND would be equivalent.

First, the implementation of proxy-ARP and AS-MAC is described without static black-hole MAC addresses. MAC filters will be required to drop or redirect traffic, but these are not shown in the example. Configuring MAC filters and applying them on SAP/SDP-bindings is labor-intensive and can be error-prone. Afterward, the implementation with AS-MAC as static black-hole is described.

Services with proxy-ARP and AS-MAC - no static black-hole MAC

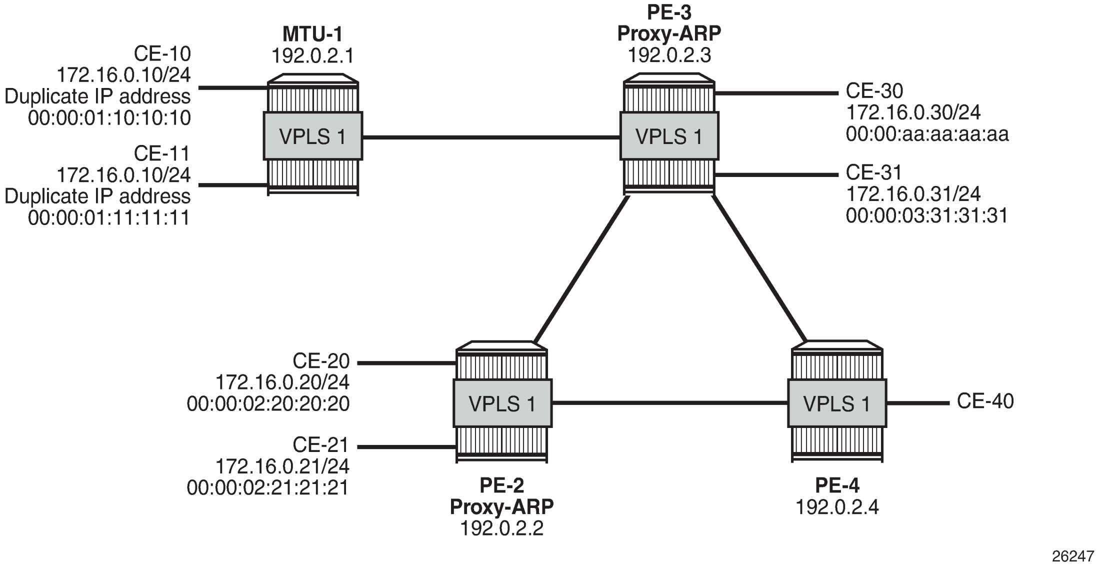
IP duplication works when the IP address moves between:

- Dynamic (learned on SAP) and EVPN
- EVPN and dynamic
- Dynamic and dynamic

The following example shows IP address moves from dynamic to dynamic between SAP 1/2/1:1 (to CE-10) and SAP 1/2/1:2 (to CE-11) in VPLS 1 on MTU-1. However, the duplicate IP address could have been in PE-3 and MTU-1 instead (EVPN or dynamic) and still the IP address would have been detected as duplicate.

Figure 40: VPLS 1 with proxy-ARP and AS-MAC shows the example setup with duplicate IP address 172.16.0.10/24 for CE-10 and CE-11. VPLS 1 is configured with proxy-ARP with duplicate IP detection in PE-2 and PE-3 (and possibly also in other PEs). MAC address 00:00:bb:bb:bb:bb is configured as AS-MAC, which will be used when a duplicate IP address has been detected.

Figure 40: VPLS 1 with proxy-ARP and AS-MAC



For IP duplication detection, the following parameters can be customized so that the system can react to particular conditions in the network. The syntax is as follows:

```
*A:PE-3>config>service>vpls>proxy-arp$ dup-detect ?
  - dup-detect [anti-spoof-mac <mac-address>] window <minutes> num-moves <count>
               hold-down <minutes|max>
  - dup-detect anti-spoof-mac <mac-address> window <minutes> num-moves <count>
               hold-down <minutes|max> [static-black-hole]

 <mac-address>        : xx-xx-xx-xx-xx-xx or xx:xx:xx:xx:xx:xx (hex chars)
 <minutes>            : [1..15] minutes - default:3
 <count>              : [3..10] - default:5
 <minutes|max>        : [2..60] default=9 | max - permanent hold
 <static-black-hole>  : keyword
```

In VPLS 1 on PE-3, a proxy-ARP with duplicate IP detection is configured, including an optional anti-spoof MAC (AS-MAC) 00:00:bb:bb:bb:bb for offending IP addresses, as follows:

```
#on PE-3:
configure
    service
        vpls "VPLS 1"
            proxy-arp
                dup-detect window 3 num-moves 3 hold-down max
                           anti-spoof-mac 00:00:bb:bb:bb:bb
                dynamic-arp-populate
                static 172.16.0.20 00:00:02:20:20:20
                no shutdown
            exit
```

The proxy-ARP table contains one static entry (for IP 172.16.0.20). In this case, dynamic ARP populate is enabled. Therefore, the proxy-ARP table will be updated with ARP entries for IP 172.16.0.10 and MAC 00:00:01:10:10:10 or MAC 00:00:01:11:11:11 for frames originating from CE-10 or CE-11.

When a duplicate IP is detected for IP 172.16.0.10 (after three changes of MAC for IP 172.16.0.10 in a period of three minutes), the corresponding ARP entry contains the duplicate IP address 172.16.0.10 and the AS-MAC 00:00:bb:bb:bb:bb and its type is duplicate (dup). Therefore, this ARP entry is always active until it is removed. Until now, this configuration does not include a static black-hole MAC, and this option is by default disabled. This configuration for duplicate IP detection can be used in combination with MAC filters. The configuration with static black-hole MAC is shown in the section Services with proxy-ARP and AS-MAC configured as static black-hole MAC.

The configured AS-MAC will be advertised in an EVPN-MAC route with the sticky/static bit set and without any IP address (because there is no IP duplication detected yet), as follows:

```
# on PE-3:
28 2021/03/24 16:01:30.827 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 124
    Flag: 0x90 Type: 14 Len: 79 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-MAC Len: 33 RD: 192.0.2.3:1 ESI: ESI-0, tag: 0, mac len: 48
                       mac: 02:17:ff:00:03:3a, IP len: 0, IP: NULL, label1: 8388544
        Type: EVPN-MAC Len: 33 RD: 192.0.2.3:1 ESI: ESI-0, tag: 0, mac len: 48
                       mac: 00:00:bb:bb:bb:bb, IP len: 0, IP: NULL, label1: 8388544
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
```

```
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:1
        bgp-tunnel-encap:MPLS
        mac-mobility:Seq:0/Static
"
```

Without the option static black-hole, the configured AS-MAC is not added to the local FDB, but this MAC address is treated as a local MAC. The FDB on PE-3 does not contain AS-MAC 00:00:bb:bb:bb:bb, as follows:

```
*A:PE-3# show service id 1 fdb mac 00:00:bb:bb:bb:bb

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

Debugging is enabled for proxy-ARP for IP address 172.16.0.10 in VPLS 1 on PE-3, as follows:

```
# on PE-3:
debug
    service
        id 1
            proxy-arp ip 172.16.0.10
```

When traffic is sent from CE-11 to CE-21, a dynamic ARP entry for IP address 172.16.0.10 and MAC 00:00:01:11:11:11 is added to the proxy-ARP table for VPLS 1 in PE-3, and an EVPN-MAC update is sent to the peer PEs, as follows:

```
# on PE-3:
35 2021/03/24 16:06:02.919 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 mac: 00:00:01:11:11:11 evpn advertise"

36 2021/03/24 16:06:02.919 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 type: Dyn mac: 00:00:01:11:11:11 Added"

37 2021/03/24 16:06:02.919 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 85
    Flag: 0x90 Type: 14 Len: 48 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-MAC Len: 37 RD: 192.0.2.3:1 ESI: ESI-0, tag: 0, mac len: 48
                   mac: 00:00:01:11:11:11, IP len: 4, IP: 172.16.0.10, label1: 8388544
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:1
        bgp-tunnel-encap:MPLS
"
```

There is no duplicate IP detected yet.

CE-10 and CE-11 have the same IP address for different MAC addresses. When CE-10 sends traffic to CE-20, the ARP entry for IP 172.16.0.10 changes MAC from 00:00:01:11:11:11 to 00:00:01:10:10:10, and an EVPN-MAC withdraw message is sent, as follows:

```
# on PE-3:
38 2021/03/24 16:06:07.137 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 mac: 00:00:01:11:11:11 evpn withdraw"

39 2021/03/24 16:06:07.137 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 Mac Change: 00:00:01:11:11:11->00:00:01:10:10:10 "

40 2021/03/24 16:06:07.137 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 46
    Flag: 0x90 Type: 15 Len: 42 Multiprotocol Unreachable NLRI:
        Address Family EVPN
        Type: EVPN-MAC Len: 37 RD: 192.0.2.3:1 ESI: ESI-0, tag: 0, mac len: 48
                   mac: 00:00:01:11:11:11, IP len: 4, IP: 172.16.0.10, label1: 0
"
```

When the MAC changes, the system sends an ARP request for confirmation of the old MAC 00:00:01:11:11:11 for IP 172.16.0.10, as follows:

```
# on PE-3:
41 2021/03/24 16:06:07.290 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 mac: 00:00:01:11:11:11 confirm"
```

When MAC 00:00:01:11:11:11 is confirmed, the MAC in the ARP entry is changed once again to 00:00:01:11:11:11 and another ARP request is sent asking to confirm MAC 00:00:01:10:10:10 for IP 172.16.0.10, as follows:

```
# on PE-3:
42 2021/03/24 16:06:11.147 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 Mac Change: 00:00:01:10:10:10->00:00:01:11:11:11 "

43 2021/03/24 16:06:11.290 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 mac: 00:00:01:10:10:10 confirm"
```

When CE-10 confirms MAC 00:00:01:10:10:10 for IP 172.16.0.10, IP duplication is detected for IP address 172.16.0.10 (after three MAC moves in a detection period of three minutes), and the following message is raised in log 99 after a duplicate proxy-ARP entry was detected for IP 172.16.0.10:

```
# log "99" on PE-3:
136 2021/03/24 16:06:15.358 UTC MINOR: SVCMGR #2346 Base
"A duplicate proxy ARP entry was detected with new MAC 00:00:01:10:10:10 for entry
IP 172.16.0.10 MAC 00:00:01:11:11:11 in service 1"
```

The following proxy-ARP debug messages show that the ARP entry for IP 172.16.0.10 in the proxy-ARP table changed MAC to the AS-MAC 00:00:bb:bb:bb:bb, and the type from dynamic to duplicate:

```
# on PE-3:
44 2021/03/24 16:06:15.357 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 mac: 00:00:bb:bb:bb:bb evpn advertise"

45 2021/03/24 16:06:15.358 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 Mac Change: 00:00:01:11:11:11->00:00:bb:bb:bb:bb Type Change: Dyn->Dup "

46 2021/03/24 16:06:15.358 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 type: Dup    Dup Detected"
```

If a duplicate IP is detected, AS-MAC 00:00:bb:bb:bb:bb is advertised with duplicate IP address 172.16.0.10 in an EVPN-MAC update to the BGP peers with the sticky/static bit set, as follows:

```
47 2021/03/24 16:06:15.358 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 93
    Flag: 0x90 Type: 14 Len: 48 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-MAC Len: 37 RD: 192.0.2.3:1 ESI: ESI-0, tag: 0, mac len: 48
                   mac: 00:00:bb:bb:bb:bb, IP len: 4, IP: 172.16.0.10, label1: 8388544
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:1
        bgp-tunnel-encap:MPLS
        mac-mobility:Seq:0/Static
"
```

The difference with the first EVPN-MAC update for AS-MAC is the IP address. Immediately after the AS-MAC was configured, it was also advertised to the BGP-EVPN peers, but without any IP address.

The proxy-ARP entry is shown with type duplicate (dup) and active status in the proxy-ARP table for VPLS 1 on PE-3, as follows:

```
*A:PE-3# show service id 1 proxy-arp detail
-------------------------------------------------------------------------------
Proxy Arp
-------------------------------------------------------------------------------
Admin State       : enabled
Dyn Populate      : enabled
Age Time          : disabled            Send Refresh      : disabled
Table Size        : 250                 Total             : 2
Static Count      : 1                   EVPN Count        : 0
Dynamic Count     : 0                   Duplicate Count   : 1

Dup Detect
-------------------------------------------------------------------------------
Detect Window     : 3 mins              Num Moves         : 3
Hold down         : max
Anti Spoof MAC    : 00:00:bb:bb:bb:bb
```

```
EVPN
-------------------------------------------------------------------------------
Garp Flood         : enabled             Req Flood         : enabled
Static Black Hole : disabled
EVPN Route Tag     : 0
-------------------------------------------------------------------------------

===============================================================================
VPLS Proxy Arp Entries
===============================================================================
IP Address          Mac Address         Type      Status    Last Update
-------------------------------------------------------------------------------
172.16.0.10         00:00:bb:bb:bb:bb   dup       active    03/24/2021 16:06:15
172.16.0.20         00:00:02:20:20:20   stat      inActv    03/24/2021 16:01:31
-------------------------------------------------------------------------------
Number of entries : 2
===============================================================================
```

A duplicate entry is always active, regardless of the AS-MAC. When the entry with the duplicate IP address and the AS-MAC address are installed in the proxy-ARP table as active, every ARP request for the duplicate IP address will be replied by the system. The entry in the proxy-ARP table is treated as active, even if the AS-MAC address is not in the FDB (AS-MAC addresses do not consume FDB space). The AS-MAC address, along with the duplicate IP address, is advertised in EVPN with the sticky/static bit set, as shown earlier. GARP messages with AS-MAC/IP information are flooded locally to make the CEs update their ARP caches to use the AS-MAC address for traffic to the duplicate IP 172.16.0.10, as follows:

```
# on PE-3:
48 2021/03/24 16:06:15.490 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 type: Dup mac: 00:00:bb:bb:bb:bb Gratuitous Update"
```

Note:
The AS-MAC address will always be "unique" in the system. When the AS-MAC is configured, the system will flush any entry with the same MAC address learned through EVPN or dynamic sources. Conditional static MAC addresses or OAM MAC addresses with the same value as the AS-MAC address are only allowed when they are configured as black-hole, which is not the case yet.

When the duplicate proxy-ARP entry is cleared from the list (hold-down timer expires, or clear command, or replacement of the duplicate entry for a static entry), an ARP request asking who has IP 172.16.0.10 is flooded by the proxy-ARP agent. This ARP refresh triggers an ARP reply from the IP owner, which will be learned in the proxy-ARP table and advertised in EVPN. The system will also send a GARP to local SAP/SDP-bindings. This will correct all host ARP caches in the network. In this example, the duplicate proxy-ARP entry is manually cleared, as follows:

```
*A:PE-3# clear service id 1 proxy-arp duplicate
```

Log "99" shows that the clear function has been run and the duplicate proxy-ARP entry 172.16.0.10 is cleared. The system forces a refresh and, if the condition with the duplicate IP address remains, this is detected almost immediately and a message is logged that a duplicate proxy-ARP entry was detected, as follows:

```
# on PE-3:
137 2021/03/24 16:08:51.964 UTC INDETERMINATE: LOGGER #2010 Base Clear SVCMGR
"Clear function clearSvcIdProxyArpDups has been run with parameters: svc-id="1"
ip-address="".  The completion result is: success.  Additional error text, if any, is: "
```

```
138 2021/03/24 16:08:51.965 UTC MINOR: SVCMGR #2347 Base
"A duplicate proxy ARP entry 172.16.0.10 is cleared in service 1"

139 2021/03/24 16:08:52.193 UTC MINOR: SVCMGR #2346 Base
"A duplicate proxy ARP entry was detected with new MAC 00:00:01:11:11:11 for entry
IP 172.16.0.10 MAC 00:00:01:10:10:10 in service 1"
```

The following debug messages for proxy-ARP on PE-3 show the process in more detail. Initially, an EVPN-MAC route withdraw message is sent and the proxy-ARP entry is deleted.

```
# on PE-3:
49 2021/03/24 16:08:51.964 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 mac: 00:00:bb:bb:bb:bb evpn withdraw"

50 2021/03/24 16:08:51.964 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 type: Dup mac: 00:00:bb:bb:bb:bb Deleted"
```

The following BGP-EVPN MAC update is sent by PE-3 to indicate that the AS-MAC is withdrawn for IP 172.16.0.10 (multiprotocol unreachable NLRI):

```
# on PE-3:
52 2021/03/24 16:08:51.965 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 46
    Flag: 0x90 Type: 15 Len: 42 Multiprotocol Unreachable NLRI:
        Address Family EVPN
        Type: EVPN-MAC Len: 37 RD: 192.0.2.3:1 ESI: ESI-0, tag: 0, mac len: 48
                       mac: 00:00:bb:bb:bb:bb, IP len: 4, IP: 172.16.0.10, label1: 0
"
```

Removing the active duplicate entry from the proxy-ARP table triggers an ARP flooding request asking who has IP 172.16.0.10 in VPLS 1, as follows:

```
# on PE-3:
51 2021/03/24 16:08:51.964 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 flood request"
```

The result of the ARP flooding request is that the IP owners reply with their MAC, at the local or a remote PE. In this case, the reply from CE-10 is received first (IP 172.16.0.10 - MAC 00:00:01:10:10:10), a dynamic proxy-ARP entry is added, and the MAC/IP route is advertised, as follows:

```
# on PE-3:
53 2021/03/24 16:08:51.967 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 mac: 00:00:01:10:10:10 evpn advertise"

54 2021/03/24 16:08:51.967 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 type: Dyn mac: 00:00:01:10:10:10 Added"
```

When CE-11 answers with its MAC 00:00:01:11:11:11, the MAC/IP route is withdrawn for IP 172.16.0.10, and the MAC address in the proxy-ARP entry for IP 172.16.0.10 is changed from MAC 00:00:01:10:10:10 to MAC 00:00:01:11:11:11, as follows:

```
# on PE-3:
55 2021/03/24 16:08:51.967 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 mac: 00:00:01:10:10:10 evpn withdraw"

56 2021/03/24 16:08:51.967 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 Mac Change: 00:00:01:10:10:10->00:00:01:11:11:11 "
```

Any change of MAC address in a proxy-ARP entry triggers an ARP request asking for confirmation of the old MAC address for IP 172.16.0.10, in this case for MAC 00:00:01:10:10:10, as follows:

```
# on PE-3:
57 2021/03/24 16:08:52.090 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 mac: 00:00:01:10:10:10 confirm"
```

MAC address 00:00:01:10:10:10 is confirmed for IP address 172.16.0.10; therefore, the MAC address is changed in the proxy-ARP entry from 00:00:01:11:11:11 to 00:00:01:10:10:10, and an ARP confirmation is asked for the old MAC address 00:00:01:11:11:11, as follows:

```
# on PE-3:
58 2021/03/24 16:08:52.093 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 Mac Change: 00:00:01:11:11:11->00:00:01:10:10:10 "

59 2021/03/24 16:08:52.190 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 mac: 00:00:01:11:11:11 confirm"
```

MAC address 00:00:01:11:11:11 is confirmed and, therefore, three MAC moves occurred within three minutes. Duplicate IP 172.16.0.10 is detected and the proxy-ARP entry has the AS-MAC 00:00:bb:bb:bb:bb and type duplicate (Dup), as follows:

```
# on PE-3:
60 2021/03/24 16:08:52.193 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 mac: 00:00:bb:bb:bb:bb evpn advertise"

61 2021/03/24 16:08:52.193 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 Mac Change: 00:00:01:10:10:10->00:00:bb:bb:bb:bb Type Change: Dyn->Dup "

62 2021/03/24 16:08:52.193 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 type: Dup    Dup Detected"

63 2021/03/24 16:08:52.193 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 93
    Flag: 0x90 Type: 14 Len: 48 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-MAC Len: 37 RD: 192.0.2.3:1 ESI: ESI-0, tag: 0, mac len: 48
                   mac: 00:00:bb:bb:bb:bb, IP len: 4, IP: 172.16.0.10, label1: 8388544
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:1
        bgp-tunnel-encap:MPLS
        mac-mobility:Seq:0/Static
"
```

A GARP update is sent for IP 172.16.0.10 and AS-MAC 00:00:bb:bb:bb:bb, as follows:

```
# on PE-3:
64 2021/03/24 16:08:52.290 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.10 type: Dup mac: 00:00:bb:bb:bb:bb Gratuitous Update"
```

The AS-MAC address is optionally configured and populates all the host ARP caches when a duplicate IP address is detected. All traffic destined to the suspicious IP address 172.16.0.10 will have the AS-MAC address 00:00:bb:bb:bb:bb as MAC DA. The user can configure MAC filters on all SAP/SDP-bindings where the CEs are connected to drop, log, or redirect traffic destined to the AS-MAC. This will block any interception or man-in-the-middle attack (due to ARP-spoofing) in the network.

The AS-MAC address is independently configured on each PE for the same service. When a different AS-MAC address is configured per PE for the same service, the user will need to filter all the AS-MAC addresses in the service at each PE, which increases the complexity of the filters. Nokia recommends using the same AS-MAC address for the same service in all the PES where duplicate detect is active and MAC filters need to be configured. However, this recommendation is suspended when the AS-MAC address is configured as static black-hole MAC address, as described in the following section.

Services with proxy-ARP and AS-MAC configured as static black-hole MAC

With the AS-MAC address configured as static black-hole MAC address, MAC-filters do not need to be configured to discard frames with MAC DA equal to the AS-MAC address. Instead, the user can decide whether to use the same AS-MAC address on all the PEs. This scalability is not limited by the number of filters, but by the number of FDB entries.

The **static-black-hole** parameter is optional and disabled by default. In the example, the static-black-hole option is not configured yet for the AS-MAC address and the behavior is as follows:

- The AS-MAC address is added to the MAC DB as local, but not programmed in the FDB.
- The AS-MAC address is advertised in EVPN (initially without an IP address, and with an IP address as soon as the IP is detected as duplicate).
- The AS-MAC address cannot be overridden by any other MAC address.
- The AS-MAC address value cannot be configured on a static MAC address, because that MAC address is reserved for the proxy-ARP, as follows:

```
*A:PE-3>conf>serv>vpls>static-mac# mac 00:00:bb:bb:bb:bb create sap 1/2/3:1 monitor fwd-status
MINOR: SVCMGR #7875 Cannot create conditional static mac - Mac reserved by proxy

*A:PE-3>conf>serv>vpls>static-mac# mac 00:00:bb:bb:bb:bb create black-hole
MINOR: SVCMGR #7875 Cannot create conditional static mac - Mac reserved by proxy
```

When the static-black-hole option is not configured, the AS-MAC address is considered as a local MAC address and cannot be overridden. The MAC address priority is as follows:

1. Local MAC address (including AS-MAC addresses without static-black-hole, es-bmacs, src-bmacs, OAM, and so on)
2. Conditional static MAC addresses (including AS-MAC addresses with static-black-hole)
3. Auto-Learn Protected MAC addresses
4. EVPN-MAC addresses with sticky/static bit set
5. Data plane learned MAC addresses (regular learning on SAP/SDP-binding)
6. EVPN-MAC addresses without sticky/static bit set

To configure an AS-MAC address with static-black-hole option, a static black-hole MAC address needs to be configured first. The following error is raised when no static black-hole MAC has been configured for AS-MAC 00:00:bb:bb:bb:bb:

```
*A:PE-3>config>service>vpls>proxy-arp# dup-detect window 3 num-moves 5 hold-down max anti-
spoof-mac 00:00:bb:bb:bb:bb static-black-hole
MINOR: SVCMGR #8007 Cannot modify proxy arp - black-hole mac not configured on service
```

In that case, the AS-MAC address needs to be removed from the proxy-ARP configuration, as follows:

```
# on PE-2, PE-3:
configure
    service
        vpls "VPLS 1"
            proxy-arp
                shutdown
                dup-detect window 3 num-moves 5 hold-down max
            exit
```

Then, the static black-hole MAC address can be created as follows:

```
# on PE-2, PE-3:
configure
    service
        vpls "VPLS 1"
            static-mac
                mac 00:00:bb:bb:bb:bb create black-hole
```

After the conditional static black-hole MAC address is configured, duplicate IP address detection cannot be configured with AS-MAC address, unless the static-black-hole option is added, as follows:

```
*A:PE-3>config>service>vpls>proxy-arp# dup-detect window 3 num-moves 5 hold-down max anti-
spoof-mac 00:00:bb:bb:bb:bb
MINOR: SVCMGR #8007 Cannot modify proxy arp - conditional static mac configured on service
```

When the static black-hole MAC 00:00:bb:bb:bb:bb is configured, the AS-MAC address can only be configured with the **static-black-hole** option in VPLS 1 on PE-2 and PE-3, as follows:

```
# on PE-2, PE-3:
configure
    service
        vpls "VPLS 1"
            static-mac
                mac 00:00:bb:bb:bb:bb create black-hole
            exit
            proxy-arp
                dup-detect window 3 num-moves 5 hold-down max
                           anti-spoof-mac 00:00:bb:bb:bb:bb static-black-hole
                dynamic-arp-populate
                static 172.16.0.20 00:00:02:20:20:20
                no shutdown
            exit
```

When the AS-MAC address is configured with the static black-hole option, the AS-MAC will be added not only to the MAC DB, but also to the FDB as CStatic, and associated with a black-hole endpoint, as follows:

```
*A:PE-3# show service id 1 fdb mac 00:00:bb:bb:bb:bb
===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:bb:bb:bb:bb black-hole              CStatic: 03/24/21 16:13:59
                                                     P
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

Any frame with MAC DA equal to the AS-MAC with static black-hole will be dropped, regardless of the ingress endpoint and without any need for a filter. This mechanism is the only way to filter MAC DAs on EVPN endpoints, because MAC filters cannot be configured on EVPN endpoints.

The AS-MAC with static black-hole will be advertised in EVPN with the sticky/static bit set, as follows:

```
# on PE-3:
75 2021/03/24 16:13:58.752 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 89
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-MAC Len: 33 RD: 192.0.2.3:1 ESI: ESI-0, tag: 0, mac len: 48
                       mac: 00:00:bb:bb:bb:bb, IP len: 0, IP: NULL, label1: 8388544
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:1
        bgp-tunnel-encap:MPLS
        mac-mobility:Seq:0/Static
"
```

When a duplicate IP address is detected, the EVPN-MAC update contains the IP address 172.16.0.10, as follows:

```
90 2021/03/24 16:15:36.266 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 93
    Flag: 0x90 Type: 14 Len: 48 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-MAC Len: 37 RD: 192.0.2.3:1 ESI: ESI-0, tag: 0, mac len: 48
                  mac: 00:00:bb:bb:bb:bb, IP len: 4, IP: 172.16.0.10, label1: 8388544
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:1
        bgp-tunnel-encap:MPLS
        mac-mobility:Seq:0/Static
"
```

The local CEs receive a GARP update with the AS-MAC address. The ARP table of CE-30 and CE-31 have an entry for the duplicate IP address 172.16.0.10 with the AS-MAC address 00:00:bb:bb:bb:bb, as follows:

```
*A:PE-3# show router 10 arp

===============================================================================
ARP Table (Service: 10)
===============================================================================
IP Address      MAC Address       Expiry    Type   Interface
-------------------------------------------------------------------------------
172.16.0.10     00:00:bb:bb:bb:bb 03h43m02s Dyn[I] int-CE-30-PE-3
172.16.0.30     00:00:aa:aa:aa:aa 00h00m00s Oth[I] int-CE-30-PE-3
-------------------------------------------------------------------------------
No. of ARP Entries: 2
===============================================================================
```

```
*A:PE-3# show router 11 arp

===============================================================================
ARP Table (Service: 11)
===============================================================================
IP Address      MAC Address       Expiry    Type   Interface
-------------------------------------------------------------------------------
172.16.0.10     00:00:bb:bb:bb:bb 03h47m43s Dyn[I] int-CE-31-PE-3
172.16.0.31     00:00:03:31:31:31 00h00m00s Oth[I] int-CE-31-PE-3
-------------------------------------------------------------------------------
No. of ARP Entries: 2
===============================================================================
```

CE-30 and CE-31 cannot reach CE-10 or CE-11, because the MAC DA will be the AS-MAC address and all traffic to this MAC DA is black-holed instead of forwarded to SAP 1/2/3:1 toward CE-10 or CE-11. When 1000 ICMP packets are sent by CE-30, they arrive in SAP 1/2/1:1 on PE-3 and are then discarded, as follows:

```
*A:PE-3# clear port 1/[1..2]/[1..4] statistics

*A:PE-3# ping router 10 172.16.0.10 rapid count 1000
PING 172.16.0.10 56 data bytes
---snip---
---- 172.16.0.10 PING Statistics ----
1000 packets transmitted, 0 packets received, 100% packet loss

*A:PE-3# show port 1/[1..2]/[1..4] statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
```

```
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/2                                    13                            1274
                                         13                            1274
===============================================================================

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/3                                    16                            1537
                                         16                            1537
===============================================================================

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/2/1                                  1000                          106000
                                          0                               0
===============================================================================
```

No packets were forwarded to SAP 1/2/3:1 toward MTU-1; therefore, there are no statistics for port 1/2/3.

Conclusion

Static black-hole MAC addresses can be applied in EVPN for security as a scalable alternative to MAC filters. Static black-hole MAC addresses are programmed in the FDB and all frames with MAC DA equal to the static black-hole MAC address are dropped, regardless of how the frame arrived at the system (SAP/SDP-binding or EVPN endpoint). Also, static black-hole MAC addresses are treated like protected MAC addresses and, in combination with RPS(-DF), filtering on MAC SA is performed in the data plane. Black-hole MAC addresses can be an option for an AS-MAC address in services with proxy-ARP/ND enabled, which simplifies the configuration because MAC filters are not required.

Data Center Interconnect Using Dual EVPN-VXLAN Instance VPLS

This chapter provides information about Data Center Interconnect using dual EVPN-VXLAN instance VPLS.

Topics in this chapter include:

Applicability

This chapter was initially written for SR OS Release 16.0.R7, but the CLI in the current edition is based on SR OS Release 21.7.R1. Dual EVPN-VXLAN instances are supported in SR OS Release 16.0.R2, or later.

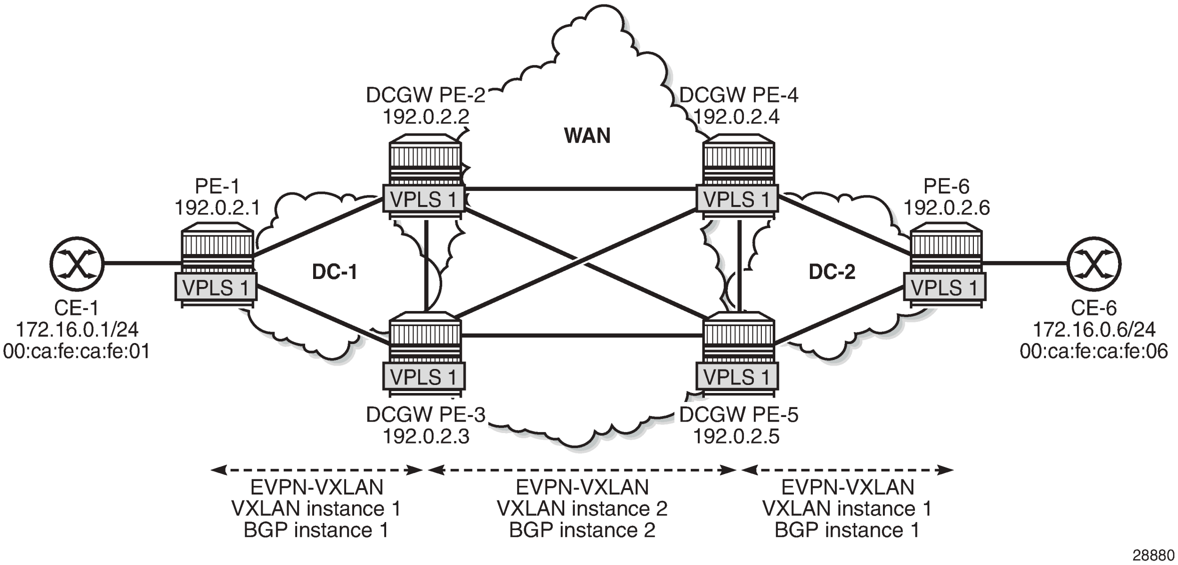
This chapter describes the redundancy based on an Anycast solution, as supported in SR OS Release 16.0, and later. For I-ES based redundancy scenarios as supported in SR OS Release 19.10, and later, see the EVPN Interconnect Ethernet Segments in Dual EVPN-VXLAN Instance VPLS Services chapter.

Overview

Chapter EVPN-MPLS Interconnect for EVPN-VXLAN VPLS Services describes a Data Center Interconnect (DCI) scenario using VXLAN in the DCs and MPLS in the WAN. This chapter describes a similar scenario, where the core is an IP network that does not use MPLS, and where end-to-end VXLAN is used instead. The DC Gateways (GWs) contain VPLS services with two EVPN-VXLAN instances and two BGP instances: one EVPN-VXLAN instance faces the DC and the other EVPN-VXLAN instance faces the WAN.

Figure 41: Dual EVPN-VXLAN instance VPLS 1 shows the example topology with two DCs. On PE-1 and PE-6, VPLS 1 is configured with one VXLAN instance and one BGP instance. On the DC GWs, VPLS 1 is configured with two VXLAN instances and two BGP instances: one toward the DC and one toward the WAN.

Figure 41: Dual EVPN-VXLAN instance VPLS 1



For example, on DC GW PE-2, VPLS 1 is configured with VXLAN instance 1 using BGP instance 1 and VXLAN 2 using BGP instance 2. In this example, the BGP instance ID matches the VXLAN instance ID, but that is not required. Each VXLAN instance has a different VNI and a different BGP instance.

```
# on PE-2:
configure
    service
        system
            bgp-auto-rd-range 10.0.0.1 comm-val 60000 to 65000
        exit
        vpls 1 name "VPLS 1" customer 1 create
            description "dual evpn-vxlan VPLS"
            vxlan instance 1 vni 11 create
            exit
            vxlan instance 2 vni 12 create
            exit
            bgp
                route-distinguisher auto-rd
                route-target export target:64500:11 import target:64500:11
            exit
            bgp 2
                route-distinguisher auto-rd
                route-target export target:64500:12 import target:64500:12
            exit
            bgp-evpn
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
                vxlan bgp 2 vxlan-instance 2
                    no shutdown
                exit
            exit
            no shutdown
```

When different BGP instances are configured, the auto-derived route distinguishers (RDs) in BGP instance 1 and BGP instance 2 are different, as follows:

```
*A:PE-2# show service id 1 bgp 1 | match "Route Dist"
Route Dist            : auto-rd
Oper Route Dist       : 10.0.0.1:60000
*A:PE-2# show service id 1 bgp 2 | match "Route Dist"
Route Dist            : auto-rd
Oper Route Dist       : 10.0.0.1:60001
```

Dual EVPN-VXLAN instance VPLSs can contain SAPs in SR OS Release 19.10.R1, and later. However, dual EVPN-VXLAN instance VPLSs cannot contain any SDP bindings in SR OS Release 21.7.R1, as follows:

```
*A:PE-2>config>service>vpls# spoke-sdp 21:1 create
MINOR: SVCMGR #1997 Cannot create sdp binding - not supported with multiple vxlan instances
```

Note:
This chapter describes the redundancy based on an Anycast solution, as supported in SR OS Release 16.0, and later. For I-ES based redundancy scenarios as supported in SR OS Release 19.10, and later, see chapter EVPN Multi-Homing on Dual EVPN-VXLAN BGP Instance VPLS.

To provide DC GW redundancy, an anycast IP address can be configured for the dual EVPN-VXLAN instance VPLSs on the DC GWs.

EVPN route types 2 and 3 are processed by dual EVPN-VXLAN VPLS services as follows:

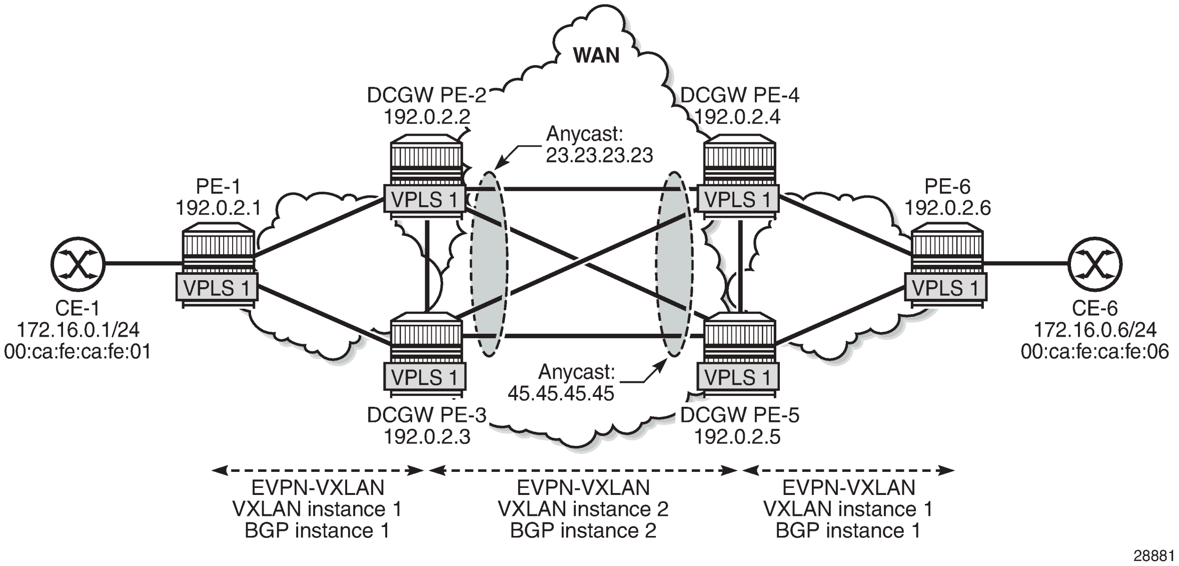
- Route type 2: MAC/IP routes
  - MAC/IP routes received in a BGP instance will be imported and—according to the selection rules—installed in the FDB.
  - Active MAC routes are re-advertised in the other BGP instance with new BGP attributes (RD, route target (RT), and so on).
    - Only the best EVPN MAC route is redistributed.
    - The MAC/IP information and the sticky bit are propagated. The only exception is the Ethernet Segment Identifier (ESI). A non-zero ESI will be reset unless the auto-disc-route-advertisement command is configured.
    - When an attribute has changed for a redistributed MAC route, the MAC route will be updated if it is still the best route. For example, an update of the sequence number or the sticky bit can trigger a redistribution.
- Route type 3: inclusive multicast routes
  - EVPN inclusive multicast routes are generated independently for each instance with the proper BGP extended communities.
  - Ingress Replication (IR) or Assisted Replication (AR) Inclusive Multicast Ethernet Tag (IMET) routes are supported.
  - The inclusive multicast originating IP can be configured with an anycast address:
    - The configured originating IP address is encoded in the originating IP field of the IMET-IR routes; the originating IP field of the IMET-AR routes is still derived from the assisted replication IP value in the service system settings for VXLAN.

- If a router receives two IMET routes with the same originating IP address, different RDs, and different next-hops, it sets up two bindings: one to each next-hop.
- If a router receives two IMET routes with the same originating IP address, the same RD, but different next-hops, it sets up one binding to the next-hop with the lowest IP address.
- If a router receives two IMET routes with the same originating IP address, different RDs, but the same next-hop, it sets up one binding to the next-hop.
- A DC GW will not set up a binding to its DC GW peer if the received originating IP equals its own originating IP, regardless of whether the local RD and the remote RD are the same or different.

Configuration

Figure 42: Example topology with VPLS 1 and anycast addresses shows the example topology. Redundancy is based on anycast: on the DC GWs PE-2 and PE-3, anycast address 23.23.23.23 is configured as inclusive multicast originating IP; on PE-4 and PE-5 in DC-2, the anycast address is 45.45.45.45. However, no Ethernet segments are used in this example.

Figure 42: Example topology with VPLS 1 and anycast addresses



The initial configuration includes:

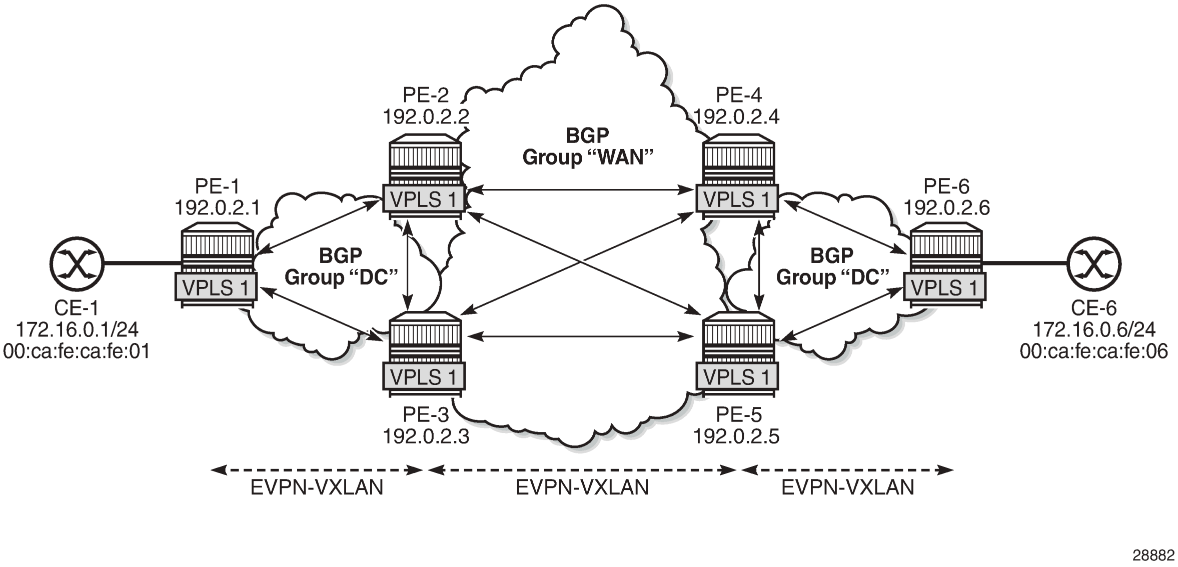
- Cards, MDAs, ports
- Router interfaces
- IS-IS as IGP (level 1 in the DCs and level 2 in the WAN)

MPLS is not configured in any of these networks.

BGP configuration

BGP is configured for the EVPN address family on all nodes. Figure 43: Example topology with BGP groups shows the BGP groups: on the DC GWs, both BGP group "DC" and "WAN" are defined. Route policies ensure that only DC routes are forwarded to DC neighbors and only WAN routes are forwarded to WAN neighbors. Also, DC GWs need to drop BGP-EVPN routes from the local peer DC GW.

Figure 43: Example topology with BGP groups



The BGP configuration on PE-1 is as follows. The configuration on PE-6 is similar.

```
# on PE-1:
configure
    router Base
        autonomous-system 64500
        bgp
            family evpn
            vpn-apply-import
            vpn-apply-export
            rapid-withdrawal
            rapid-update evpn
            group "DC"
                type internal
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
```

On the DC GWs, two BGP groups are defined: one for the DC group and one for the WAN group. Export policies ensure that only DC routes are exported to the DC group and only WAN routes are exported to the WAN group. Import policies ensure that routes from the local DC GW are dropped; for example, PE-2 drops routes from PE-3, and vice versa. The policies will be described later. The BGP configuration on PE-2 is as follows. The BGP configuration on the other DC GWs is similar.

```
# on PE-2:
configure
    router Base
        autonomous-system 64500
        bgp
            family evpn
            vpn-apply-import
            vpn-apply-export
            rapid-withdrawal
            rapid-update evpn
            group "DC"
                type internal
                import "drop SOO-DCGW-23"
                export "allow only DC and add SOO"
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.3
                exit
            exit
            group "WAN"
                type internal
                import "drop SOO-DCGW-23"
                export "allow only WAN and add SOO"
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
        exit
```

Route policies

The route policies are equivalent to the policies described in the chapter EVPN-MPLS Interconnect for EVPN-VXLAN VPLS Services. In this example, no filtering can be done based on the encapsulation extended community (VXLAN versus MPLS), because only VXLAN is used in the DCs and the WAN. Therefore, the route tag is used as a criterion instead in the route policies "allow only DC and add SOO" (Site of Origin) and "allow only WAN and add SOO". When two BGP instances for the same encapsulation are configured in a VPLS, different route tags in each BGP instance are required. In this example, route tag 11 is used in BGP instance 1 in the DCs and route tag 12 is used in BGP instance 2 in the WAN.

When redistributing to the other BGP instance, route filtering toward DC or WAN will be based on the route tags. Export policy "allow only DC and add SOO" drops routes with WAN route tag 12. Likewise, export policy "allow only WAN and add SOO" drops routes with DC route tag 11. Filtering matching on route tags on EVPN BGP instances is scalable, because only two route tags are required per PE. Filtering matching on route target (RT) is also possible, but in that case, two RTs per service are required. This does not scale well and is cumbersome.

The export policy "allow only DC and add SOO" ensures that EVPN routes with route tag 12 are dropped and only DC routes are forwarded. This route policy is applied in the BGP group "DC" context. Likewise, the export policy "allow only WAN and add SOO" drops EVPN routes with route tag 11, so that only WAN EVPN routes are forwarded.

Both policies also add a site of origin, such as "SOO-23" for PE-2 and PE-3, and "SOO-45" for PE-4 and PE-5. This SOO is used for filtering in the import policies "drop SOO-DCGW-23" and "drop SOO-DCGW-45" to ensure that, for instance, PE-2 drops routes advertised by the local peer PE-3 with the same SOO-23, and vice versa. Likewise, PE-4 drops routes advertised by its local peer PE-5 with the same SOO-45, and vice versa.

The following policies are configured on DC GWs PE-2 and PE-3:

```
# on PE-2, PE-3:
configure
    router Base
        policy-options
            begin
            community "SOO-23"
                members "origin:64500:23"
            exit
            policy-statement "drop SOO-DCGW-23"
                entry 10
                    from
                        community "SOO-23"
                        family evpn
                    exit
                    action drop
                    exit
                exit
            exit
            policy-statement "allow only DC and add SOO"
                entry 10
                    from
                        tag 12
                        family evpn
                    exit
                    action drop
                    exit
                exit
                entry 20
                    from
                        family evpn
                    exit
                    action accept
                        community add "SOO-23"
                    exit
                exit
            exit
            policy-statement "allow only WAN and add SOO"
                entry 10
                    from
                        tag 11
                        family evpn
                    exit
                    action drop
                    exit
                exit
                entry 20
                    from
                        family evpn
                    exit
                    action accept
                        community add "SOO-23"
                    exit
                exit
            exit
            commit
```

The following policies are configured on DC GWs PE-4 and PE-5:

```
# on PE-4, PE-5:
configure
    router Base
        policy-options
            begin
            community "SOO-45"
                members "origin:64500:45"
            exit
            policy-statement "drop SOO-DCGW-45"
                entry 10
                    from
                        community "SOO-45"
                        family evpn
                    exit
                    action drop
                    exit
                exit
            exit
            policy-statement "allow only DC and add SOO"
                entry 10
                    from
                        tag 12
                        family evpn
                    exit
                    action drop
                    exit
                exit
                entry 20
                    from
                        family evpn
                    exit
                    action accept
                        community add "SOO-45"
                    exit
                exit
            exit
            policy-statement "allow only WAN and add SOO"
                entry 10
                    from
                        tag 11
                        family evpn
                    exit
                    action drop
                    exit
                exit
                entry 20
                    from
                        family evpn
                    exit
                    action accept
                        community add "SOO-45"
                    exit
                exit
            exit
            commit
```

VPLS configuration

On PE-2 and PE-3, the service configuration is identical and VPLS 1 is configured as follows. For redundancy, the anycast IP address 23.23.23.23 is configured as inclusive multicast originating IP on PE-2 and PE-3. The RT is the same in all nodes: the RT is 64500:11 in BGP instance 1 of VPLS 1; in BGP instance 2, the RT is 64500:12. The RD is 64500:2311 in BGP instance 1 of VPLS 1 and 64500:2312 in BGP instance 2 of VPLS 1 on PE-2 and PE-3. The RD must be the same in PE-2 and PE-3 because they are part of the anycast group, but the RD in PE-1 must be different.

```
# on PE-2, PE-3:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            vxlan instance 1 vni 11 create
            exit
            vxlan instance 2 vni 12 create
            exit
            bgp
                route-distinguisher 64500:2311
                route-target export target:64500:11 import target:64500:11
            exit
            bgp 2
                route-distinguisher 64500:2312
                route-target export target:64500:12 import target:64500:12
            exit
            bgp-evpn
                incl-mcast-orig-ip 23.23.23.23
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    default-route-tag 11
                    no shutdown
                exit
                vxlan bgp 2 vxlan-instance 2
                    default-route-tag 12
                    no shutdown
                exit
            exit
            no shutdown
```

On PE-1, VPLS 1 is configured with VXLAN instance 1 and BGP instance 1, as follows. The RT is 64500:11 in BGP instance 1 of VPLS 1 on PE-1. The RD (64500:111) in PE-1 is different from the RD (64500:2311) in PE-2 and PE-3.

```
# on PE-1:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            vxlan instance 1 vni 11 create
            exit
            bgp
                route-distinguisher 64500:111
                route-target export target:64500:11 import target:64500:11
            exit
            bgp-evpn
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            sap 1/2/1:1 create
            exit
            no shutdown
```

On PE-4 and PE-5, the service configuration is identical and VPLS 1 is configured as follows. For redundancy, the anycast IP address 45.45.45.45 is configured as inclusive multicast originating IP.

```
# on PE-4, PE-5:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            vxlan instance 1 vni 11 create
            exit
            vxlan instance 2 vni 12 create
            exit
            bgp
                route-distinguisher 64500:4511
                route-target export target:64500:11 import target:64500:11
            exit
            bgp 2
                route-distinguisher 64500:4512
                route-target export target:64500:12 import target:64500:12
            exit
            bgp-evpn
                incl-mcast-orig-ip 45.45.45.45
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    default-route-tag 11
                    no shutdown
                exit
                vxlan bgp 2 vxlan-instance 2
                    default-route-tag 12
                    no shutdown
                exit
            exit
            no shutdown
```

Verification

On PE-2, the following EVPN inclusive multicast routes are received. The first route has RD 64500:111, so it applies to BGP instance 1; the last two have RD 64500:4512, so they apply to BGP instance 2. Toward anycast address 45.45.45.45, the lowest IP next-hop 192.0.2.4 (PE-4) is preferred over 192.0.2.5 (PE-5).

```
*A:PE-2# show router bgp routes evpn incl-mcast
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Flag  Route Dist.         OrigAddr
      Tag                 NextHop
-------------------------------------------------------------------------------
u*>i  64500:111           192.0.2.1
```

```
      0                   192.0.2.1

u*>i  64500:4512          45.45.45.45
      0                   192.0.2.4

*>i   64500:4512          45.45.45.45
      0                   192.0.2.5

-------------------------------------------------------------------------------
Routes : 3
===============================================================================
```

The following shows the VXLAN destinations on PE-2: PE-1 (192.0.2.1) is the VXLAN Tunnel Endpoint (VTEP) in VXLAN instance 1; the VTEP for VXLAN instance 2 is PE-4 (192.0.2.4).

```
*A:PE-2# show service id 1 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
Instance    VTEP Address                            Egress VNI  EvpnStatic Num
 Mcast       Oper State                              L2 PBR     SupBcasDom MACs
-------------------------------------------------------------------------------
1           192.0.2.1                               11          evpn       0
 BUM         Up                                      No          No
2           192.0.2.4                               12          evpn       0
 BUM         Up                                      No          No
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 2
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

The following shows the BGP information for VPLS 1 on PE-1. Only BGP instance 1 is configured. The RD is configured with the value 64500:111 and the RT with the value 64500:11, which are also the operational values. No VSI import or VSI export policies are configured on PE-1.

```
*A:PE-1# show service id 1 bgp

===============================================================================
BGP Information
===============================================================================
Bgp Instance         : 1
Vsi-Import           : None
Vsi-Export           : None
Route Dist           : 64500:111
Oper Route Dist      : 64500:111
Oper RD Type         : configured
Rte-Target Import    : 64500:11             Rte-Target Export: 64500:11
Oper RT Imp Origin   : configured           Oper RT Import   : 64500:11
Oper RT Exp Origin   : configured           Oper RT Export   : 64500:11

PW-Template Id       : None
-------------------------------------------------------------------------------
```

```
===============================================================================
```

On PE-2, the following information for BGP instance 1 includes the configured and operational RD 64500:2311 and RT 64500:11.

```
*A:PE-2# show service id 1 bgp 1

===============================================================================
BGP Information
===============================================================================
Vsi-Import           : None
Vsi-Export           : None
Route Dist           : 64500:2311
Oper Route Dist      : 64500:2311
Oper RD Type         : configured
Rte-Target Import    : 64500:11             Rte-Target Export: 64500:11
Oper RT Imp Origin   : configured           Oper RT Import   : 64500:11
Oper RT Exp Origin   : configured           Oper RT Export   : 64500:11
PW-Template Id       : None
-------------------------------------------------------------------------------
===============================================================================
```

On PE-2, the following information for BGP instance 1 includes the configured and operational RD 64500:2312 and RT 64500:12.

```
*A:PE-2# show service id 1 bgp 2

===============================================================================
BGP Information
===============================================================================
Vsi-Import           : None
Vsi-Export           : None
Route Dist           : 64500:2312
Oper Route Dist      : 64500:2312
Oper RD Type         : configured
Rte-Target Import    : 64500:12             Rte-Target Export: 64500:12
Oper RT Imp Origin   : configured           Oper RT Import   : 64500:12
Oper RT Exp Origin   : configured           Oper RT Export   : 64500:12
-------------------------------------------------------------------------------
===============================================================================
```

The CEs are simulated by VPRN 11 configured on PE-1 and PE-6. Connectivity between CE-1 and CE-6 is verified as follows:

```
*A:PE-1# ping router 11 172.16.0.6 rapid
PING 172.16.0.6 56 data bytes
!!!!!
---- 172.16.0.6 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 3.75ms, avg = 4.40ms, max = 5.31ms, stddev = 0.511ms
```

The following two EVPN MAC routes are accepted on PE-1, which has only BGP instance 1 and VXLAN instance 1 enabled, and the VNI is 11. The used EVPN MAC route for MAC address 00:ca:fe:ca:fe:06 has PE-2 (192.0.2.2) as next-hop. The second route for the same MAC address has PE-3 (192.0.2.3) as next-hop, but it is not preferred, so it is not used.

```
*A:PE-1# show router bgp routes evpn mac
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
```

```
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  64500:2311          00:ca:fe:ca:fe:06 ESI-0
      0                   Seq:0             VNI 11
                          n/a
                          192.0.2.2

*>i   64500:2311          00:ca:fe:ca:fe:06 ESI-0
      0                   Seq:0             VNI 11
                          n/a
                          192.0.2.3

-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

On PE-2, the following three EVPN MAC routes are accepted. The first route has PE-1 (192.0.2.1) as next-hop and is received in BGP instance 1, which corresponds to VXLAN 1 and VNI 11. The latter two routes are received in BGP instance 2 for VXLAN instance 2 with VNI 12. These routes both have RD 64500:4512 and the route with the lowest IP next-hop is preferred, so the route to PE-4 (192.0.2.4) is used. The EVPN MAC routes on PE-3 are similar.

```
*A:PE-2# show router bgp routes evpn mac
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  64500:111           00:ca:fe:ca:fe:01 ESI-0
      0                   Seq:0             VNI 11
                          n/a
                          192.0.2.1

u*>i  64500:4512          00:ca:fe:ca:fe:06 ESI-0
      0                   Seq:0             VNI 12
                          n/a
                          192.0.2.4

*>i   64500:4512          00:ca:fe:ca:fe:06 ESI-0
```

```
      0                   Seq:0            VNI 12
                          n/a
                          192.0.2.5

-------------------------------------------------------------------------------
Routes : 3
===============================================================================
```

The EVPN MAC routes on the nodes in DC-2 are also similar.

The following FDB for VPLS 1 on PE-2 shows that MAC address 00:ca:fe:ca:fe:01 is learned from an EVPN MAC route in VXLAN 1 from 192.0.2.1 (PE-1); MAC address 00:ca:fe:ca:fe:06 is learned in VXLAN 2 from 192.0.2.4 (PE-4). For routes with the same RD but different next-hops, the router processes only the route with the lowest IP next-hop.

```
*A:PE-2# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:ca:fe:ca:fe:01 vxlan-1:                Evpn     08/12/21 14:55:03
                             192.0.2.1:11
1          00:ca:fe:ca:fe:06 vxlan-2:                Evpn     08/12/21 14:55:14
                             192.0.2.4:12
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

Conclusion

With dual EVPN-VXLAN instance VPLS services, service providers can deploy DCI scenarios with end-to-end VXLAN.

Domain Path Attribute for VPRN BGP Routes

This chapter provides information about the domain path attribute for VPRN BGP routes.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

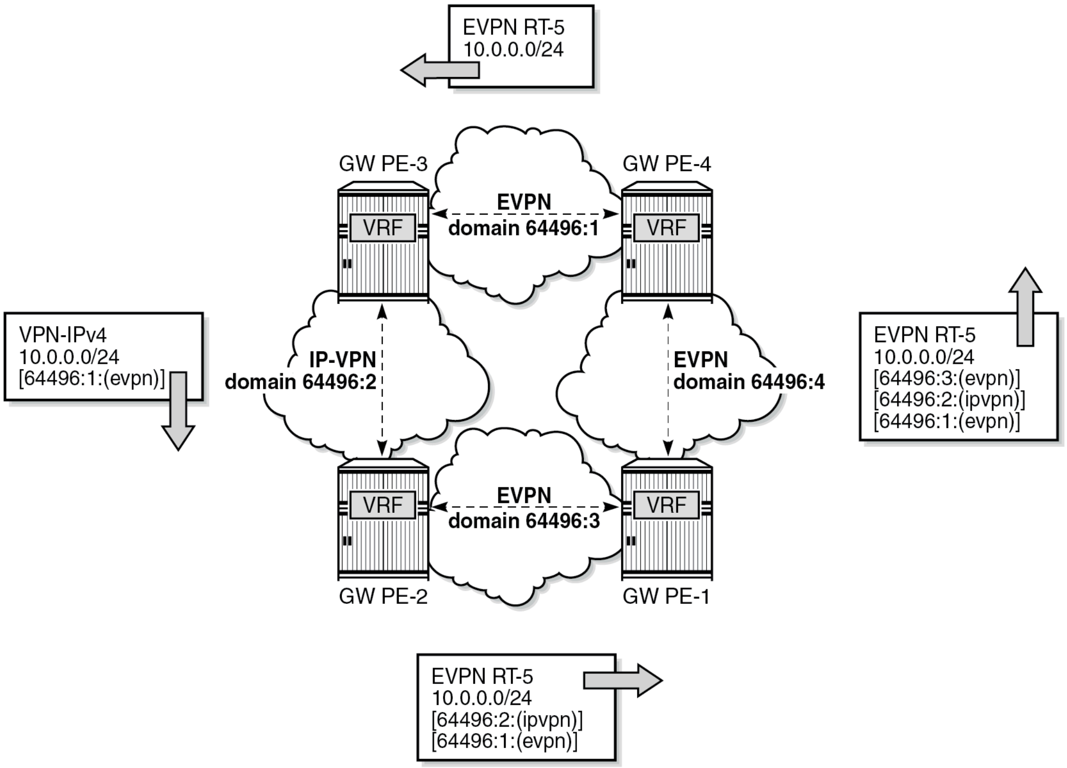
The information and configuration in this chapter are based on SR OS Release 22.7.R1. The domain path (D-path) attribute is supported in SR OS Release 21.10.R1 and later.

Overview

The D-path attribute can be used for route traceability, BGP best path selection, and loop prevention in networks that expand multiple IP-VPN and EVPN domains.

The D-path attribute is a sequence of domain segments, where each domain segment is represented by a domain ID in combination with an inter-subnet forwarding (ISF) subaddress family indicator (SAFI). The D-path attribute is added or modified by gateways (GWs) that import BGP-EVPN route type 5 (RT-5) or IP-VPN routes into a VPRN route table and export these prefixes as RT-5 or IP-VPN routes to their neighbors. Any PE that imports a prefix route does not install the route in the VPRN route table if the D-path attribute contains a domain segment where the domain ID matches a local domain ID, as shown in Figure 44: Loop prevention in networks with multiple IP-VPN and EVPN domains.

Figure 44: Loop prevention in networks with multiple IP-VPN and EVPN domains

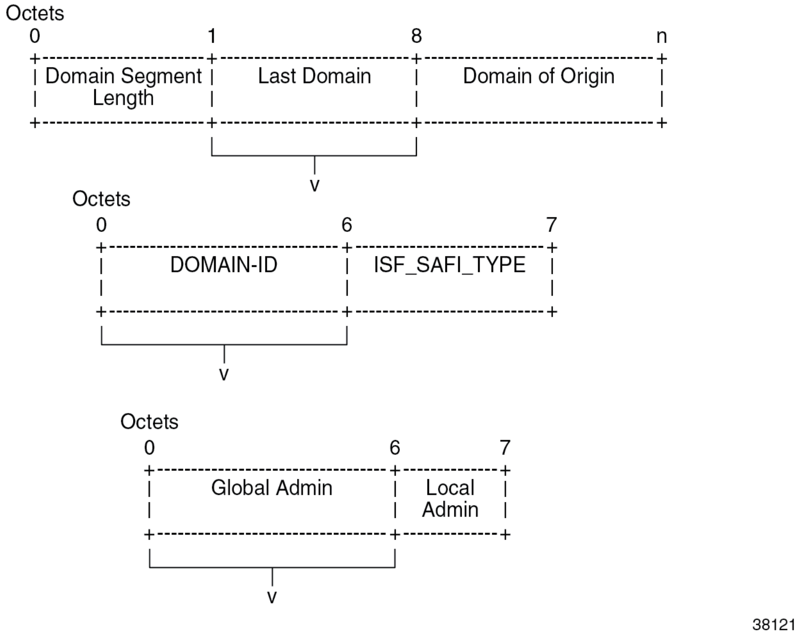


38120

All PEs in Figure 44: Loop prevention in networks with multiple IP-VPN and EVPN domains are GWs. PE-4 exports local prefix 10.0.0.0/24 as an EVPN RT-5 route without the D-path attribute when no domain ID is configured for local routes. PE-3 accepts this route. Domain ID 64496:1 is defined in PE-4 and PE-3, but the domain segment 64496:1:(evpn) is only added by GW PE-3 where the prefix is exported as an IP-VPN route instead of an EVPN RT-5 route. GW PE-2 accepts this route and modifies the D-path attribute by prepending domain segment 64496:2:(ipvpn) when exporting prefix 10.0.0.0/24 as an EVPN RT-5 route. PE-1 accepts this route. When PE-1 exports the prefix as an EVPN RT-5 route to PE-4, it prepends domain segment 64496:3:(evpn) to the D-path attribute. The VRF on PE-4 cannot import this prefix because the D-path attribute contains domain ID 64496:1, which is defined on PE-4.

Figure 45: D-path attribute shows the D-path attribute as defined in *draft-ietf-bess-evpn-ipvpn-interworking*.

Figure 45: D-path attribute



The D-path attribute is composed of a sequence of domain segments. Each domain segment consists of a domain ID and a SAFI type. The domain ID represents the domain and is composed of a 4-octet global administrator subfield and a 2-octet local administrator subfield. The global administrator subfield must have a value that is unique for the domain; for example, an autonomous system number (ASN). The 1-octet SAFI field can have the following values:

- 0 for local ISF routes
- 1 for PE-CE BGP domains
- 70 for EVPN domains
- 128 for IP-VPN domains

The domain ID can be configured on:

- VPRN BGP-EVPN MPLS and BGP-EVPN SRv6 instances (EVPN interface-less (EVPN-IFL))
- VPRN BGP-IPVPN MPLS and BGP-IPVPN SRv6 instances
- R-VPLS BGP-EVPN MPLS and BGP-EVPN VXLAN instances (EVPN interface-ful (EVPN-IFF))
- VPRN BGP neighbors (PE-CE)
- VPRN level (for local routes). When configured on the VPRN level, using the optional **local-routes-domain-id** command, the PE advertises its direct, static, or IGP routes with a D-path attribute.

Domain IDs can be modified while the service is operational. Modifying the domain ID initiates a route refresh for all address families associated with the VPRN.

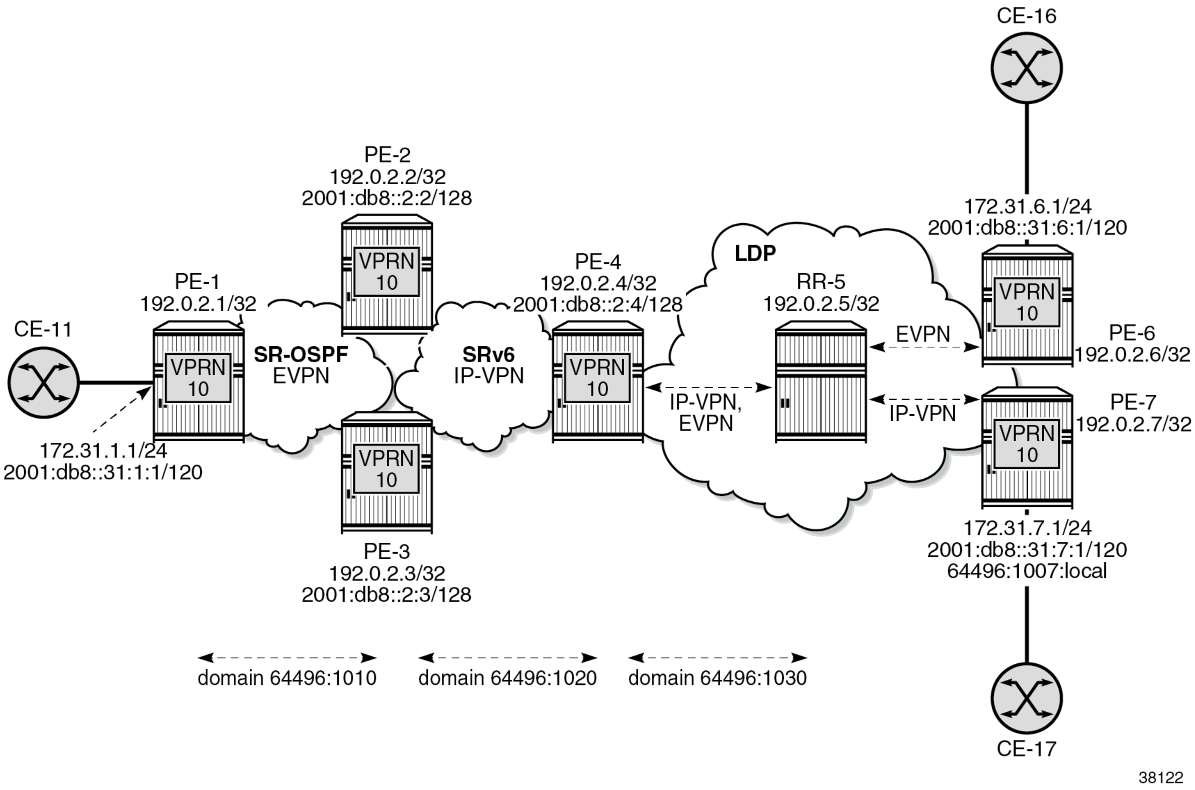
A PE receiving a prefix route with a D-path attribute containing one of its own domain IDs detects a routing loop and does not install the route in the VPRN route table.

The D-path attribute length can influence the BGP best path selection. In the BGP decision process, the shorter D-path is preferred, unless the **d-path-length-ignore** command is configured.

Configuration

Figure 46: Example topology with VPRN 10 and its domain IDs shows an example topology where PE-6 exports EVPN RT-5 routes 172.31.6.0/24 and 2001:db8::31:6:0/120 to route reflector RR-5, whereas PE-7 exports IP-VPN routes 172.31.7.0/24 and 2001:db8::31:7:0/120 to RR-5. LDP tunnels are used between PE-4, RR-5, PE-6, and PE-7; SRv6 tunnels are used between PE-2, PE-3, and PE-4; SR-OSPF tunnels are used between PE-1, PE-2, and PE-3.

Figure 46: Example topology with VPRN 10 and its domain IDs



The initial configuration includes:

- cards, MDAs, ports
- router interfaces
- OSPF as IGP on PE-1, PE-2, and PE-3
- IS-IS as IGP on PE-2, PE-3, PE-4, RR-5, PE-6, and PE-7
- SR-OSPF on PE-1, PE-2, and PE-3
- SRv6 on PE-2, PE-3, and PE-4, configured as in chapter "Segment Routing over IPv6" in the *7750 SR and 7950 XRS Segment Routing and PCE Advanced Configuration Guide for Classic CLI*.
- LDP on PE-4, RR-5, PE-6, and PE-7

The BGP configuration on PE-1 is as follows:

```
# on PE-1:
configure
    router Base
        autonomous-system 64496
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal1"
                family evpn
                type internal
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
```

```
# on PE-2 (similar configuration on PE-3):
configure
    router Base
        autonomous-system 64496
        bgp
            vpn-apply-import
            vpn-apply-export
            router-id 192.0.2.2        # on PE-3: 192.0.2.3
            advertise-inactive
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update vpn-ipv4 vpn-ipv6 evpn
            group "internal1"
                family evpn
                next-hop-self
                type internal
                local-address 192.0.2.2        # on PE-3: 192.0.2.3
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.3        # on PE-3: 192.0.2.2
                exit
            exit
            group "internal2"
                family vpn-ipv4 vpn-ipv6
                next-hop-self
                type internal
                local-address 2001:db8::2:2        # on PE-3: 2001:db8::2:3
                extended-nh-encoding ipv4 vpn-ipv4
                advertise-ipv6-next-hops vpn-ipv4 vpn-ipv6
                neighbor 2001:db8::2:3        # on PE-3: 2001:db8::2:2
                exit
                neighbor 2001:db8::2:4
                exit
            exit
```

```
# on PE-4:
configure
    router Base
        autonomous-system 64496
        bgp
            vpn-apply-import
            vpn-apply-export
            router-id 192.0.2.4
            advertise-inactive
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update vpn-ipv4 vpn-ipv6 evpn
            group "internal2"
                family vpn-ipv4 vpn-ipv6 evpn
                next-hop-self
                type internal
                local-address 2001:db8::2:4
                extended-nh-encoding ipv4 vpn-ipv4
                advertise-ipv6-next-hops vpn-ipv4 vpn-ipv6
                neighbor 2001:db8::2:2
                exit
                neighbor 2001:db8::2:3
                exit
            exit
            group "internal3"
                family vpn-ipv4 vpn-ipv6 evpn
                next-hop-self
                type internal
                local-address 192.0.2.4
                neighbor 192.0.2.5
                exit
            exit
```

```
# on RR-5: only EVPN toward PE-6; only IP-VPN toward PE-7:
configure
    router Base
        autonomous-system 64496
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update vpn-ipv4 vpn-ipv6 evpn
            group "internal3"
                type internal
                cluster 192.0.2.5
                neighbor 192.0.2.4
                    family vpn-ipv4 vpn-ipv6 evpn
                exit
                neighbor 192.0.2.6
                    family evpn
                exit
                neighbor 192.0.2.7
                    family vpn-ipv4 vpn-ipv6
                exit
            exit
```

```
# on PE-6:
configure
    router Base
        autonomous-system 64496
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal3"
                type internal
                neighbor 192.0.2.5
                    family evpn
                exit
            exit
```

```
# on PE-7:
configure
    router Base
        autonomous-system 64496
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update vpn-ipv4 vpn-ipv6
            group "internal3"
                type internal
                neighbor 192.0.2.5
                    family vpn-ipv4 vpn-ipv6
                exit
            exit
```

Domain IDs in VPRN BGP-EVPN MPLS and SRv6 instances

On PE-1, VPRN 10 is configured without domain ID in the **bgp-evpn mpls** context:

```
# on PE-1:
configure
    service
        vprn 10 name "VPRN 10" customer 1 create
            autonomous-system 64496
            interface "int-PE-1-CE-11" create
                address 172.31.1.1/24
                ipv6
                    address 2001:db8::31:1:1/120
                exit
                sap 1/1/c5/1:10 create
                exit
            exit
            bgp-evpn
                mpls
                    auto-bind-tunnel
                        resolution-filter
                            sr-ospf
                        exit
                        resolution filter
                    exit
                    route-distinguisher 192.0.2.1:10
                    vrf-target target:64496:10
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

Domain ID 64496:1010 is configured in the **bgp-evpn mpls** context on GWs PE-2 and PE-3, whereas domain ID 64496:1020 is configured in the **bgp-ipvpn segment-routing-v6** context on PE-2, PE-3, and PE-4. Domain ID 64496:1030 is configured for IP-VPN and for BGP-EVPN on PE-4.

On PE-2, VPRN 10 is configured as follows. The configuration on PE-3 is similar.

```
# on GW PE-2:
configure
    service
         vprn 10 name "VPRN 10" customer 1 create
            autonomous-system 64496
            segment-routing-v6 1 create
                locator "PE-2_loc"                       # on PE-3:"PE-3_loc"
                    function
                        end-dt4
                        end-dt6
                    exit
                exit
            exit
            bgp-ipvpn
                segment-routing-v6
                    domain-id 64496:1020
                    route-distinguisher 192.0.2.2:16     # on PE-3: 192.0.2.3:16
                    srv6-instance 1 default-locator "PE-2_loc"  # on PE-3:"PE-3_loc"
                    source-address 2001:db8::2:2         # on PE-3: 2001:db8::2:3
                    vrf-target target:64496:10
                    no shutdown
                exit
            exit
            bgp-evpn
                mpls
                    auto-bind-tunnel
                        resolution-filter
                            sr-ospf
                        exit
                        resolution filter
                    exit
                    domain-id 64496:1010
                    route-distinguisher 192.0.2.2:10     # on PE-3: 192.0.2.3:10
                    vrf-target target:64496:10
                    no shutdown
                exit
            exit
            no shutdown
```

On GW PE-4, VPRN 10 is configured with two domain IDs: domain ID 1020 for IP-VPN over SRv6 and domain ID 1030 for IP-VPN over MPLS and for EVPN over MPLS.

```
# on GW PE-4:
configure
    service
        vprn 10 name "VPRN 10" customer 1 create
            autonomous-system 64496
            segment-routing-v6 1 create
                locator "PE-4_loc"
                    function
                        end-dt4
                        end-dt6
                    exit
                exit
            exit
            bgp-ipvpn
```

```
                mpls
                    auto-bind-tunnel
                        resolution-filter
                            ldp
                        exit
                        resolution filter
                    exit
                    domain-id 64496:1030
                    route-distinguisher 192.0.2.4:10
                    vrf-target target:64496:10
                    no shutdown
                exit
                segment-routing-v6
                    domain-id 64496:1020
                    route-distinguisher 192.0.2.4:16
                    srv6-instance 1 default-locator "PE-4_loc"
                    source-address 2001:db8::2:4          ## system IP@
                    vrf-target target:64496:10
                    no shutdown
                exit
            exit
            bgp-evpn
                mpls
                    auto-bind-tunnel
                        resolution-filter
                            ldp
                        exit
                        resolution filter
                    exit
                    domain-id 64496:1030
                    route-distinguisher 192.0.2.4:10
                    vrf-target target:64496:10
                    no shutdown
                exit
            exit
            allow-export-bgp-vpn
            no shutdown
```

For completeness, the configuration on VPRN 10 on PE-6 and PE-7 is also shown. PE-6 has no domain ID configured:

```
# on PE-6:
configure
    service
        vprn 10 name "VPRN 10" customer 1 create
            autonomous-system 64496
            interface "int-PE-6-CE-16" create
                address 172.31.6.1/24
                ipv6
                    address 2001:db8::31:6:1/120
                exit
                sap 1/1/c5/1:10 create
                exit
            exit
            bgp-evpn
                mpls
                    auto-bind-tunnel
                        resolution-filter
                            ldp
                        exit
                        resolution filter
                    exit
                    route-distinguisher 192.0.2.6:10
                    vrf-target target:64496:10
                    no shutdown
                exit
            exit
            no shutdown
```

PE-7 does not have a domain ID configured in the **bgp-ipvpn mpls** context, but it has a local domain ID configured: 64496:1007:

```
# on PE-7:
configure
    service
        vprn 10 name "VPRN 10" customer 1 create
            local-routes-domain-id 64496:1007
            autonomous-system 64496
            interface "int-PE-7-CE-17" create
                address 172.31.7.1/24
                ipv6
                    address 2001:db8::31:7:1/120
                exit
                sap 1/1/c5/1:10 create
                exit
            exit
            bgp-ipvpn
                mpls
                    auto-bind-tunnel
                        resolution-filter
                            ldp
                        exit
                        resolution filter
                    exit
                    route-distinguisher 192.0.2.7:10
                    vrf-target target:64496:10
                    no shutdown
                exit
            exit
            no shutdown
```

The following commands on PE-4 display the domain ID for BGP-IPVPN and BGP-EVPN. For BGP-IPVPN, domain ID 64496:1030 is configured in the EVPN-MPLS domain and domain ID 64496:1020 is configured in the SRv6 domain:

```
*A:PE-4# show service id 10 bgp-ipvpn

===============================================================================
Service 10 BGP-IPVPN MPLS Information
===============================================================================
Admin State       : Up
VRF Import        : None
VRF Export        : None
Route Dist.       : None
Oper Route Dist   : 192.0.2.4:10
Oper RD Type      : configured
Route Target      : target:64496:10
Route Target Impor: None
Route Target Expor: None
Domain-Id         : 64496:1030
Dyn Egr Lbl Limit : Disabled

Auto-Bind Tunnel
Resolution        : disabled            Strict Tnl Tag    : False
ECMP              : 0                   Flex Algo FB      : False
```

```
Weighted ECMP     : False
BGP Instance      : 1
Filter Tunnel Type: (Not Specified)
===============================================================================

===============================================================================
Service 10 BGP-IPVPN Segment-Routing-V6 Information
===============================================================================

Admin State       : Up
VRF Import        : None
VRF Export        : None
Route Dist.       : 192.0.2.4:16
Oper Route Dist   : 192.0.2.4:16
Oper RD Type      : configured
Route Target      : target:64496:10
Route Target Expor: None
Route Target Impor: None
Def Route Tag     : 0x0
Route Resolution  : route-table

Srv6 Instance     : 1
Default Locator   : PE-4_loc
Source Address    : 2001:db8::2:4
Domain-Id         : 64496:1020

===============================================================================
```

For BGP-EVPN, domain ID 64496:1030 is configured in the EVPN-MPLS domain:

```
*A:PE-4# show service id 10 bgp-evpn

===============================================================================
BGP EVPN MPLS Table
===============================================================================
Admin State        : Up
VRF Import         : None
VRF Export         : None
Route Dist.        : 192.0.2.4:10
Oper Route Dist.   : 192.0.2.4:10
Oper RD Type       : configured
Route Target       : target:64496:10
Route Target Import: None
Route Target Export: None
Default Route Tag  : None
Domain-Id          : 64496:1030
Dyn Egr Lbl Limit  : Disabled

Advertise          : Disabled
Weighted ECMP      : Disabled

Auto-Bind Tunnel
Resolution         : filter                Strict Tnl Tag : False
ECMP               : 1                     Flex Algo FB   : False
BGP Instance       : 1
Filter Tunnel Types: ldp

Tunnel Encap
MPLS               : True                  MPLSoUDP       : False
===============================================================================
```

VPRN BGP routes for prefix 172.31.6.0/24

PE-6 advertises prefix 172.31.6.0/24 as an EVPN-IFL route without the D-path attribute, as follows:

```
# on PE-6:
1 2022/09/05 14:07:07.846 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 82
    Flag: 0x90 Type: 14 Len: 45 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.6
        Type: EVPN-IP-PREFIX Len: 34 RD: 192.0.2.6:10, ESI: ESI-0, tag: 0, ip_prefix:
 172.31.6.0/24 gw_ip 0.0.0.0 Label: 8388528 (Raw Label: 0x7fffb0)
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64496:10
        bgp-tunnel-encap:MPLS
```

RR-5 forwards prefix 172.31.6.0/24 as an EVPN-IFL route without the D-path attribute, as follows:

```
# on RR-5:
34 2022/09/05 14:07:11.660 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 156
    Flag: 0x90 Type: 14 Len: 105 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.6
        Type: EVPN-IP-PREFIX Len: 34 RD: 192.0.2.6:10, ESI: ESI-0, tag: 0, ip_prefix:
 172.31.6.0/24 gw_ip 0.0.0.0 Label: 8388528 (Raw Label: 0x7fffb0)
        Type: EVPN-IP-PREFIX Len: 58 RD: 192.0.2.6:10, ESI: ESI-0, tag: 0, ip_prefix:
 2001:db8::31:6:0/120 gw_ip :: Label: 8388528 (Raw Label: 0x7fffb0)
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.6
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        192.0.2.5
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64496:10
        bgp-tunnel-encap:MPLS
"
```

PE-4 adds a D-path attribute when advertising prefix 172.31.6.0/24 as a VPN-IPv4 route to PE-2 (or PE-3):

```
53 2022/09/05 14:07:11.662 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:2
"Peer 1: 2001:db8::2:2: UPDATE
Peer 1: 2001:db8::2:2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 98
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family VPN_IPV4
        NextHop len 24 NextHop 2001:db8::2:4
        172.31.6.0/24 RD 192.0.2.4:10 Label 524280 (Raw label 0x7fff81)
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.6
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        192.0.2.5
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64496:10
    Flag: 0xc0 Type: 36 Len: 8 D-PATH:[64496:1030:(evpn)]
"
```

PE-2 prepends domain segment 64496:1020:(ipvpn) to the D-path attribute when advertising prefix 172.31.6.0/24 in an EVPN-IFL route to PE-1:

```
# on PE-2:
40 2022/09/05 14:07:11.662 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 115
    Flag: 0x90 Type: 14 Len: 45 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-IP-PREFIX Len: 34 RD: 192.0.2.2:10, ESI: ESI-0, tag: 0, ip_prefix:
 172.31.6.0/24 gw_ip 0.0.0.0 Label: 8388528 (Raw Label: 0x7fffb0)
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.6
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        192.0.2.5
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64496:10
        bgp-tunnel-encap:MPLS
    Flag: 0xc0 Type: 36 Len: 16 D-PATH:[64496:1020:(ipvpn)][64496:1030:(evpn)]
"
```

Figure 47: VPRN BGP routes for prefix 172.31.6.0/24 shows the D-path attribute in the BGP routes for prefix 172.31.6.0/24:

Figure 47: VPRN BGP routes for prefix 172.31.6.0/24

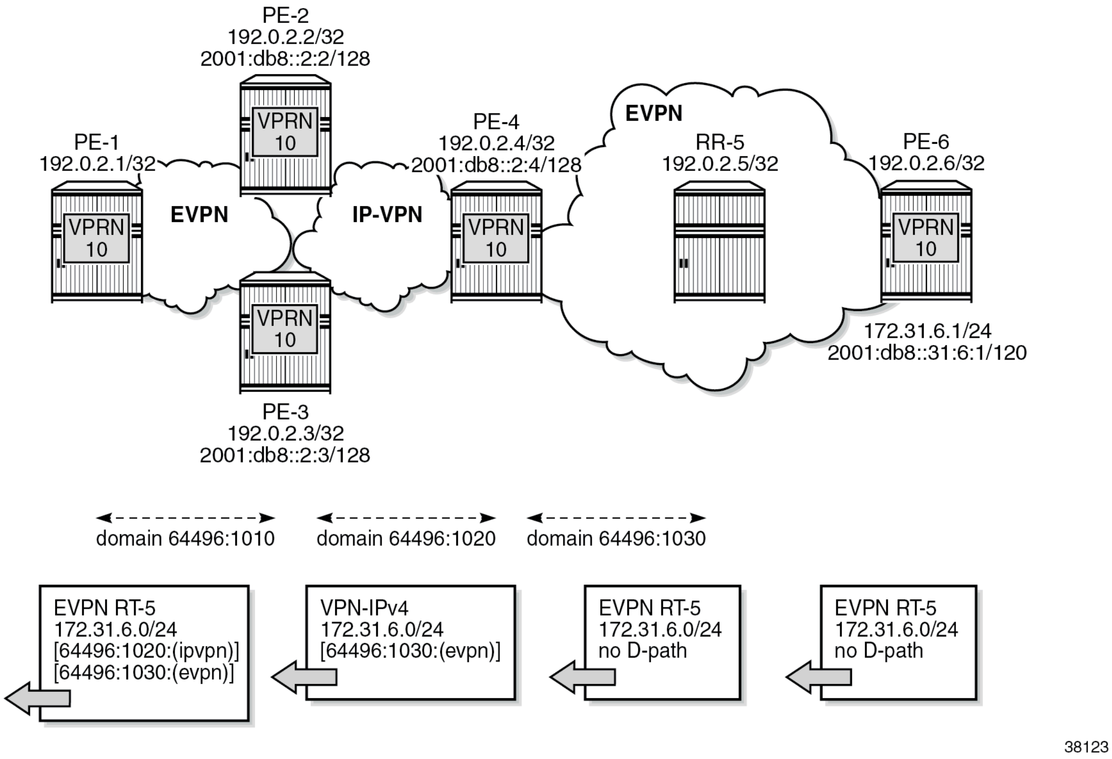
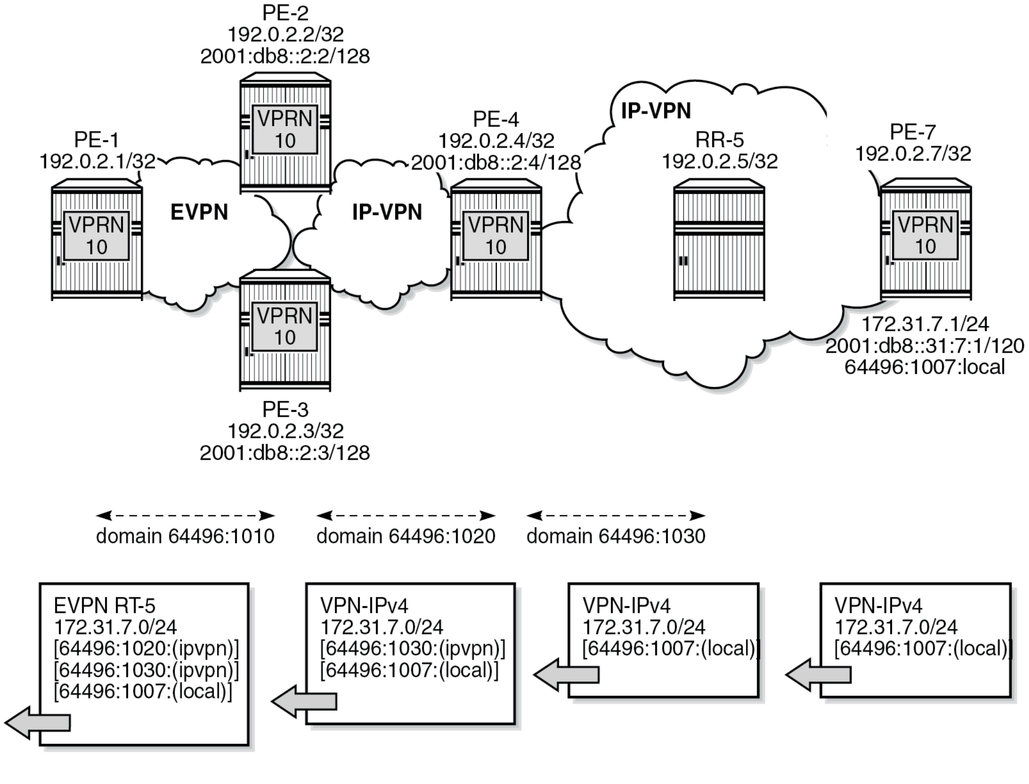


Figure 48: VPRN BGP routes for prefix 172.31.7.0/24 similarly shows the D-path attribute in the BGP routes for prefix 172.31.7.0/24:

Figure 48: VPRN BGP routes for prefix 172.31.7.0/24



In VPRN 10 on PE-6, no local domain ID is configured, whereas in VPRN 10 on PE-7, the local domain ID 64496:1007 is configured for the routes local to PE-7.

The following BGP update shows that PE-7 advertises prefix 172.31.7.0/24 as a VPN-IPv4 route with a D-path attribute containing the domain segment 64496:1007:(local).

```
# on PE-7:
1 2022/09/05 14:07:07.879 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 72
    Flag: 0x90 Type: 14 Len: 32 Multiprotocol Reachable NLRI:
        Address Family VPN_IPV4
        NextHop len 12 NextHop 192.0.2.7
        172.31.7.0/24 RD 192.0.2.7:10 Label 524283 (Raw label 0x7fffb1)
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64496:10
    Flag: 0xc0 Type: 36 Len: 8 D-PATH:[64496:1007:(local)]
"
```

RR-5 advertises prefix 172.31.7.0/24 as a VPN-IPv4 route with the same D-path attribute. PE-4 prepends the domain segment 64496:1030:(ipvpn) to the D-path attribute of the VPN-IPv4 routes for prefix

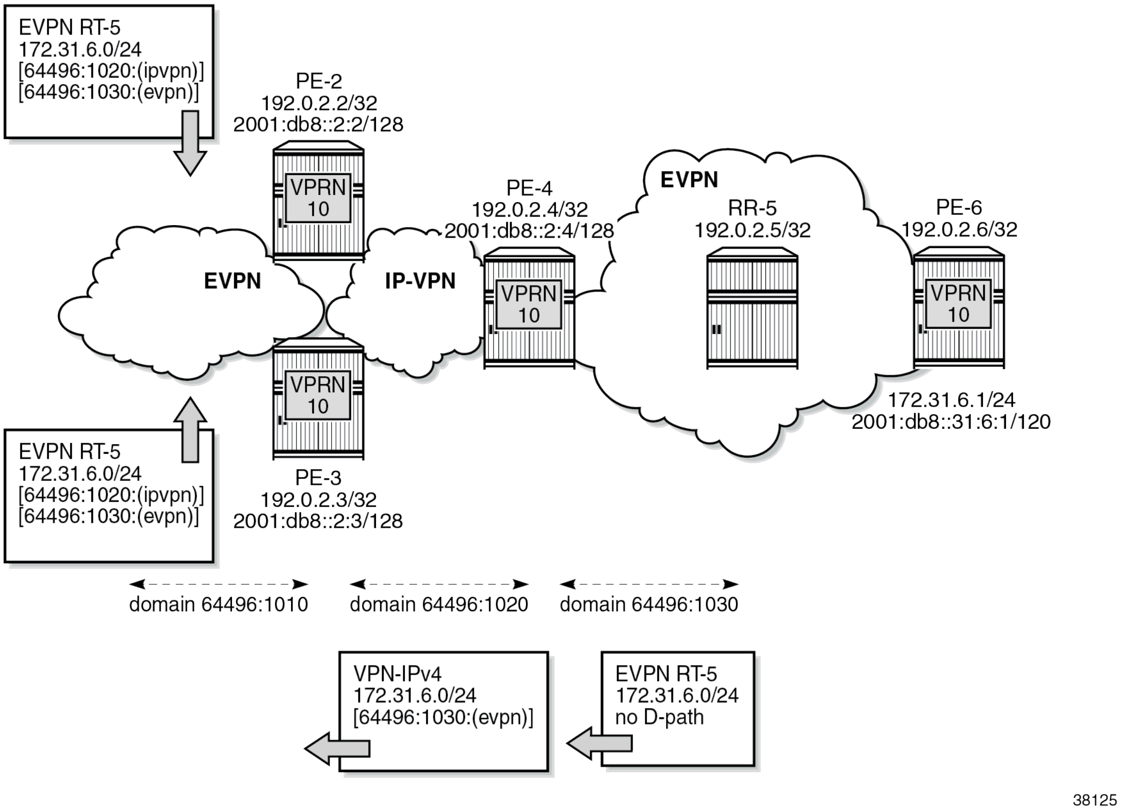
172.31.7.0/24 to PE-2 (and PE-3). PE-2 advertises prefix 172.31.7.0/24 as an EVPN-IFL route to PE-1 with domain segment 64496:1020:(ipvpn) added to the D-path attribute:

```
# on PE-2:
41 2022/09/05 14:07:11.662 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 123
    Flag: 0x90 Type: 14 Len: 45 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-IP-PREFIX Len: 34 RD: 192.0.2.2:10, ESI: ESI-0, tag: 0, ip_prefix:
 172.31.7.0/24 gw_ip 0.0.0.0 Label: 8388528 (Raw Label: 0x7fffb0)
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.7
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        192.0.2.5
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64496:10
        bgp-tunnel-encap:MPLS
    Flag: 0xc0 Type: 36 Len: 24 D-PATH:[64496:1020:(ipvpn)][64496:1030:(ipvpn)][64496:1007:
(local)]
"
```

Loop prevention

Besides traceability, the D-path attribute provides loop prevention in the control plane. Redundant GWs PE-2 and PE-3 cause routing loops and the D-path attribute helps preventing these loops. When PE-2 receives the EVPN-IFL route from PE-3 with a D-path containing domain IDs configured on PE-2, such as 64496:1020, it does not install the route in the VPRN route table, as shown in Figure 49: Loop prevention between PE-2 and PE-3:

Figure 49: Loop prevention between PE-2 and PE-3



The following command on PE-2 shows that in the EVPN-IFL route for prefix 172.31.6.0/24 that was received from PE-3, a D-path loop has been detected in VPRN 10:

```
*A:PE-2# show router bgp routes evpn ip-prefix prefix 172.31.6.0/24 hunt
===============================================================================
 BGP Router ID:192.0.2.2        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : n/a
Nexthop        : 192.0.2.3
Path Id        : None
From           : 192.0.2.3
Res. Nexthop   : 192.168.23.2
Local Pref.    : 100                    Interface Name : int-PE-2-PE-3
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
```

```
Community      : target:64496:10 bgp-tunnel-encap:MPLS
Cluster        : 192.0.2.5
Originator Id  : 192.0.2.6              Peer Router Id : 192.0.2.3
Flags          : Valid Best IGP
Route Source   : Internal
AS-Path        : No As-Path
D-Path         : [64496:1020:(ipvpn)][64496:1030:(evpn)]
EVPN type      : IP-PREFIX
ESI            : ESI-0
Tag            : 0
Gateway Address: 00:00:00:00:00:00
Prefix         : 172.31.6.0/24
Route Dist.    : 192.0.2.3:10
MPLS Label     : LABEL 524283
Route Tag      : 0
Neighbor-AS    : n/a
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h24m27s
DPath Loop VRFs:  10
---snip---
```

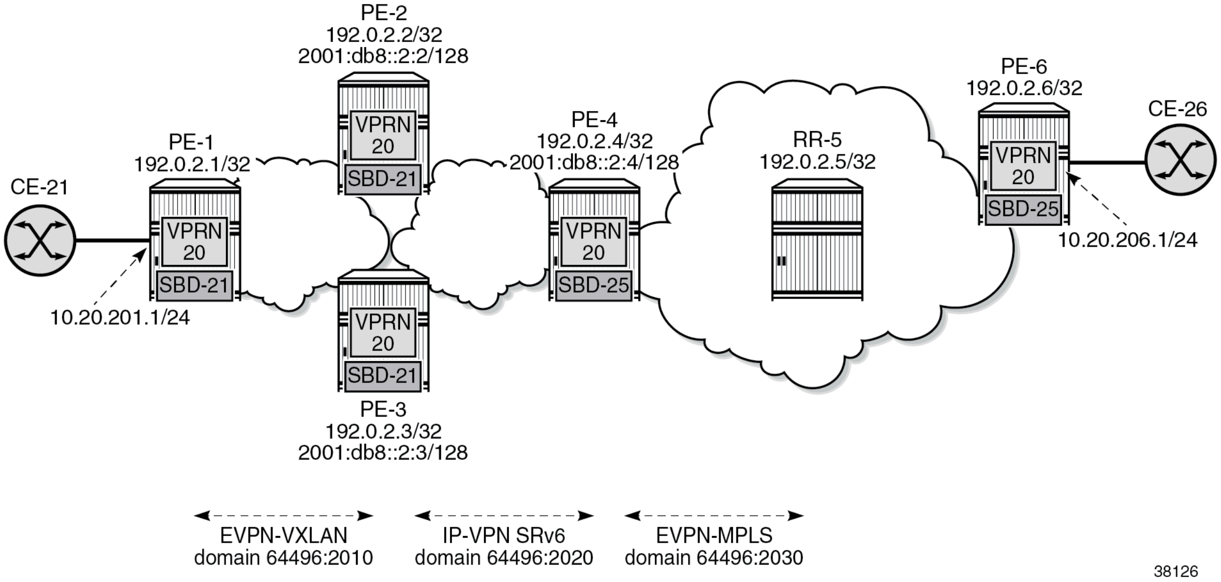
The preceding EVPN-IFL route from PE-3 for prefix 172.31.6.0/24 is not installed in the VPRN route table and is not forwarded to other PEs. The route table for VPRN 10 on PE-2 only has an IP-VPN route for prefix 172.31.6.0/24 with next hop PE-4:

```
*A:PE-2# show router 10 route-table

===============================================================================
Route Table (Service: 10)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
172.31.1.0/24                                 Remote  EVPN-IFL  00h26m24s  170
       192.0.2.1 (tunneled:SR-OSPF:524290)                          10
172.31.6.0/24                                 Remote  BGP VPN   00h26m24s  170
       2001:db8:aaaa:104:7fff:b000:: (tunneled:SRV6)                20
172.31.7.0/24                                 Remote  BGP VPN   00h26m24s  170
       2001:db8:aaaa:104:7fff:b000:: (tunneled:SRV6)                20
-------------------------------------------------------------------------------
No. of Routes: 3
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

Domain IDs in R-VPLS BGP-EVPN MPLS and BGP-EVPN VXLAN instances

Loops can also be prevented in Layer 3 EVPN data center gateway (DC GW) scenarios where EVPN-IFF routes are translated into IP-VPN routes, and vice versa. Because redundant GWs are used, the scenario is subject to Layer 3 routing loops and the D-path attribute helps preventing these loops without the need for extra routing policies to tag or drop routes. Figure 50: Example topology with R-VPLS shows a slightly modified example topology with R-VPLS with PE-2 and PE-3 acting as redundant DC GWs. PE-1 advertises an EVPN-IFF route for prefix 10.20.201.0/24 and PE-6 advertises an EVPN-IFF route for prefix 10.20.206.0/24.

Figure 50: Example topology with R-VPLS



The service configuration on PE-1 does not include a domain ID, as follows:

```
# on PE-1:
configure
    service
        vprn 20 name "VPRN 20" customer 1 create
            autonomous-system 64496
            interface "int-SBD-21" create
                vpls "SBD-21"
                    evpn-tunnel
                exit
            exit
            interface "int-PE-1-CE-21" create
                address 10.20.201.1/24
                sap 1/1/c5/1:20 create
                exit
            exit
            no shutdown
        exit
        vpls 21 name "SBD-21" customer 1 create
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 1 create
            exit
            bgp
            exit
            bgp-evpn
                ip-route-advertisement
                evi 21
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
```

On DC GW PE-2, domain ID 64496:2010 is configured in VPLS "SBD-21" whereas domain ID 64496:2020 is configured in VPRN 20. The configuration on DC GW PE-3 is similar.

```
# on PE-2:
configure
    service
        vprn 20 name "VPRN 20" customer 1 create
            autonomous-system 64496
            interface "int-SBD-21" create
                vpls "SBD-21"
                    evpn-tunnel
                exit
            exit
            segment-routing-v6 1 create
                locator "PE-2_loc"        # on PE-3: "PE3_loc"
                    function
                        end-dt46
                    exit
                exit
            exit
            bgp-ipvpn
                segment-routing-v6
                    domain-id 64496:2020
                    route-distinguisher 192.0.2.2:26        # on PE-3; 192.0.2.3:26
                    srv6-instance 1 default-locator "PE-2_loc"  # on PE-3: "PE3_loc"
                    source-address 2001:db8::2:2            # on PE-3: 2001:db8::2:3
                    vrf-target target:64496:20
                    no shutdown
                exit
            exit
            no shutdown
        exit
        vpls 21 name "SBD-21" customer 1 create
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 1 create
            exit
            bgp
            exit
            bgp-evpn
                ip-route-advertisement domain-id 64496:2010
                evi 21
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
```

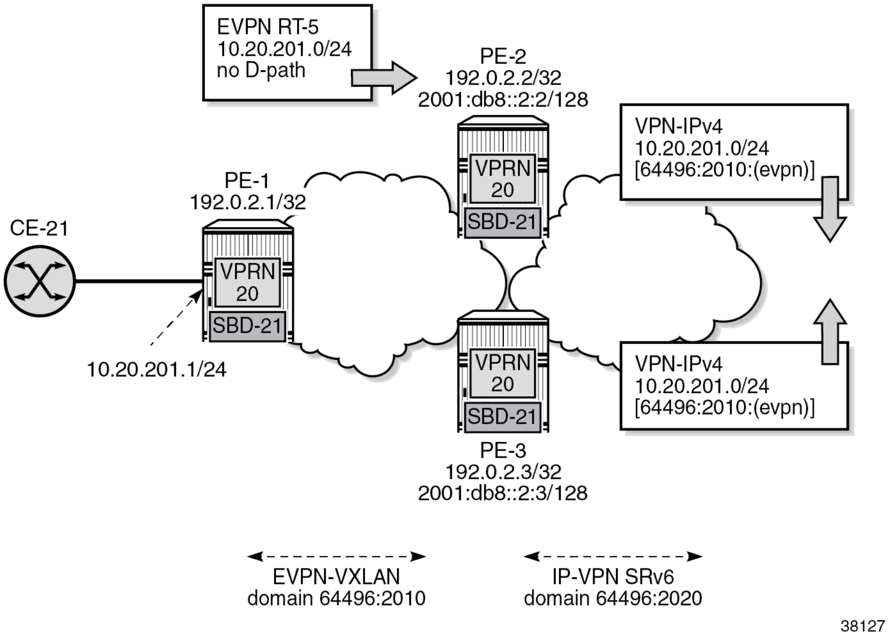
The service configuration examples for PE-1, PE-2, and PE-3 show how a loop is detected at the DC GWs in VPN-IPv4 routes for prefix 10.20.201.0/24 received from the other DC GW. The following command on DC GW PE-2 shows that a D-path loop is detected in VPRN 20 in a VPN-IPv4 route for prefix 10.20.201.0/24 received from DC GW PE-3:

```
*A:PE-2# show router bgp routes vpn-ipv4 rd 192.0.2.3:26 hunt
===============================================================================
 BGP Router ID:192.0.2.2        AS:64496       Local AS:64496
```

```
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP VPN-IPv4 Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 10.20.201.0/24
Nexthop        : 2001:db8::2:3
Route Dist.    : 192.0.2.3:26           VPN Label      : 524286
Path Id        : None
From           : 2001:db8::2:3
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : int-PE-2-PE-3
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64496:20
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.3
Fwd Class      : None                   Priority       : None
Flags          : Valid Best IGP
Route Source   : Internal
AS-Path        : No As-Path
D-Path         : [64496:2010:(evpn)]
Route Tag      : 0
Neighbor-AS    : n/a
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h07m49s
SRv6 TLV Type  : SRv6 L3 Service TLV (5)
SRv6 SubTLV    : SRv6 SID Information (1)
Sid            : 2001:db8:aaaa:103::
Full Sid       : 2001:db8:aaaa:103:7fff:e000::
Behavior       : End.DT46 (20)
SRv6 SubSubTLV : SRv6 SID Structure (1)
Loc-Block-Len  : 48                     Loc-Node-Len   : 16
Func-Len       : 20                     Arg-Len        : 0
Tpose-Len      : 20                     Tpose-offset   : 64
VPRN Imported  : None
DPath Loop VRFs:  20

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

Figure 51: Loop prevention between DC GW PE-2 and DC GW PE-3 shows that PE-1 sends an EVPN-IFF route for prefix 10.20.201.0/24 without D-path attribute to PE-2 and PE-3. Both PE-2 and PE-3 re-advertise prefix 10.20.201.0/24 as a VPN-IPv4 route with D-path attribute 64496:2010:(evpn). When PE-2 receives this VPN-IPv4 route from PE-3, it detects a loop based on the D-path attribute with domain segment 64496:2010:(evpn) and does not install the route in the VPRN route table. Likewise, PE-3 receives the VPN-IPv4 route from PE-2 and does not install it in the VPRN route table.

Figure 51: Loop prevention between DC GW PE-2 and DC GW PE-3



PE-2 does not use the VPN-IPv4 route for prefix 10.20.201.0/24 from PE-3. The VPRN route table on PE-2 contains the EVPN-IFF route received from PE-1 for prefix 10.20.201.0/24:

```
*A:PE-2# show router 20 route-table

===============================================================================
Route Table (Service: 20)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.20.201.0/24                                Remote  EVPN-IFF  00h18m36s  169
       int-SBD-21 (ET-02:0f:ff:ff:ff:52)                            0
10.20.206.0/24                                Remote  BGP VPN   00h18m36s  170
       2001:db8:aaaa:104:7fff:9000:: (tunneled:SRV6)                20
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

Conclusion

The D-path attribute provides traceability for VPRN BGP routes and can be used for BGP best path selection. The D-path attribute for VPRN routes also helps preventing loops without the need for dedicated routing policies to tag and drop routes.

Dual EVPN-MPLS Instance VPLS Services

This chapter provides information about the dual EVPN-MPLS instance VPLS services.

Topics in this chapter include:

Applicability

The information and configuration in this chapter are based on SR OS Release 22.10.R1. Dual EVPN-MPLS instance in VPLS is supported in SR OS Release 21.10.R1 and later.

Overview

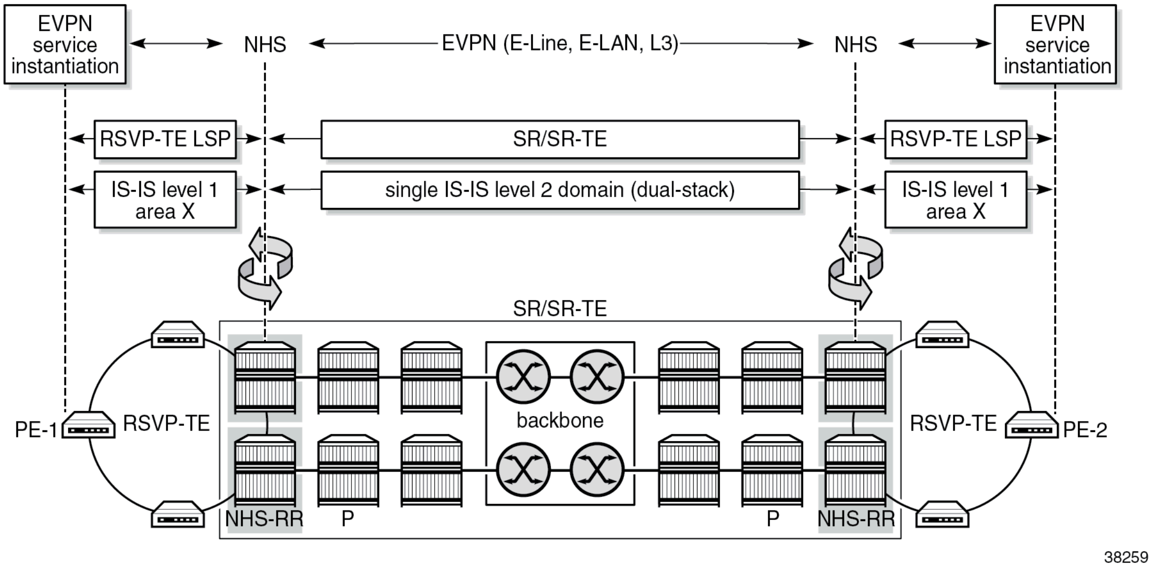
One of the scale issues that low-scale access nodes or leaf PEs face in high-scale architectures is the limited number of EVPN/IP-VPN next hops, tunnels, and service labels that they support.

The following solutions reduce the number of EVPN next hops exposed to the access nodes:

- inter-AS model B, as described in the "Inter-AS VPRN Model B" chapter in the *7450 ESS, 7750 SR, and 7950 XRS Layer 2 Services and EVPN Advanced Configuration Guide for Classic CLI*
- next-hop-self route reflectors (NHS-RRs)

Figure 52: Access nodes receive next hops from the NHS-RRs shows the NHS-RR solution reducing the number of EVPN next hops that are sent to the low-scale access nodes PE-1 and PE-2. Only the two NHS-RRs are exposed as next hops to PE-1.

Figure 52: Access nodes receive next hops from the NHS-RRs



The number of EVPN next hops is reduced, but the number of service labels to be learned is not. PE-1 still learns one service label per remote PE for each service it is attached to. In case of EVPN E-LAN services and broadcast, unknown unicast, and multicast (BUM) traffic, the ingress PE still needs one copy of every BUM packet per egress PE that exists in the remote domains, even if all the BUM traffic goes through one of the two NHS-RRs (or ASBRs in the case of model B).

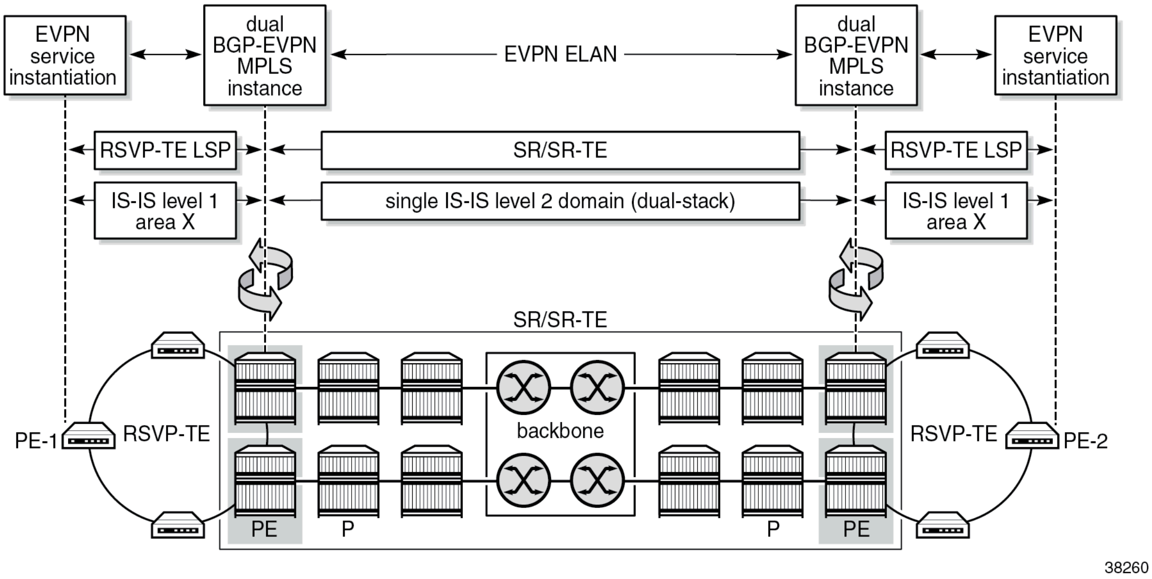
The following solutions reduce the number of service labels:

- VPRN services on the NHS-RRs with **allow-export-bgp-vpn** configured
- dual EVPN-MPLS instance VPLS services on the NHS-RRs

The **allow-export-bgp-vpn** command applies to VPRN services using EVPN-IFL, VPN-IPv4, and VPN-IPv6 families. Routes from the WAN are imported to the VPRN service and exported to the access nodes as new VPN-IP routes. The values of the service labels, route targets (RTs), and BGP next hops of the re-advertised routes are based on the configuration of the exporting VPRN.

Figure 53: Access nodes receive one service label per service from each NHS-RR shows a dual EVPN-MPLS instance VPLS service on the NHS-RRs, which offers a similar solution for EVPN-VPLS services to the **allow-export-bgp-vpn** solution for VPRN services. EVPN-MPLS routes received from the WAN are imported to the network EVPN-MPLS instance and redistributed to the access EVPN-MPLS instance with a new route distinguisher (RD), next hop, service label, and possibly a new RT. The ingress PE learns only one service label for each NHS-RR per service, as opposed to one service label per remote PE that is attached to the same EVPN service. With this solution, the replication of BUM traffic is also optimized because the ingress PE sends a single copy of each BUM packet to the NHS-RR, as opposed to one copy per egress PE.

Figure 53: Access nodes receive one service label per service from each NHS-RR



In the example, redundant NHS-RRs are used. Redundancy is handled via anycast multihoming, which implies that two or more PEs are configured with the same service parameters as part of the same redundancy group: identical route distinguishers and RTs per instance, and the same anycast IP address. The ingress PEs set up EVPN destinations to only one PE in the anycast group for a specific service. EVPN BUM destinations are not established between PEs in the same anycast group because the received anycast peer inclusive multicast Ethernet tag (IMET) routes have the same local originating IP address. In anycast multihoming scenarios, policies are required to prevent control-plane loops.

Configuration

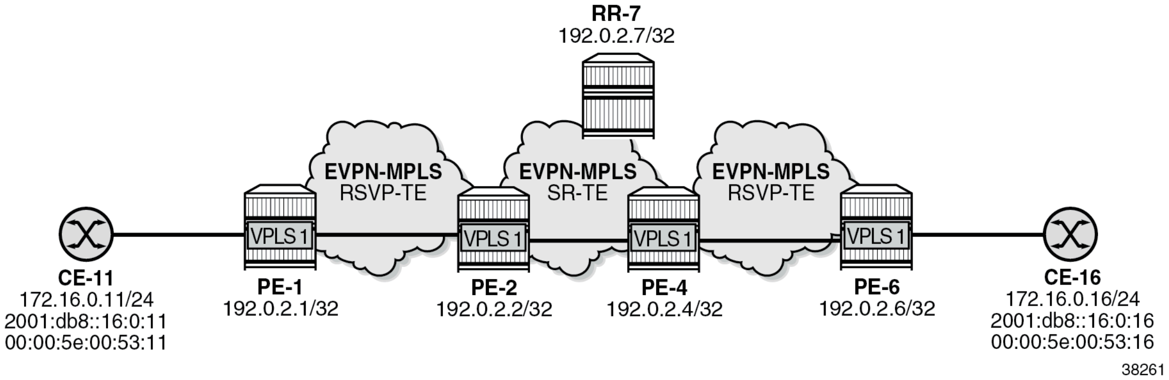
The following scenarios are described in this section:

- dual EVPN-MPLS instance VPLS without multihoming
- dual EPVN-MPLS instance VPLS with anycast multihoming

Dual EVPN-MPLS instance VPLS without multihoming

Figure 54: Example topology 1 shows EVPN-MPLS VPLS 1 configured on four PEs. PE-2 and PE-4 are EVPN gateways (GWs). RR-7 is the route reflector for PE-2 and PE-4 in the WAN network.

Figure 54: Example topology 1



The initial configuration includes:

- cards, MDAs, ports
- router interfaces
- IS-IS level 1 between PE-1 and PE-2 and between PE-4 and PE-6
- IS-IS level 2 between PE-2, PE-4, and RR-7
- SR-TE tunnels between PE-2 and PE-4
- MPLS LSPs between PE-1 and PE-2 and between PE-4 and PE-6

BGP is configured on all nodes for the EVPN address family. PE-1 peers with the dual-homed EVPN GW PE-2. In a similar way, PE-6 peers with EVPN GW PE-4. The BGP configuration on PE-1 is as follows:

```
# on PE-1:
configure
    router Base
        autonomous-system 64496
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "access1"
                family evpn
                peer-as 64496
                neighbor 192.0.2.2
                exit
            exit
```

EVPN GW PE-2 peers with PE-1 in BGP group "access1" and with RR-7 in BGP group "WAN":

```
# on PE-2:
configure
    router Base
        autonomous-system 64496
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "access1"
                family evpn
                next-hop-self
                cluster 192.0.2.2
                export "drop-tag-20"
                peer-as 64496
                neighbor 192.0.2.1
                exit
            exit
            group "WAN"
                family evpn
                next-hop-self
                export "drop-tag-10"
                peer-as 64496
                neighbor 192.0.2.7
                exit
            exit
```

The BGP configuration on PE-4 is similar. The export policies use tags to avoid loops in topologies with redundant EVPN GWs, as described in the section Dual EVPN-MPLS instance VPLS with anycast multihoming.

RR-7 peers with PE-2 and PE-4 in BGP group "WAN":

```
# on RR-7:
configure
    router Base
        autonomous-system 64496
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "WAN"
                family evpn
                cluster 192.0.2.7
                peer-as 64496
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.4
                exit
            exit
```

On PE-1, VPLS 1 is configured with a single EVPN-MPLS instance. The RD 192.0.2.1:1 for BGP 1 is auto-derived from the values for the IPv4 system address and the EVI. PE-1 imports and exports routes with RT 64496:101.

```
# on PE-1:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            bgp
                # route-distinguisher 192.0.2.1:1  # will be auto-derived
                route-target export target:64496:101 import target:64496:101
            exit
            bgp-evpn
```

```
                evi 1
                mpls bgp 1
                    auto-bind-tunnel
                        resolution-filter
                            rsvp
                        exit
                        resolution filter
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/1/c10/1:1 create
                no shutdown
            exit
            no shutdown
        exit
```

On PE-2, VPLS 1 is configured with two EVPN-MPLS instances: instance 1 is configured with multihoming mode access and instance 2 with the (default) multihoming mode network, as follows:

```
# on PE-2:
configure
    service
        system
            bgp-auto-rd-range 192.0.2.2 comm-val 2000 to 2999
        exit
        vpls 1 name "VPLS 1" customer 1  create
            description "dual BGP-EVPN MPLS instance VPLS 1"
            bgp
                # route-distinguisher 192.0.2.2:1    # will be auto-derived
                route-target export target:64496:101 import target:64496:101
            exit
            bgp 2
                route-distinguisher auto-rd    # different RD (must be configured)
                route-target export target:64496:100 import target:64496:100
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    mh-mode access
                    auto-bind-tunnel
                        resolution-filter
                            rsvp
                        exit
                        resolution filter
                    exit
                    no shutdown
                exit
                mpls bgp 2
                    # mh-mode network               # default MH mode
                    auto-bind-tunnel
                        resolution-filter
                            sr-te
                        exit
                        resolution filter
                    exit
                    no shutdown
                exit
            exit
            stp
```

```
                shutdown
            exit
            no shutdown
```

Note: The RD for BGP 1 can be auto-derived from the values for the IPv4 system address and the EVI, for example, 192.0.2.2:1 on PE-2. The RD for BGP 2 cannot be auto-derived from the values for the IPv4 system address and the EVI, because the RD for BGP 2 must be different from the RD for BGP 1, so it must be configured manually or with **auto-rd**.

On PE-4, the configuration is similar:

```
# on PE-4:
configure
    service
        system
            bgp-auto-rd-range 192.0.2.4 comm-val 2000 to 2999
        exit
        vpls 1 name "VPLS 1" customer 1 create
            description "dual BGP-EVPN MPLS instance VPLS"
            bgp
                # route-distinguisher 192.0.2.4:1    # will be auto-derived
                route-target export target:64496:102 import target:64496:102
            exit
            bgp 2
                route-distinguisher auto-rd    # different RD
                route-target export target:64496:100 import target:64496:100
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    mh-mode access
                    auto-bind-tunnel
                        resolution-filter
                            rsvp
                        exit
                        resolution filter
                    exit
                    no shutdown
                exit
                mpls bgp 2
                    # mh-mode network               # default MH mode
                    auto-bind-tunnel
                        resolution-filter
                            sr-te
                        exit
                        resolution filter
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
```

The following command on PE-2 shows BGP instances 1 and 2 in VPLS 1. RD 192.0.2.2:1 for BGP instance 1 is auto-derived from the IPv4 system address and the EVI; the RD for BGP instance 2 is configured with **auto-rd** and has the value 192.0.2.2:2000. The RT values are configured.

```
*A:PE-2# show service id 1 bgp
```

```
===============================================================================
BGP Information
===============================================================================
Bgp Instance         : 1
Vsi-Import           : None
Vsi-Export           : None
Route Dist           : None
Oper Route Dist      : 192.0.2.2:1
Oper RD Type         : derivedEvi
Rte-Target Import    : 64496:101            Rte-Target Export: 64496:101
Oper RT Imp Origin   : configured           Oper RT Import   : 64496:101
Oper RT Exp Origin   : configured           Oper RT Export   : 64496:101
ADV Service MTU      : -1

Bgp Instance         : 2
Vsi-Import           : None
Vsi-Export           : None
Route Dist           : auto-rd
Oper Route Dist      : 192.0.2.2:2000
Oper RD Type         : auto
Rte-Target Import    : 64496:100            Rte-Target Export: 64496:100
Oper RT Imp Origin   : configured           Oper RT Import   : 64496:100
Oper RT Exp Origin   : configured           Oper RT Export   : 64496:100
ADV Service MTU      : -1

PW-Template Id       : None
-------------------------------------------------------------------------------
===============================================================================
```

The following command on PE-2 shows EVPN destination 192.0.2.1 in EVPN-MPLS instance 1:

```
*A:PE-2# show service id 1 evpn-mpls instance 1

===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address                    Egr Label     Num.   Mcast Last Change
                               Transport:Tnl MACs         Sup BCast Domain
-------------------------------------------------------------------------------
192.0.2.1                      524286        1      bum   12/09/2022 09:59:58
                               rsvp:1                     No
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                       Num. Macs              Last Change
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

The following command on PE-2 shows EVPN destination 192.0.2.4 in EVPN-MPLS instance 2:

```
*A:PE-2# show service id 1 evpn-mpls instance 2

===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address                    Egr Label     Num.   Mcast Last Change
                               Transport:Tnl MACs         Sup BCast Domain
```

```
-------------------------------------------------------------------------------
192.0.2.4                      524282        1      bum   12/09/2022 10:00:04
                               sr-te:655362               No
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                       Num. Macs               Last Change
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

When traffic is sent between CE-11 and CE-16, MAC address 00:00:5e:00:53:11 of CE-11 is learned on the local SAP in VPLS 1 on PE-1 and MAC address 00:00:5e:00:53:16 of CE-16 is learned on the local SAP in VPLS 1 on PE-6. EVPN MAC routes are advertised to the BGP-EVPN peers.

The forwarding database (FDB) on PE-1 is as follows:

```
*A:PE-1# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:5e:00:53:11 sap:1/1/c10/1:1         L/0      12/09/22 10:06:17
1          00:00:5e:00:53:16 mpls-1:                 Evpn     12/09/22 10:06:17
                             192.0.2.2:524284
           rsvp:1
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The FDB on PE-2 shows that an EVPN MAC route is received in EVPN-MPLS instance 1 for address 00:00:5e:00:53:11 whereas an EVPN MAC route is received in EVPN-MPLS instance 2 for address 00:00:5e:00:53:16.

```
*A:PE-2# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:5e:00:53:11 mpls-1:                 Evpn     12/09/22 10:06:17
                             192.0.2.1:524286
           rsvp:1
1          00:00:5e:00:53:16 mpls-2:                 Evpn     12/09/22 10:06:17
                             192.0.2.4:524282
           sr-te:655362
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
```

```
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

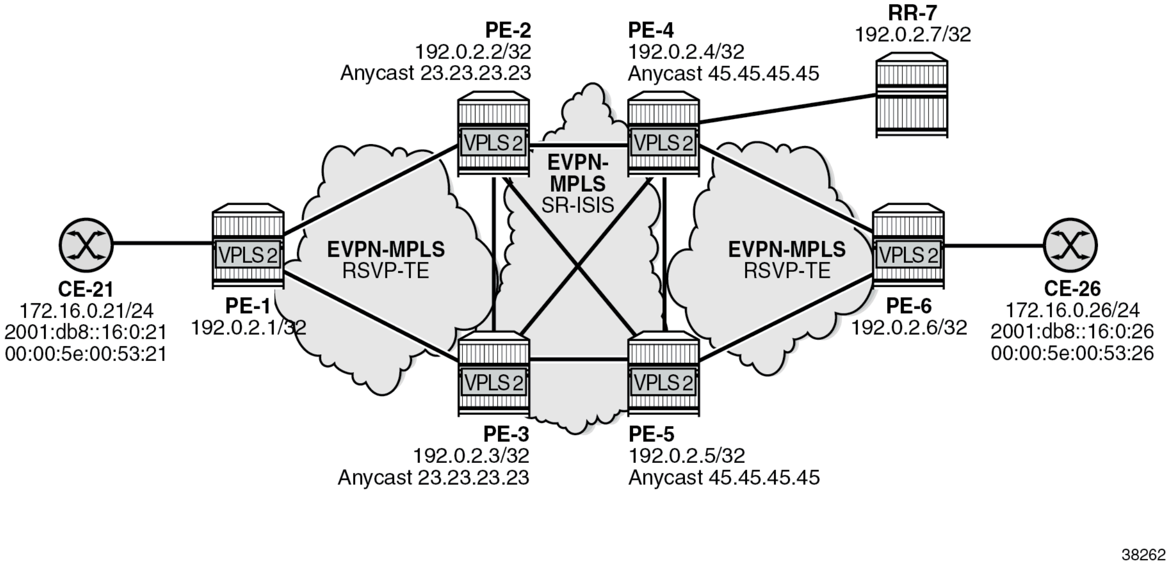
The following command shows the received EVPN-MAC routes on PE-2 for MAC address 00:00:5e:00:53:16. The route with RD 192.0.2.4:2000 is used:

```
*A:PE-2# show router bgp routes evpn mac mac-address 00:00:5e:00:53:16
===============================================================================
 BGP Router ID:192.0.2.2        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.4:2000      00:00:5e:00:53:16 ESI-0
      0                   Seq:0             LABEL 524282
                          n/a
                          192.0.2.4

*>i   192.0.2.6:1         00:00:5e:00:53:16 ESI-0
      0                   Seq:0             LABEL 524286
                          n/a
                          192.0.2.6

-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

Dual EVPN-MPLS instance VPLS with anycast multihoming

Figure 55: Example topology 2 shows example topology 2 with VPLS 2 configured on six PEs. PE-2 and PE-3 are redundant EVPN GWs with anycast address 23.23.23.23; PE-4 and PE-5 are redundant EVPN GWs with anycast address 45.45.45.45. RR-7 is the route reflector for PE-2, PE-3, PE-4, and PE-5 in the WAN network.

Figure 55: Example topology 2



The initial configuration includes:

- cards, MDAs, ports
- router interfaces
- IS-IS level 1 between PE-1, PE-2, and PE-3
- IS-IS level 1 between PE-4, PE-5, and PE-6
- IS-IS level 2 between PE-2, PE-3, PE-4, PE-5, and RR-7
- SR-ISIS between PE-2, PE-3, PE-4, and PE-5
- MPLS LSPs between PE-1 and PE-2, between PE-1 and PE-3, between PE-4 and PE-6, and between PE-5 and PE-6

The BGP configuration on PE-1 and PE-6 is similar.

```
# on PE-1:
configure
    router Base
        autonomous-system 64496
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "access1"
                family evpn
                peer-as 64496
                neighbor 192.0.2.2        # on PE-6: 192.0.2.4
                exit
                neighbor 192.0.2.3        # on PE-6: 192.0.2.5
                exit
            exit
```

The BGP configuration on PE-3 is:

```
# on PE-3:
configure
    router Base
        autonomous-system 64496
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "WAN"
                family evpn
                next-hop-self
                export "drop-tag-10"
                peer-as 64496
                neighbor 192.0.2.7
                exit
            exit
            group "access1"
                family evpn
                next-hop-self
                cluster 192.0.2.3
                export "drop-tag-20"
                peer-as 64496
                neighbor 192.0.2.1
                exit
            exit
```

The BGP configuration on PE-2, PE-4, and PE-5 is similar.

On PE-1, VPLS 2 is configured with a single EVPN-MPLS instance. PE-1 imports and exports routes with RT 64496:501. The configuration is as follows:

```
# on PE-1:
configure
    service
        vpls 2 name "VPLS 2" customer 1 create
            bgp
                # route-distinguisher 192.0.2.1:2  # will be auto-derived
                route-target export target:64496:501 import target:64496:501
            exit
            bgp-evpn
                evi 2
                mpls bgp 1
                    auto-bind-tunnel
                        resolution-filter
                            rsvp
                        exit
                        resolution filter
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/1/c10/1:2 create
                no shutdown
            exit
            no shutdown
        exit
```

On PE-2 and PE-3, the following policies are used in VPLS 2:

- Export policy "vsi-501-export" adds the communities "SOO-23" for the site of origin (SOO) and "RT64496:501" for the RT.
- Export policy "vsi-502-export" adds the communities "SOO-23" and "RT64496:502".
- Import policy "vsi-501-import" prevents loops based on the SOO and accepts routes with RT 64496:501.
- Import policy "vsi-502-import" prevent loops based on the SOO and accepts routes with RT 64496:502.

```
# on PE-2, PE-3:
configure
    router Base
        policy-options
            begin
            community "SOO-23"
                members "origin:23:23"
            exit
            community "RT64496:501"
                members "target:64496:501"
            exit
            community "RT64496:502"
                members "target:64496:502"
            exit
            policy-statement "vsi-501-export"
                default-action accept
                    community add "RT64496:501" "SOO-23"
                exit
            exit
            policy-statement "vsi-501-import"
                entry 10
                    from
                        community "SOO-23"
                        family evpn
                    exit
                    action drop
                    exit
                exit
                entry 20
                    from
                        community "RT64496:501"
                        family evpn
                    exit
                    action accept
                    exit
                exit
            exit
            policy-statement "vsi-502-export"
                default-action accept
                    community add "RT64496:502" "SOO-23"
                exit
            exit
            policy-statement "vsi-502-import"
                entry 10
                    from
                        community "SOO-23"
                        family evpn
                    exit
                    action drop
                    exit
                exit
                entry 20
                    from
                        community "RT64496:502"
                        family evpn
                    exit
                    action accept
                    exit
                exit
            exit
            commit
```

On PE-2 and PE-3, VPLS 2 is configured with two EVPN-MPLS instances: instance 1 is configured with multihoming mode access and instance 2 with multihoming mode network. For redundancy, anycast multihoming is configured with anycast address 23.23.23.23 and identical RDs and RTs for the same instance. The RD for BGP 1 is 192.0.2.23:2 and the RD for BGP 2 is 192.0.2.32:2. The **default-route-tag 10** command is configured for service instance 1, while **default-route-tag 20** is configured for service instance 2. These route tags are used in the BGP peer export policies to differentiate the different routes.On PE-2 and PE-3, VPLS 2 is configured as follows:

```
# on PE-2, PE-3:
configure
    service
        vpls 2 name "VPLS 2" customer 1 create
            description "dual BGP-EVPN MPLS instance VPLS"
            bgp
                route-distinguisher 192.0.2.23:2
                vsi-export "vsi-501-export"
                vsi-import "vsi-501-import"
            exit
            bgp 2
                route-distinguisher 192.0.2.32:2
                vsi-export "vsi-502-export"
                vsi-import "vsi-502-import"
            exit
            bgp-evpn
                incl-mcast-orig-ip 23.23.23.23
                evi 2
                mpls bgp 1
                    mh-mode access
                    auto-bind-tunnel
                        resolution-filter
                            rsvp
                        exit
                        resolution filter
                    exit
                    default-route-tag 10
                    no shutdown
                exit
                mpls bgp 2
                    # mh-mode network               # default MH mode
                    auto-bind-tunnel
                        resolution-filter
                            sr-isis
                        exit
                        resolution filter
                    exit
                    default-route-tag 20
                    no shutdown
                exit
            exit
            stp
```

```
                shutdown
            exit
            no shutdown
```

Note: For anycast multihoming, the RDs must be identical, so all RDs are configured manually.

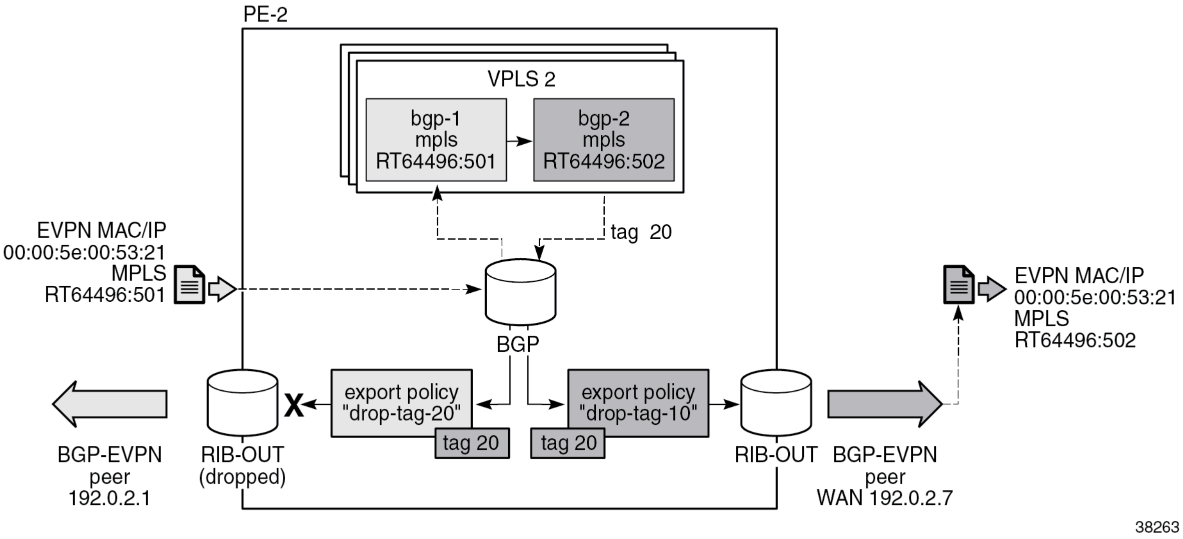
In datacenter GWs (DC GWs) with EVPN-VXLAN and EVPN-MPLS instances, route policies can match on the encapsulation type VXLAN or MPLS. In DC GWs with two EVPN-MPLS instances, the default route tag is used instead. The default route tag prevents a MAC/IP route that is installed in instance 1 (access) from being readvertised back to the access peers. In a similar way, MAC/IP routes installed in instance 2 are not readvertised back to peers in instance 2. On PE-2 and PE-3, the BGP peer export policy "drop-tag-10" drops routes with tag 10 and is configured in BGP group "WAN" with neighbor RR-7; BGP peer export policy "drop-tag-20" drops routes with tag 20 and is configured in BGP group "access1" with neighbor PE-1.

```
# on PE-2, PE-3:
configure
    router Base
        policy-options
            begin
            policy-statement "drop-tag-10"
                description "used as export policy toward WAN BGP peers"
                entry 10
                    from
                        tag 10
                    exit
                    action drop
                    exit
                exit
                default-action accept
                exit
            exit
            policy-statement "drop-tag-20"
                description "used as export policy toward DC BGP peers"
                entry 10
                    from
                        tag 20
                    exit
                    action drop
                    exit
                exit
                default-action accept
                exit
            exit
            commit
        exit
        bgp
            group "access1"
                export "drop-tag-20"
            exit
            group "WAN"
                export "drop-tag-10"
            exit
```

Figure 56: Export policies on PE-2 drop routes based on tag shows an incoming EVPN MAC route on PE-2 for CE-21's MAC address 00:00:5e:00:53:21. PE-2 receives the EVPN MAC route with RT target:64496:501 from PE-1 (BGP-EVPN peer 192.0.2.1). On PE-2, BGP 1 in VPLS 2 imports routes with this RT and the MAC address is installed in the FDB. The EVPN MAC route is redistributed to BGP 2

where the communities "RT64496:502" and "SOO-23", as well as internal tag 20, are added to the route. When PE-2’s BGP process sends an EVPN MAC route with tag 20 to BGP peer PE-1, the BGP export policy "drop-tag-20" drops the route, preventing PE-2 from re-advertising the EVPN MAC route back to the access peer 192.0.2.1. PE-2 can only send the EVPN MAC route to WAN neighbor 192.0.2.7 because the BGP export policy toward the WAN only drops the routes with tag 10, not the ones with tag 20.

Figure 56: Export policies on PE-2 drop routes based on tag



For completeness, the configuration on PE-4 and PE-5 is as follows:

```
# on PE-4, PE-5:
configure
    router Base
        policy-options
            begin
            community "SOO-45"
                members "origin:45:45"
            exit
            community "RT64496:502"
                members "target:64496:502"
            exit
            community "RT64496:503"
                members "target:64496:503"
            exit
            policy-statement "drop-tag-20"
                description "used as export policy toward DC BGP peers"
                entry 10
                    from
                        tag 20
                    exit
                    action drop
                    exit
                exit
                default-action accept
                exit
            exit
            policy-statement "drop-tag-30"
                description "used as export policy toward WAN BGP peers"
                entry 10
```

```
                    from
                        tag 30
                    exit
                    action drop
                    exit
                exit
                default-action accept
                exit
            exit
            policy-statement "vsi-502-export"
                default-action accept
                    community add "RT64496:502" "SOO-45"
                exit
            exit
            policy-statement "vsi-502-import"
                entry 10
                    from
                        community "SOO-45"
                        family evpn
                    exit
                    action drop
                    exit
                exit
                entry 20
                    from
                        community "RT64496:502"
                        family evpn
                    exit
                    action accept
                    exit
                exit
            exit
            policy-statement "vsi-503-export"
                default-action accept
                    community add "RT64496:503" "SOO-45"
                exit
            exit
            policy-statement "vsi-503-import"
                entry 10
                    from
                        community "SOO-45"
                        family evpn
                    exit
                    action drop
                    exit
                exit
                entry 20
                    from
                        community "RT64496:503"
                        family evpn
                    exit
                    action accept
                    exit
                exit
            exit
            commit
        exit
    exit
    service
        vpls 2 name "VPLS 2" customer 1 create
            description "dual BGP-EVPN MPLS instance VPLS"
            bgp
                route-distinguisher 192.0.2.45:2
                vsi-export "vsi-503-export"
                vsi-import "vsi-503-import"
            exit
            bgp 2
                route-distinguisher 192.0.2.54:2
                vsi-export "vsi-502-export"
                vsi-import "vsi-502-import"
            exit
            bgp-evpn
                incl-mcast-orig-ip 45.45.45.45
                evi 2
                mpls bgp 1
                    mh-mode access
                    auto-bind-tunnel
                        resolution-filter
                            rsvp
                        exit
                        resolution filter
                    exit
                    default-route-tag 30
                    no shutdown
                exit
                mpls bgp 2
                    # mh-mode network               # default MH mode
                    auto-bind-tunnel
                        resolution-filter
                            sr-isis
                        exit
                        resolution filter
                    exit
                    default-route-tag 20
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
```

The following command on PE-2 shows BGP instances 1 and 2 in VPLS 2. RD 192.0.2.23:2 is configured in BGP instance 1; RD 192.0.2.32:2 is configured in BGP instance 2. The RTs are defined by virtual switching instance (VSI) policies.

```
*A:PE-2# show service id 2 bgp

===============================================================================
BGP Information
===============================================================================
Bgp Instance         : 1
Vsi-Import           : vsi-501-import
Vsi-Export           : vsi-501-export
Route Dist           : 192.0.2.23:2
Oper Route Dist      : 192.0.2.23:2
Oper RD Type         : configured
Rte-Target Import    : None                 Rte-Target Export: None
Oper RT Imp Origin   : vsi                  Oper RT Import   : Policy Based
Oper RT Exp Origin   : vsi                  Oper RT Export   : Policy Based
ADV Service MTU      : -1

Bgp Instance         : 2
Vsi-Import           : vsi-502-import
Vsi-Export           : vsi-502-export
Route Dist           : 192.0.2.32:2
```

```
Oper Route Dist      : 192.0.2.32:2
Oper RD Type         : configured
Rte-Target Import    : None                 Rte-Target Export: None
Oper RT Imp Origin   : vsi                  Oper RT Import   : Policy Based
Oper RT Exp Origin   : vsi                  Oper RT Export   : Policy Based
ADV Service MTU      : -1

PW-Template Id       : None
-------------------------------------------------------------------------------
===============================================================================
```

The following command shows that EVPN destination 192.0.2.1 is reachable via an RSVP tunnel and EVPN destination 192.0.2.4 via an SR-ISIS tunnel. In EVPN-MPLS instance 2 of VPLS 2 on PE-2, the EVPN destination 192.0.2.4 is reachable via an SR-ISIS tunnel:

```
*A:PE-2# show service id 2 evpn-mpls

===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address                    Egr Label     Num.   Mcast Last Change
                               Transport:Tnl MACs         Sup BCast Domain
-------------------------------------------------------------------------------
192.0.2.1                      524284        1      bum   12/09/2022 10:11:04
                               rsvp:1                     No
192.0.2.4                      524278        1      bum   12/09/2022 10:11:17
                               isis:524291                No
-------------------------------------------------------------------------------
Number of entries : 2
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                       Num. Macs               Last Change
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

When traffic is sent between CE-21 and CE-26, the FDB in PE-1 shows that traffic toward MAC address 00:00:5e:00:53:26 is sent via RSVP tunnel 1 toward PE-2:

```
*A:PE-1# show service id 2 fdb detail

===============================================================================
Forwarding Database, Service 2
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
2          00:00:5e:00:53:21 sap:1/1/c10/1:2         L/120    12/09/22 10:10:20
2          00:00:5e:00:53:26 mpls-1:                 Evpn     12/09/22 10:11:36
                             192.0.2.2:524281
           rsvp:1
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The following command on PE-1 shows that only the EVPN MAC route received from PE-2 is used, not the one from PE-3 in the same anycast group. This is due to the best path selection done by BGP for the two routes, which have the same route key:

```
*A:PE-1# show router bgp routes evpn mac mac-address 00:00:5e:00:53:26
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.23:2        00:00:5e:00:53:26 ESI-0
      0                   Seq:0             LABEL 524281
                          n/a
                          192.0.2.2

*>i   192.0.2.23:2        00:00:5e:00:53:26 ESI-0
      0                   Seq:0             LABEL 524283
                          n/a
                          192.0.2.3

-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

The FDB for VPLS 2 on PE-2 shows that MAC address 00:00:5e:00:53:21 can be reached using EVPN-MPLS instance 1 whereas MAC address 00:00:5e:00:53:26 can be reached using EVPN-MPLS instance 2:

```
*A:PE-2# show service id 2 fdb detail

===============================================================================
Forwarding Database, Service 2
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
2          00:00:5e:00:53:21 mpls-1:                 Evpn     12/09/22 10:11:04
                             192.0.2.1:524284
           rsvp:1
2          00:00:5e:00:53:26 mpls-2:                 Evpn     12/09/22 10:11:36
                             192.0.2.4:524278
           isis:524291
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The FDB for VPLS 2 on PE-4 is as follows:

```
*A:PE-4# show service id 2 fdb detail
```

```

===============================================================================
Forwarding Database, Service 2
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
2          00:00:5e:00:53:21 mpls-2:                 Evpn     12/09/22 10:11:20
                             192.0.2.2:524280
           isis:524290
2          00:00:5e:00:53:26 mpls-1:                 Evpn     12/09/22 10:11:36
                             192.0.2.6:524284
           rsvp:1
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The FDB for VPLS 2 on PE-6 is as follows:

```
*A:PE-6# show service id 2 fdb detail

===============================================================================
Forwarding Database, Service 2
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
2          00:00:5e:00:53:21 mpls-1:                 Evpn     12/09/22 10:11:39
                             192.0.2.4:524279
           rsvp:1
2          00:00:5e:00:53:26 sap:1/1/c10/1:2         L/30     12/09/22 10:11:36
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

Conclusion

Dual-instance EVPN-MPLS GWs reduce the number of service labels to be learned at the access nodes, and optimizes the replication of BUM traffic from the access nodes.

EVPN E-LAN Services with SRv6 Transport

This chapter provides information about SRv6 support for distributed EVPN-enabled VPLS Layer 2 multipoint overlay services.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

The information and configuration in this chapter are based on SR OS Release 22.10.R1. SRv6 support for distributed EVPN-enabled VPLS Layer 2 multipoint overlay services is supported on FP-based platforms with FP4-based network ports in SR OS Release 22.7.R1 and later.

Overview

On FP-based platforms with FP4-based network ports, SR OS provides SRv6 support for distributed EVPN-enabled VPLS Layer 2 multipoint overlay services. SRv6 tunnels carry EVPN data between the PEs on which the EVPN service is provisioned. As usual in EVPN services, a full mesh of SRv6 tunnels is set up among all PEs that participate in the EVPN-enabled VPLS service. This supports the flooding of Broadcast, Unknown unicast, or Multicast (BUM) traffic to all remote destinations in the service, while ensuring that the PEs receive the traffic without looping or duplication of frames. Two or more routers may participate in a single EVPN-enabled VPLS service; a single router may participate in multiple EVPN-enabled VPLS services. The PE routers attached to an EVPN-enabled VPLS service with SRv6 transport use SRv6 End.DT2U behavior to terminate and forward unicast traffic, and SRv6 End.DT2M behavior to terminate and forward BUM traffic.

An SRv6 L2 Service TLV, which is carried in a BGP Prefix-SID attribute, signals the SRv6 Service SID for the End.DT2U or End.DT2M behavior for an EVPN-enabled VPLS Layer 2 overlay service, as per RFC 9252. The SRv6 Service SID is equivalent to an MPLS label for EVPN service routes in RFC 7432.

When a PE is attached to an EVPN-enabled VPLS service with SRv6 transport, the PE advertises its originating IP address in an Inclusive Multicast Ethernet Tag (IMET) route (also known as en EVPN type 3 route), along with the service attributes and the SRv6 SID corresponding to the End.DT2M behavior for the service. A remote PE attached to the same EVPN-enabled VPLS service imports the IMET route based on the import route target and adds an SRv6 destination entry to its flooding list for the EVPN-enabled VPLS service. In this way, all PEs that participate in an EVPN-enabled VPLS service learn about each other.

As in any other type of EVPN-enabled VPLS service, a PE learns the MAC address of a locally connected CE, either via data plane MAC learning or static provisioning. In the case of data plane MAC learning, a PE learns the source MAC address from data frames that it receives from the CE and adds a temporary entry for it in a VPLS forwarding database (FDB), which, on each PE, is private for each EVPN-enabled VPLS service.

A local MAC address is advertised in an EVPN MAC/IP advertisement route (EVPN type 2 route) for the EVPN-enabled VPLS service, along with the service parameters and an SRv6 SID corresponding to the End.DT2U behavior for the service. A remote PE that imports the EVPN MAC/IP advertisement route adds an entry for the advertised MAC addresses to the FDB, pointing at an SRv6 destination based on the received SRv6 SID. In this way, remote PEs that participate in an EVPN-enabled VPLS service with SRv6 transport learn how to unicast return traffic to the remote (source) MAC address.

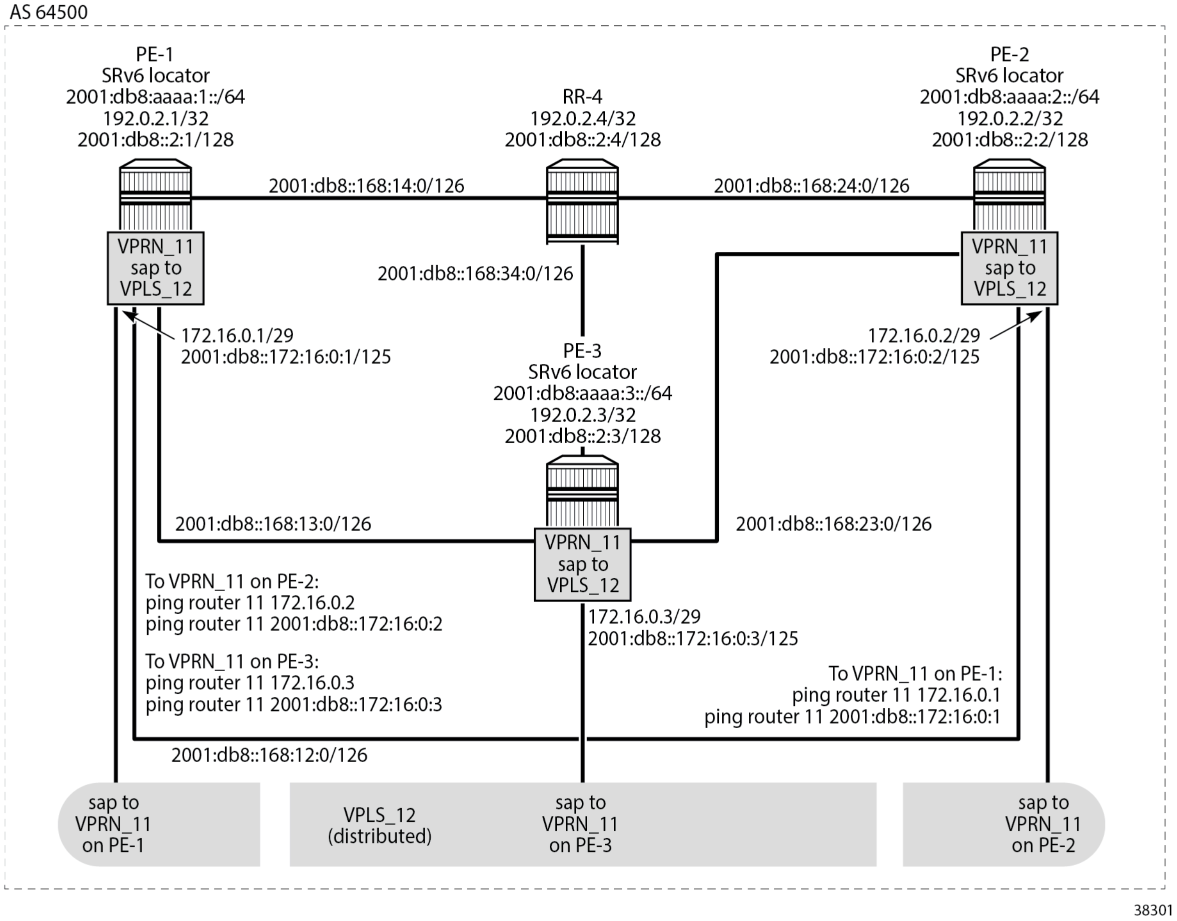
The **locator** command in the **service vpls <service-id> segment-routing-v6 <instance>** context configures the SRv6 locator that the PE uses to terminate SRv6 traffic for the EVPN-enabled VPLS service.

The base SRv6 configuration is as described in the "SRv6 Encapsulation in the Base Routing Instance" chapter in the *7750 SR and 7950 XRS Segment Routing and PCE Advanced Configuration Guide for Classic CLI*.

Configuration

Figure 57: Example topology shows the example topology with three PE routers. The SRv6-enabled network that it represents comprises PE-1, PE-2, and PE-3 in the control and data planes, and a BGP route reflector RR-4 in the control plane only. The SRv6-enabled network has only IPv6 addresses and interfaces. IS-IS and BGP are configured on all routers. The system interfaces have also an IPv4 address, from which a unique router-id is automatically derived for IS-IS and BGP respectively.

Figure 57: Example topology



For the traffic of data frames from the EVPN-enabled VPLS service on a local PE to the same EVPN-enabled VPLS service on a remote PE, the local PE acts as the SRv6 ingress PE node, while the remote PE acts as the SRv6 egress PE node. SRv6 and forwarding port extensions (FPE) are configured only on the PE routers.

The **ping** commands between IPv4 and IPv6 interface addresses in the EVPN-enabled VPLS service simulate IPv4 and IPv6 data traffic respectively.

Configure the router

This configuration includes:

- on PE-1, PE-2, PE-3, and RR-4:
  - ports, IPv6-only interfaces, and system interfaces
  - IS-IS:
    - level 2 capability with wide metrics (for the 128-bit identifiers)
    - native IPv6 routing

        - the **traffic-engineering** and **traffic-engineering-options** commands, as a best practice to advertise the router capability within the autonomous system (AS)
    - BGP, with internal group "gr_v6_internal" that includes:
        - the EVPN family
        - BGP neighbor system IPv6 addresses
- on PE-1, PE-2, and PE-3, port cross-connect (PXC), using internal loopbacks on an FP4 MAC chip, as described in the "Segment Routing over IPv6" chapter in the *7750 SR and 7950 XRS Segment Routing and PCE Advanced Configuration Guide for Classic CLI*

The following example configuration applies for PE-1. A similar configuration applies for PE-2, PE-3, and RR-4. RR-4 has PE-1, PE-2 and PE-3 as BGP neighbors in a cluster.

```
*A:PE-1# configure
    port 1/1/c2/1
        ethernet
            mode hybrid
            encap-type dot1q
        exit
        no shutdown
    exit
    port 1/1/c3/1
        ethernet
            mode hybrid
            encap-type dot1q
        exit
        no shutdown
    exit
    port 1/1/c4/1
        ethernet
            mode hybrid
            encap-type dot1q
        exit
        no shutdown
    exit
    port 1/1/c1/1
        ethernet
            mode hybrid
            encap-type dot1q
        exit
        no shutdown
    exit
    port 1/1/c1/2
        ethernet
            mode hybrid
            encap-type dot1q
        exit
        no shutdown
    exit
    router Base
        interface "int-PE-1-PE-2"
            description "interface between PE-1 and PE-2"
            port 1/1/c2/1:1000
            ipv6
                address 2001:db8::168:12:1/126
            exit
            no shutdown
        exit
        interface "int-PE-1-PE-3"
            description "interface between PE-1 and PE-3"
            port 1/1/c3/1:1000
            ipv6
                address 2001:db8::168:13:1/126
            exit
            no shutdown
        exit
        interface "int-PE-1-RR-4"
            description "interface between PE-1 and RR-4"
            port 1/1/c4/1:1000
            ipv6
                address 2001:db8::168:14:1/126
            exit
            no shutdown
        exit
        interface "system"
            address 192.0.2.1/32
            description "system interface of PE-1"
            ipv6
                address 2001:db8::2:1/128
            exit
            no shutdown
        exit
        autonomous-system 64500
        isis 0
            level-capability level-2
            area-id 49.0001
            traffic-engineering
            traffic-engineering-options
                ipv6
                application-link-attributes
                exit
            exit
            advertise-router-capability as
            ipv6-routing native
            level 2
                wide-metrics-only
            exit
            interface "system"
                passive
                no shutdown
            exit
            interface "int-PE-1-PE-2"
                interface-type point-to-point
                no shutdown
            exit
            interface "int-PE-1-PE-3"
                interface-type point-to-point
                no shutdown
            exit
            interface "int-PE-1-RR-4"
                interface-type point-to-point
                no shutdown
            exit
            no shutdown
        exit
        bgp
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "gr_v6_internal"
                description "internal bgp group on PE-1"
                family evpn
                peer-as 64500
                neighbor 2001:db8::2:4
                exit
            exit
            no shutdown
        exit
    exit
exit
```

Configure the VPRNs to simulate CEs

On each PE, the VPRN configuration includes an IPv4 address and an IPv6 address for an interface from the local VPRN to the EVPN-enabled VPLS service. These IPv4 and IPv6 addresses must be in the same address range on all PEs, because the same EVPN-enabled VPLS service is provisioned on each PE. Each interface to the (local) EVPN-enabled VPLS service also includes a SAP.

The VPRNs are introduced only to simulate CEs from where the **ping** commands can be launched.

The following example configuration applies for VPRN 11 on PE-1. A similar configuration applies for VPRN 11 on PE-2 and for VPRN 11 on PE-3.

```
*A:PE-1# configure
    service
        vprn 11 name "VPRN_11" customer 1 create
            description "CE_1"
            interface "local" create
                mac 00:00:5e:00:53:01
                address 172.16.0.1/29
                ipv6
                    address 2001:db8::172:16:0:1/125
                exit
                sap 1/1/c1/2:11 create
                exit
            exit
            no shutdown
        exit
    exit all
```

For example, VPRN 11 on PE-2 has the following interface, with corresponding IPv4 and IPv6 addresses. Similar output applies for VPRN 11 on PE-1 and for VPRN 11 on PE-3.

```
*A:PE-2# show router 11 interface

===============================================================================
Interface Table (Service: 11)
===============================================================================
Interface-Name                   Adm       Opr(v4/v6)  Mode    Port/SapId
   IP-Address                                                  PfxState
-------------------------------------------------------------------------------
local                            Up        Up/Up       VPRN    1/1/c2/2:11
   172.16.0.2/29                                               n/a
   2001:db8::172:16:0:2/125                                    PREFERRED
   fe80::200:22ff:fe22:2222/64                                 PREFERRED
-------------------------------------------------------------------------------
Interfaces : 1
===============================================================================
```

VPRN 11 on PE-1 has the following IPv4 and IPv6 routes. Similar output applies for VPRN 11 on PE-2 and for VPRN 11 on PE-3.

For IPv4:

```
*A:PE-1# show router 11 route-table

===============================================================================
Route Table (Service: 11)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
172.16.0.0/29                                 Local   Local     00h15m44s  0
       local                                                        0
-------------------------------------------------------------------------------
No. of Routes: 1
---snip---
===============================================================================
```

For IPv6:

```
*A:PE-1# show router 11 route-table ipv6

===============================================================================
IPv6 Route Table (Service: 11)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
2001:db8::172:16:0:0/125                      Local   Local     00h15m43s  0
       local                                                        0
-------------------------------------------------------------------------------
No. of Routes: 1
---snip---
===============================================================================
```

VPRN 11 on PE-1 has one locally learned MAC address for the locally configured interface. Similar output applies for VPRN 11 on PE-2 and for VPRN 11 on PE-3.

```
*A:PE-1# show router 11 arp summary

============================================================
ARP Table Summary (Service: 11)
============================================================
Local ARP Entries    : 1
---snip---
Dynamic ARP Entries  : 0
---snip---
------------------------------------------------------------
No. of ARP Entries   : 1
============================================================
```

The **show router 11 arp** command shows the association between the IP address and the MAC address, and the interface that the MAC address belongs to. The MAC address for the local interface to the EVPN-enabled VPLS service corresponds with that of the SAP that is configured for it in VPRN 11. Because the interface is statically configured, the association between the IP address and the MAC address does not expire. Similar output applies for PE-2 and for PE-3.

```
*A:PE-1# show router 11 arp

===============================================================================
ARP Table (Service: 11)
```

```
===============================================================================
IP Address      MAC Address       Expiry    Type   Interface
-------------------------------------------------------------------------------
172.16.0.1      00:00:5e:00:53:01 00h00m00s Oth[I] local
-------------------------------------------------------------------------------
No. of ARP Entries: 1
===============================================================================
```

Configure data path support, FPE, and SRv6

Configure data path support (PXC) and FPE identically on PE-1, PE2, and PE-3.

```
*A:PE-1# configure
    card 1
        mda 1
            xconnect
                mac 1 create
                    loopback 1 create
                    exit
                    loopback 2 create
                    exit
                exit
            exit
            no shutdown
        exit
    exit
    port-xc
        pxc 1 create
            port 1/1/m1/1
            no shutdown
        exit
        pxc 2 create
            port 1/1/m1/2
            no shutdown
        exit
    exit
    port pxc-1.a
        no shutdown
        exit
    port pxc-1.b
        no shutdown
    exit
    port pxc-2.a
        no shutdown
    exit
    port pxc-2.b
        no shutdown
    exit
    port 1/1/m1/1
        no shutdown
    exit
    port 1/1/m1/2
        no shutdown
    exit
    fwd-path-ext
        fpe 1 create
            path pxc 1
            srv6 origination
                interface-a
                exit
                interface-b
        exit
      exit
    exit
    fpe 2 create
      path pxc 2
      srv6 termination
        interface-a
        exit
        interface-b
        exit
      exit
    exit
  exit all
```

Configure the SRv6 locator *"PE-1_loc_VPLS"* with **ip-prefix** *2001:db8:aaaa:1::/64* in the **router Base segment-routing segment-routing-v6** context on PE-1 and similar on PE-2, with **ip-prefix** *2001:db8:aaaa:2::/64* for SRv6 locator *"PE-2_loc_VPLS"*, and on PE-3, with **ip-prefix** *2001:db8:aaaa:3::/64* for SRv6 locator *"PE-3_loc_VPLS"*.

```
*A:PE-1# configure router Base segment-routing segment-routing-v6
        source-address 2001:db8::2:1
        locator "PE-1_loc_VPLS"
          block-length 48
          prefix
            ip-prefix 2001:db8:aaaa:1::/64
          exit
          no shutdown
        exit all
```

Use FPE 1 as the SRv6 origination FPE and FPE 2 as the SRv6 termination FPE on PE-1, and similar on PE-2 for SRv6 locator *"PE-2_loc_VPLS"*, and on PE-3 for SRv6 locator *"PE-3_loc_VPLS"*. For more information, see the "Segment Routing over IPv6" chapter in the *7750 SR and 7950 XRS Segment Routing and PCE Advanced Configuration Guide for Classic CLI*.

```
*A:PE-1# config# router Base segment-routing
    segment-routing-v6
      origination-fpe 1
      locator "PE-1_loc_VPLS"
        termination-fpe 2
        no shutdown
      exit
    exit all
```

Advertise the SRv6 locator *"PE-1_loc_VPLS"* in IS-IS while ensuring level 2 capability on PE-1, and similar on PE-2 for SRv6 locator *"PE-2_loc_VPLS"*, and on PE-3 for SRv6 locator *"PE-3_loc_VPLS"*.

```
*A:PE-1# configure router Base
    isis 0
      segment-routing-v6
        locator "PE-1_loc_VPLS"
          level-capability level-2
          level 1
          exit
          level 2
          exit
        exit
        no shutdown
      exit
```

```
        exit all
```

Verify the IS-IS data base on PE-1 with the **show router isis 0 database detail** command. The output of this command (shortened here for PE-1, PE-3 and RR-4) provides information about each IS-IS-enabled router. For each uniquely identified IS-IS-enabled router, the SRv6 information indicates:

- the IS-IS-advertised router capabilities
- the IS-IS topology details
- the IPv4 and IPv6 reachability details
- the advertised SRv6 locator TLV
- the advertised configured SRv6 End SID and SRv6 End-X SIDs

```
*A:PE-1# show router isis 0 database detail

===============================================================================
Rtr Base ISIS Instance 0 Database (detail)
===============================================================================

Displaying Level 1 database
-------------------------------------------------------------------------------
Level (1) LSP Count : 0

Displaying Level 2 database
-------------------------------------------------------------------------------
LSP ID    : PE-1.00-00                                  Level     : L2
---snip---

-------------------------------------------------------------------------------
LSP ID    : PE-2.00-00                                  Level     : L2
Sequence  : 0x6                    Checksum  : 0xfb48   Lifetime  : 968
Version   : 1                      Pkt Type  : 20       Pkt Ver   : 1
Attributes: L1L2                   Max Area  : 3        Alloc Len : 432
SYS ID    : 1920.0000.2002         SysID Len : 6        Used Len  : 432

TLVs :
  Area Addresses:
    Area Address : (3) 49.0001
  Supp Protocols:
    Protocols     : IPv4
    Protocols     : IPv6
  IS-Hostname   : PE-2
  Router ID   :
    Router ID   : 192.0.2.2
  TE Router ID v6  :
    Router ID   : 2001:db8::2:2
  Router Cap : 192.0.2.2, D:0, S:0
    TE Node Cap : B E M  P
    SRv6 Cap: 0x0000
    SR Alg: metric based SPF
    Node MSD Cap: BMI : 0 SRH-MAX-SL : 10 SRH-MAX-END-POP : 9 SRH-MAX-H-ENCAPS : 3 SRH-MAX-END-
D : 9
  I/F Addresses :
    I/F Address   : 192.0.2.2
  I/F Addresses IPv6 :
    IPv6 Address    : 2001:db8::2:2
    IPv6 Address    : 2001:db8::168:12:2
    IPv6 Address    : 2001:db8::168:23:1
    IPv6 Address    : 2001:db8::168:24:1
  TE IS Nbrs   :
    Nbr   : PE-1.00
    Default Metric  : 10
    Sub TLV Len     : 36
    IPv6 Addr : 2001:db8::168:12:2
    Nbr IPv6  : 2001:db8::168:12:1
  TE IS Nbrs   :
    Nbr   : PE-3.00
    Default Metric  : 10
    Sub TLV Len     : 36
    IPv6 Addr : 2001:db8::168:23:1
    Nbr IPv6  : 2001:db8::168:23:2
  TE IS Nbrs   :
    Nbr   : RR-4.00
    Default Metric  : 10
    Sub TLV Len     : 18
    IPv6 Addr : 2001:db8::168:24:1
  TE IP Reach   :
    Default Metric  : 0
    Control Info:    , prefLen 32
    Prefix   : 192.0.2.2
  IPv6 Reach:
    Metric: ( I ) 0
    Prefix   : 2001:db8::2:2/128
    Metric: ( I ) 10
    Prefix   : 2001:db8::168:12:0/126
    Metric: ( I ) 10
    Prefix   : 2001:db8::168:23:0/126
    Metric: ( I ) 10
    Prefix   : 2001:db8::168:24:0/126
    Metric: ( I ) 0
    Prefix   : 2001:db8:aaaa:2::/64
  SRv6 Locator  :
    MT ID : 0
    Metric: ( ) 0 Algo:0
    Prefix   : 2001:db8:aaaa:2::/64

-------------------------------------------------------------------------------
LSP ID    : PE-3.00-00                                Level     : L2
---snip---

-------------------------------------------------------------------------------
LSP ID    : RR-4.00-00                                Level     : L2
---snip---

Level (2) LSP Count : 4
-------------------------------------------------------------------------------
---snip---
===============================================================================
```

PE-1 learns the remote SRv6 locators that PE-2 and PE-3 advertise and installs a route for them in the IPv6 routing table. This route uses an SRv6 tunnel. Similar output applies for PE-2 and for PE-3.

```
*A:PE-1# show router route-table ipv6

===============================================================================
IPv6 Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
---snip---
2001:db8::2:2/128                             Remote  ISIS      00h03m58s  18
       fe80::60e:1ff:fe01:1-"int-PE-1-PE-2"                         10
2001:db8::2:3/128                             Remote  ISIS      00h03m58s  18
```

```
       fe80::612:1ff:fe01:1-"int-PE-1-PE-3"                         10
---snip---
2001:db8:aaaa:1::/64                      Local   SRV6      00h05m05s  3
       fe80::201-"_tmnx_fpe_2.a"                                    0
2001:db8:aaaa:2::/64                      Remote  ISIS      00h03m50s  18
       2001:db8:aaaa:2::/64 (tunneled:SRV6-ISIS)                    10
2001:db8:aaaa:3::/64                      Remote  ISIS      00h03m43s  18
       2001:db8:aaaa:3::/64 (tunneled:SRV6-ISIS)                    10
-------------------------------------------------------------------------------
No. of Routes: 13
---snip---
===============================================================================
```

Next to its own local locator prefix, PE-1 also learns the remote locator prefixes that PE-2 and PE-3 advertise. Similar output applies for PE-2 and for PE-3.

```
*A:PE-1# show router isis 0 segment-routing-v6 locator

===============================================================================
Rtr Base ISIS Instance 0 SRv6 Locator Table
===============================================================================
Prefix                         AdvRtr                        MT     Lvl/Typ
 AttributeFlags                 Tag                           Flags  Algo
-------------------------------------------------------------------------------
2001:db8:aaaa:1::/64           PE-1                          0      2/Int.
  -                              0                             -      0
2001:db8:aaaa:2::/64           PE-2                          0      2/Int.
  -                              0                             -      0
2001:db8:aaaa:3::/64           PE-3                          0      2/Int.
  -                              0                             -      0
-------------------------------------------------------------------------------
No. of Locators: 3
-------------------------------------------------------------------------------
---snip---
===============================================================================
```

From PE-1, the remote locator prefix 2001:db8:aaaa:2::/64 is routable via a next hop using the "int-PE-1-PE-2" interface. Similar output applies for the remote locator prefix 2001:db8:aaaa:3::/64 using the "int-PE-1-PE-3" interface. Similar output applies from PE-2 and from PE-3.

```
*A:PE-1# show router isis 0 routes

===============================================================================
Rtr Base ISIS Instance 0 Route Table
===============================================================================
Prefix[Flags]                 Metric     Lvl/Typ     Ver.  SysID/Hostname
  NextHop                                            MT     AdminTag/SID[F]
-------------------------------------------------------------------------------
---snip---
2001:db8::2:2/128             10         2/Int.      12    PE-2
   fe80::60e:1ff:fe01:1-"int-PE-1-PE-2"                0       0
2001:db8::2:3/128             10         2/Int.      12    PE-3
---snip---
2001:db8:aaaa:1::/64          0          2/Int.      14    PE-1
   ::                                                  0       0
2001:db8:aaaa:2::/64          10         2/Int.      13    PE-2
   fe80::60e:1ff:fe01:1-"int-PE-1-PE-2"                0       0
2001:db8:aaaa:3::/64          10         2/Int.      14    PE-3
   fe80::612:1ff:fe01:1-"int-PE-1-PE-3"                0       0
-------------------------------------------------------------------------------
No. of Routes: 14 (14 paths)
-------------------------------------------------------------------------------
```

```
---snip---
===============================================================================
```

PE-1 transports IPv4 and IPv6 data to the remote SRv6 locator prefixes in an SRv6 encapsulated tunnel. For each SRv6 locator prefix destination, PE-1 sets up a different SRv6 tunnel with its specific label (TunnelId). Similar output applies for PE-2 and for PE-3.

```
*A:PE-1# show router tunnel-table ipv6

===============================================================================
IPv6 Tunnel Table (Router: Base)
===============================================================================
Destination                                     Owner     Encap TunnelId  Pref
Nexthop                                         Color           Metric
-------------------------------------------------------------------------------
2001:db8:aaaa:2::/64                            srv6-isis SRV6  524289    0
  fe80::60e:1ff:fe01:1-"int-PE-1-PE-2"                            10
2001:db8:aaaa:3::/64                            srv6-isis SRV6  524290    0
  fe80::612:1ff:fe01:1-"int-PE-1-PE-3"                            10
-------------------------------------------------------------------------------
---snip---
===============================================================================
```

The **show router fp-tunnel-table 1 ipv6** command in PE-1 shows the local endpoints of the SRv6 tunnels in PE-1. Similar output applies for the local endpoints of the SRv6 tunnels in PE-2 and for the local endpoints of the SRv6 tunnels in PE-3.

```
*A:PE-1# show router fp-tunnel-table 1 ipv6

===============================================================================
IPv6 Tunnel Table Display
---snip---
===============================================================================
Destination                                  Protocol         Tunnel-ID
  Lbl/SID
    NextHop                                                   Intf/Tunnel
  Lbl/SID (backup)
    NextHop   (backup)
-------------------------------------------------------------------------------
2001:db8:aaaa:2::/64                         SRV6              524289
  -
    fe80::60e:1ff:fe01:1-"int-PE-1-PE-2"                     1/1/c2/1:1000
2001:db8:aaaa:3::/64                         SRV6              524290
  -
    fe80::612:1ff:fe01:1-"int-PE-1-PE-3"                     1/1/c3/1:1000
-------------------------------------------------------------------------------
Total Entries : 2
-------------------------------------------------------------------------------
===============================================================================
```

Verify data traffic

At this point, verify that IPv4 and IPv6 data traffic is not possible between the local VPRN 11 on PE-1 and the remote VPRN 11 on PE-2 and PE-3. PE-1 is not aware of the remote MAC addresses that are associated with IPv4 address 172.16.0.2 and IPv4 address 172.16.0.3 (or IPv6 address 2001:db8::172:16:0:2 and IPv6 address 2001:db8::172:16:0:3), because only interfaces that are locally connected to the EVPN-enabled VPLS service on PE-1 reply on the ARP request. Perform a similar verification for IPv4 and IPv6 data traffic between the local VPRN 11 on PE-2 and the remote VPRN 11 on PE-1 and PE-3, and for IPv4 and IPv6 data traffic between the local VPRN 11 on PE-3 and the remote VPRN 11 on PE-1 and PE-2.

For example, for IPv4 data traffic to the remote VPRN 11 on PE-2:

```
*A:PE-1# ping router 11 172.16.0.2
PING 172.16.0.2 56 data bytes
Request timed out. icmp_seq=1.
---snip---
---- 172.16.0.2 PING Statistics ----
5 packets transmitted, 0 packets received, 100% packet loss
```

For example, for IPv6 data traffic to the remote VPRN 11 on PE-2:

```
*A:PE-1# ping router 11 2001:db8::172:16:0:2
PING 2001:db8::172:16:0:2 56 data bytes

---- 2001:db8::172:16:0:2 PING Statistics ----
5 packets transmitted, 5 packets bounced, 0 packets received, 100% packet loss
```

Configure the EVPN- and SRv6-enabled VPLS service on PE-1, PE-2, and PE-3

On each PE, this configuration includes a SAP to the local VPRN.

On PE-1, create an SRv6 instance *1* for the EVPN-enabled VPLS service. Use the SRv6 locator *"PE-1_loc_VPLS"* from the **router Base segment-routing segment-routing-v6** context in the **service vpls 12 segment-routing-v6 1** context and configure End.DT2U and End.DT2M behavior for it.

Use the configured SRv6 locator *"PE-1_loc_VPLS"* as the default locator in the **service vpls 12 bgp-evpn segment-routing-v6 1** context. In the **service vpls 12 bgp-evpn segment-routing-v6 1 locator "PE-1_loc_VPLS"** context, use the unique PE-1 system IPv6 address as the route next hop. This configuration can be verified with the **show service id 12 bgp** command (not shown). Perform a similar configuration on PE-2 (and PE-3), with the configured SRv6 locator *"PE-2_loc_VPLS"* (*"PE-3_loc_VPLS"*) as the default locator, and the PE-2 (PE-3) system IPv6 address as route next hop.

```
*A:PE-1# configure service
        vpls 12 name "VPLS_12" customer 1 create
            description "VPLS_12 on PE-1"
            segment-routing-v6 1 create
                locator "PE-1_loc_VPLS"
                    function
                        end-dt2u
                        end-dt2m
                    exit
                exit
            exit
            bgp
            exit
            bgp-evpn
                evi 1
                segment-routing-v6 bgp 1 srv6-instance 1 default-locator "PE-1_loc_VPLS" create
                    route-next-hop system-ipv6
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/1/c1/1:11 create
                description "sap to VPRN_11 on PE-1"
                no shutdown
            exit
            no shutdown
        exit all
```

The **show service id 12 fdb expiry** command shows that MAC learning and MAC aging are enabled. For example, the VPLS FDB entries that are locally learned expire after 300 seconds.

```
*A:PE-1# show service id 12 fdb expiry

===============================================================================
Forwarding Database, Service 12
===============================================================================
---snip---
Table Size        : 250                 Allocated Count   : 0
Total In Use      : 0
Learned Count     : 0                   Static Count      : 0
---snip---
BGP EVPN Count    : 0                   EVPN Static Cnt   : 0
---snip---
Remote Age        : 900                 Local Age         : 300
---snip---
Mac Learning      : Enabled             Discard Unknown   : Disabled
Mac Aging         : Enabled             Relearn Only      : False
---snip---
===============================================================================
```

The **show service id 12 bgp-evpn** command shows how BGP EVPN behavior is configured. MAC advertisement for EVPN MAC/IP advertisement routes (for **ping** commands) and inclusive multicast advertisement for EVPN IMET routes (for flooding and BUM traffic) are enabled. The next hop corresponds with the local system IPv6 address. The route resolution uses the route table of the VPRN that has a local interface to the EVPN-enabled VPLS service. Similar output applies for PE-2 and for PE-3.

```
*A:PE-1# show service id 12 bgp-evpn

===============================================================================
BGP EVPN Table
===============================================================================
EVI                : 1
Creation Origin    : manual

MAC/IP Routes
MAC Advertisement  : Enabled            Unknown MAC Route  : Disabled
CFM MAC Advertise  : Disabled

Multicast Routes
Sel Mcast Advert   : Disabled
Ing Rep Inc McastAd: Enabled

---snip---
===============================================================================
Segment Routing v6 Instance 1 Service 12
===============================================================================
Admin State             : Enabled
Srv6 Instance           : 1
Default Locator         : PE-1_loc_VPLS

Oper Group              : (Not Specified)
```

```
Default Route Tag       : 0x0
Source Address          : (Not Specified)
ECMP                    : 1
Force Vlan VC Fwd       : disabled
Next Hop Type           : system-ipv6
Evi 3-byte Auto-RT      : disabled
Route Resolution        : route-table
Force QinQ VC Fwd       : none
MH Mode                 : network
Rest Prot Src Mac       : disabled
Split Horizon Group     : n/a
===============================================================================
```

The configuration of the SRv6 End.DT2U and End.DT2M behavior for the SRv6 locator that is used in the EVPN-enabled VPLS service results in corresponding SRv6 full SIDs. For example, the **show service id 12 segment-routing-v6 instance 1** command on PE-2 shows them. For the SRv6 End.DT2U behavior, the SRv6 function is 524288 (0x80000) and the corresponding SRv6 full SID is 2001:db8::aaaa:2:8000::. For the SRv6 End.DT2M behavior, the SRv6 function is 524287 (0x7ffff) and the corresponding SRv6 full SID is 2001:db8::aaaa:2:7fff:f000::. Similar output applies for PE-1 and for PE-3.

```
*A:PE-2# show service id 12 segment-routing-v6 instance 1

===============================================================================
Segment Routing v6 Instance 1 Service 12
===============================================================================
Locator
  Type         Function  SID                                      Status
-------------------------------------------------------------------------------
PE-2_loc_VPLS
  End.DT2U       *524288 2001:db8:aaaa:2:8000::                   ok
  End.DT2M       *524287 2001:db8:aaaa:2:7fff:f000::              ok
===============================================================================
Legend: * - System allocated
```

The **show router segment-routing-v6 local-sid** command shows that the SRv6 local SIDs belong to the VPLS context. Similar output applies for PE-1 and for PE-3.

```
*A:PE-2# show router segment-routing-v6 local-sid

===============================================================================
Segment Routing v6 Local SIDs
===============================================================================
SID                                               Type           Function
  Locator
  Context
-------------------------------------------------------------------------------
2001:db8:aaaa:2:7fff:f000::                       End.DT2M       524287
  PE-2_loc_VPLS
  SvcId: 12 Name: VPLS_12
2001:db8:aaaa:2:8000::                            End.DT2U       524288
  PE-2_loc_VPLS
  SvcId: 12 Name: VPLS_12
-------------------------------------------------------------------------------
SIDs : 2
-------------------------------------------------------------------------------
===============================================================================
```

Enabling the SRv6 End.DT2M behavior allows the exchange of EVPN IMET BGP update messages for the EVPN family. The **show log log-id <log-id>** command on PE-1 shows the BGP update message that PE-1 receives from PE-2, via the RR. It indicates the remote source address (orig_addr:

2001:db8::2:2), and the route distinguisher (RD: 192.0.2.2:1), tag (tag: 0), route target (Extended Community: target:64500:1), and next hop (Global NextHop 2001:db8::2:2) that PE-1 must use while sending IPv4 or IPv6 data traffic to PE-2. In addition, it indicates the Provicer Multicast Service Interface (PMSI) information about tunnel type (Tunnel-type Ingress Replication), MPLS label (MPLS Label 8388592 (0x7fffff)), and tunnel endpoint (Tunnel-Endpoint 2001:db8::2:2). Finally, it indicates that PE-1 must send the frames to the SRv6 locator (SRv6 SID: 2001:db8:aaaa:2::) with End.DT2M behavior (Behavior: 0x18 (24)). Similar output applies for the BGP update that PE-1 receives from PE-3, via the RR. PE-1 advertises a similar BGP update message to the RR, which forwards it to PE-2 and PE-3 (not shown here). PE-2 and PE-3 receive and advertise similar BGP update messages.

```
*A:PE-1# show log log-id 2

===============================================================================
Event Log 2 log-name log_2
===============================================================================
Description : (Not Specified)
Memory Log contents  [size=100   next event=4  (not wrapped)]
---snip---
2 2022/12/20 12:50:55.367 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:4
"Peer 1: 2001:db8::2:4: UPDATE
Peer 1: 2001:db8::2:4 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 159
    Flag: 0x90 Type: 14 Len: 52 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 16 Global NextHop 2001:db8::2:2
        Type: EVPN-INCL-MCAST Len: 29 RD: 192.0.2.2:1, tag: 0, orig_addr len: 128, orig_addr:
 2001:db8::2:2
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.2
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        4.4.4.4
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64500:1
    Flag: 0xc0 Type: 22 Len: 21 PMSI:
        Tunnel-type Ingress Replication (6)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf: not required]
        MPLS Label 8388592
        Tunnel-Endpoint 2001:db8::2:2
    Flag: 0xc0 Type: 40 Len: 37 Prefix-SID-attr:
       SRv6 Services TLV (37 bytes):-
           Type: SRV6 L2 Service TLV (6)
           Length: 34 bytes, Reserved: 0x0
         SRv6 Service Information Sub-TLV (33 bytes)
             Type: 1 Len: 30 Rsvd1: 0x0
             SRv6 SID: 2001:db8:aaaa:2::
             SID Flags: 0x0 Endpoint Behavior: 0x18 Rsvd2: 0x0
             SRv6 SID Sub-Sub-TLV
                Type: 1 Len: 6
                BL:48 NL:16 FL:20 AL:0 TL:20 TO:64
"
---snip---
```

The reception of the EVPN IMET BGP update messages triggers PE-1 to install learned inclusive multicast routes as shown with the **show router bgp neighbor <ip-address> received-routes evpn** command. Because PE-1 receives EVPN IMET BGP update messages from PE-2 and from PE-3 with different route distinguishers, PE-1 installs a learned inclusive multicast route for each one of them. Similar output applies for PE-2 and for PE-3. The BGP EVPN inclusive multicast routes that are received, can also be displayed with the **show router bgp routes evpn incl-mcast** command.

```
*A:PE-1# show router bgp neighbor 2001:db8::2:4 received-routes evpn
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
---snip---
===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Flag  Route Dist.         OrigAddr
      Tag                 NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.2:1         2001:db8::2:2
      0                   2001:db8::2:2

u*>i  192.0.2.3:1         2001:db8::2:3
      0                   2001:db8::2:3

-------------------------------------------------------------------------------
Routes : 2
===============================================================================
---snip---
========================================================
```

The **show router bgp neighbor <ip-address> advertised-routes evpn** command on PE-2 shows the local inclusive multicast routes on PE-2. PE-2 advertises them to its BGP neighbors. Similar output applies for PE-1 and for PE-3.

```
*A:PE-2# show router bgp neighbor 2001:db8::2:4 advertised-routes evpn
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
---snip---
===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Flag  Route Dist.         OrigAddr
      Tag                 NextHop
-------------------------------------------------------------------------------
i     192.0.2.2:1         2001:db8::2:2
      0                   2001:db8::2:2

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
---snip---
================
```

From the received EVPN IMET BGP update messages, PE-1 learns the SRv6 tunnel endpoints for multicast traffic, as shown with the **show service id 12 segment-routing-v6 instance 1 destinations** command. The segment ID (SRv6 SID) corresponds with the expected End.DT2M behavior on PE-2 and PE-3 respectively. Similar output applies for PE-2 and for PE-3.

```
*A:PE-1# show service id 12 segment-routing-v6 instance 1 destinations

===============================================================================
TEP, SID
===============================================================================
Instance  TEP Address                Segment Id                 SupBcasDom Num
 Mcast                                                                     MACs
-------------------------------------------------------------------------------
1         2001:db8::2:2              2001:db8:aaaa:2:7fff:f000:  No        0
                                     :
 BUM
1         2001:db8::2:3              2001:db8:aaaa:3:7fff:f000:  No        0
                                     :
 BUM
-------------------------------------------------------------------------------
Number of TEP, SID: 2
-------------------------------------------------------------------------------
===============================================================================
---snip---
===============================================================================
```

The list of next hops for the EVPN family can be shown with the **show router bgp next-hop evpn** command. For each next hop, the details can be shown. The **show router bgp next-hop evpn 2001:db8::2:2 detail** command on PE-1 shows the details on PE-1 for next hop 2001:db8::2:2. It indicates that IPv4 and IPv6 data for the EVPN family uses the SRv6 tunnel for locator 2001:db8:aaaa:2::/64 and is sent to the next hop 2001:db8::2:2 via the resolved next hop fe80::60e:1ff:fe01:1, which corresponds with the "int-PE-1-PE-2" interface on PE-1. Similar output applies on PE-1 for next hop 2001:db8::2:3. Similar output applies for PE-2 and for PE-3.

```
*A:PE-1# show router bgp next-hop evpn 2001:db8::2:2 detail
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================

===============================================================================
BGP VPN Next Hop
===============================================================================
-------------------------------------------------------------------------------
VPN Next Hop          : 2001:db8::2:2
Autobind              : gre/rtm
Labels                : --
User-labels           : 1
Admin-tag-policy      : --
Strict-tunnel-tagging : N
Color                 : --
Locator               : 2001:db8:aaaa:2::/64
Created               : 00h01m18s
Last-modified         : 00h01m18s
-------------------------------------------------------------------------------
Resolving Prefix : 2001:db8::2:2/128
Preference       : 18                   Metric           : 10
Reference Count  : 1                    Owner            : GRE
Fib Programmed   : Y
Resolved Next Hop: fe80::60e:1ff:fe01:1
Egress Label     : n/a                  TunnelId         : 4294967293
Locator State    : Resolved
```

```
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Next Hops : 1
===============================================================================
```

The **show router bgp routes evpn incl-mcast hunt** command shows a consolidated view on the inclusive multicast routes for the EVPN family. On PE-1, in the RIB In Entries section, it shows for each learned next hop how PE-1 must handle the BUM traffic destined for it. In the RIB Out Entries section, it shows for each local next hop how PE-1 expects the remote routers to handle BUM traffic destined for it. Similar output applies for PE-2 and for PE-3.

```
*A:PE-1# show router bgp routes evpn incl-mcast hunt
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
---snip---
===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : n/a
Nexthop        : 2001:db8::2:2
Path Id        : None
From           : 2001:db8::2:4
Res. Nexthop   : fe80::60e:1ff:fe01:1
Local Pref.    : 100                    Interface Name : int-PE-1-PE-2
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64500:1
Cluster        : 4.4.4.4
Originator Id  : 192.0.2.2              Peer Router Id : 192.0.2.4
Flags          : Used Valid Best IGP
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : INCL-MCAST
Tag            : 0
Originator IP  : 2001:db8::2:2
Route Dist.    : 192.0.2.2:1
Route Tag      : 0
Neighbor-AS    : n/a
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h01m17s
SRv6 TLV Type  : SRv6 L2 Service TLV (6)
SRv6 SubTLV    : SRv6 SID Information (1)
Sid            : 2001:db8:aaaa:2::
Full Sid       : 2001:db8:aaaa:2:7fff:f000::
Behavior       : End.DT2M (24)
SRv6 SubSubTLV : SRv6 SID Structure (1)
Loc-Block-Len  : 48                     Loc-Node-Len   : 16
Func-Len       : 20                     Arg-Len        : 0
Tpose-Len      : 20                     Tpose-offset   : 64
-------------------------------------------------------------------------------
PMSI Tunnel Attributes :
Tunnel-type    : Ingress Replication
Flags          : Type: RNVE(0) BM: 0 U: 0 Leaf: not required
MPLS Label     : 8388592
```

```
Tunnel-Endpoint: 2001:db8::2:2
-------------------------------------------------------------------------------

Network        : n/a
Nexthop        : 2001:db8::2:3
---snip---

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
Network        : n/a
Nexthop        : 2001:db8::2:1
Path Id        : None
To             : 2001:db8::2:4
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : n/a
Connector      : None
Community      : target:64500:1
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.4
Origin         : IGP
AS-Path        : No As-Path
EVPN type      : INCL-MCAST
Tag            : 0
Originator IP  : 2001:db8::2:1
Route Dist.    : 192.0.2.1:1
Route Tag      : 0
Neighbor-AS    : n/a
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
SRv6 TLV Type  : SRv6 L2 Service TLV (6)
SRv6 SubTLV    : SRv6 SID Information (1)
Sid            : 2001:db8:aaaa:1::
Full Sid       : 2001:db8:aaaa:1:7fff:f000::
Behavior       : End.DT2M (24)
SRv6 SubSubTLV : SRv6 SID Structure (1)
Loc-Block-Len  : 48                     Loc-Node-Len   : 16
Func-Len       : 20                     Arg-Len        : 0
Tpose-Len      : 20                     Tpose-offset   : 64
-------------------------------------------------------------------------------
PMSI Tunnel Attributes :
Tunnel-type    : Ingress Replication
Flags          : Type: RNVE(0) BM: 0 U: 0 Leaf: not required
MPLS Label     : 8388592
Tunnel-Endpoint: 2001:db8::2:1
-------------------------------------------------------------------------------

-------------------------------------------------------------------------------
Routes : 3
===============================================================================
```

Verify data traffic

At this point, verify that IPv4 and IPv6 data traffic is possible between the local VPRN 11 on PE-1 and the remote VPRN 11 on PE-2 and PE-3. Perform a similar verification for IPv4 and IPv6 data traffic between the local VPRN 11 on PE-2 and the remote VPRN 11 on PE-1 and PE-3, and for IPv4 and IPv6 data traffic between the local VPRN 11 on PE-3 and the remote VPRN 11 on PE-1 and PE-2.

For example, for IPv4 data traffic to the remote VPRN 11 on PE-2:

```
*A:PE-1# ping router 11 172.16.0.2
PING 172.16.0.2 56 data bytes
64 bytes from 172.16.0.2: icmp_seq=1 ttl=64 time=6.89ms.
---snip---
---- 172.16.0.2 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 1.94ms, avg = 3.04ms, max = 6.89ms, stddev = 1.93ms
```

For example, for IPv6 data traffic to the remote VPRN 11 on PE-2:

```
*A:PE-1# ping router 11 2001:db8::172:16:0:2
PING 2001:db8::172:16:0:2 56 data bytes
64 bytes from 2001:db8::172:16:0:2 icmp_seq=1 hlim=64 time=14.4ms.
---snip---
---- 2001:db8::172:16:0:2 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 2.09ms, avg = 4.63ms, max = 14.4ms, stddev = 4.88ms
```

When the SRv6 End.DT2U behavior is enabled, the sending of IPv4 or IPv6 data traffic triggers the exchange of EVPN MAC/IP BGP update messages for the EVPN family. The **show log log-id <log-id>** command on PE-1 shows the BGP update message that PE-1 receives from PE-2, via the RR. It indicates the remote MAC address (mac: 00:00:5e:00:53:02), and the route distinguisher (RD: 192.0.2.2:1), ESI (ESI: ESI-0), tag (tag: 0), label (label1: 8388608 (0x800000)), route target (Extended Community: target:64500:1), and next hop (Global NextHop 2001:db8::2:2) that PE-1 must use while sending IPv4 or IPv6 data traffic to PE-2. In addition, it indicates that PE-1 must send the frames to the SRv6 locator (SRv6 SID: 2001:db8:aaaa:2::) with End.DT2U behavior (Behavior: 0x17 (23)). PE-1 derives the SRv6 full SID that is needed for this (2001:db8:aaaa:2:8000::). Similar output applies for the BGP update that PE-1 receives from PE-3, via the RR. PE-1 advertises a similar BGP update message to the RR, which forwards it to PE-2 and PE-3 (not shown here). PE-2 and PE-3 receive and advertise similar BGP update messages.

```
*A:PE-1# show log log-id 2

===============================================================================
Event Log 2 log-name log_2
===============================================================================
Description : (Not Specified)
Memory Log contents  [size=100   next event=4  (not wrapped)]
---snip---
2 2022/12/20 12:53:36.016 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:4
"Peer 1: 2001:db8::2:4: UPDATE
Peer 1: 2001:db8::2:4 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 139
    Flag: 0x90 Type: 14 Len: 56 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 16 Global NextHop 2001:db8::2:2
        Type: EVPN-MAC Len: 33 RD: 192.0.2.2:1 ESI: ESI-0, tag: 0, mac len: 48 mac:
 00:00:5e:00:53:02, IP len: 0, IP: NULL, label1: 8388608
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.2
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        4.4.4.4
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64500:1
    Flag: 0xc0 Type: 40 Len: 37 Prefix-SID-attr:
       SRv6 Services TLV (37 bytes):-
```

```
           Type: SRV6 L2 Service TLV (6)
           Length: 34 bytes, Reserved: 0x0
         SRv6 Service Information Sub-TLV (33 bytes)
             Type: 1 Len: 30 Rsvd1: 0x0
             SRv6 SID: 2001:db8:aaaa:2::
             SID Flags: 0x0 Endpoint Behavior: 0x17 Rsvd2: 0x0
             SRv6 SID Sub-Sub-TLV
                Type: 1 Len: 6
                BL:48 NL:16 FL:20 AL:0 TL:20 TO:64
"
---snip---
```

The reception of the EVPN MAC/IP BGP update messages triggers PE-1 to install learned MAC routes, as shown with the **show router bgp neighbor <ip-address> received-routes evpn** command. In contrast to the learned inclusive multicast routes, the learned MAC routes expire in accordance with the configuration that is shown with the **show service id 12 fdb expiry** command. PE-1 installs a learned MAC/IP route for each of the remote CEs. PE-1 derives the SRv6 function (524288) from the received label field. The earlier installed inclusive multicast routes remain in place (not shown). The BGP EVPN MAC/IP advertisement routes that are received, can also be displayed with the **show router bgp routes evpn mac** command. Similar output applies for PE-2 and for PE-3.

```
*A:PE-1# show router bgp neighbor 2001:db8::2:4 received-routes evpn
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
---snip---
===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.2:1         00:00:5e:00:53:02 ESI-0
      0                   Seq:0             LABEL 524288
                          n/a
                          2001:db8::2:2

u*>i  192.0.2.3:1         00:00:5e:00:53:03 ESI-0
      0                   Seq:0             LABEL 524288
                          n/a
                          2001:db8::2:3

-------------------------------------------------------------------------------
Routes : 2
===============================================================================

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
---snip---
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
---snip---
```

```
========================================================
```

The **show router bgp neighbor <ip-address> advertised-routes evpn** command on PE-2 shows the local MAC routes on PE-2. PE-2 advertises them to its BGP neighbors. Similar output applies for PE-1 and for PE-3.

```
*A:PE-2# show router bgp neighbor 2001:db8::2:4 advertised-routes evpn
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
---snip---
===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
i     192.0.2.2:1         00:00:5e:00:53:02 ESI-0
      0                   Seq:0             524288
                          n/a
                          2001:db8::2:2

-------------------------------------------------------------------------------
Routes : 1
===============================================================================

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Flag  Route Dist.         OrigAddr
      Tag                 NextHop
-------------------------------------------------------------------------------
i     192.0.2.2:1         2001:db8::2:2
      0                   2001:db8::2:2

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
---snip---
========================================================
```

From the received EVPN MAC/IP BGP update messages, PE-1 learns the SRv6 tunnel endpoints for unicast traffic, as shown with the **show service id 12 segment-routing-v6 instance 1 destinations** command. The segment ID (SRv6 SID) corresponds with the expected End.DT2U behavior on PE-2 and PE-3 respectively. The earlier learned SRv6 tunnel endpoints for BUM traffic remain in place. Similar output applies for PE-2 and for PE-3.

```
*A:PE-1# show service id 12 segment-routing-v6 instance 1 destinations

===============================================================================
TEP, SID
===============================================================================
Instance  TEP Address                Segment Id                SupBcasDom Num
```

```
 Mcast                                                                     MACs
--------------------------------------------------------------------------------
1         2001:db8::2:2              2001:db8:aaaa:2:7fff:f000:  No        0
                                     :
 BUM
1         2001:db8::2:2              2001:db8:aaaa:2:8000::      No        1
 -
1         2001:db8::2:3              2001:db8:aaaa:3:7fff:f000:  No        0
                                     :
 BUM
1         2001:db8::2:3              2001:db8:aaaa:3:8000::      No        1
 -
--------------------------------------------------------------------------------
Number of TEP, SID: 4
--------------------------------------------------------------------------------
================================================================================
---snip---
================================================================================
```

The **show service id 12 fdb expiry** command on PE-1 shows that PE-1 learns one MAC address locally, while PE-1 learns two remote MAC addresses via BGP EVPN.

```
*A:PE-1# show service id 12 fdb expiry

================================================================================
Forwarding Database, Service 12
================================================================================
---snip---
Table Size        : 250                 Allocated Count   : 3
Total In Use      : 3
Learned Count     : 1                   Static Count      : 0
---snip---
BGP EVPN Count    : 2                   EVPN Static Cnt   : 0
---snip---
Remote Age        : 900                 Local Age         : 300
---snip---
Mac Learning      : Enabled             Discard Unknown   : Disabled
Mac Aging         : Enabled             Relearn Only      : False
---snip---
================================================================================
```

The locally learned MAC address belongs to the originator of the **ping** commands in the VPRN 11 context on PE-1, while the BGP EVPN learned MAC addresses belong to the destinations for those **ping** commands, which are in the VPRN 11 context on PE-2 and in the VPRN 11 context on PE-3 respectively. The Transport:Tnl-Id (for example 2001:db8:aaaa:2:8000::) indicates that PE-1 transports frames to the destination (on or connected to PE-2) via the SRv6 full SID to PE-2 for the End.DT2U behavior. The VPLS FDB entries that PE-1 learns locally expire after 300 seconds. The removal of a locally learned entry from the local VPLS FDB triggers the removal of the corresponding BGP EVPN learned entries in the remote VPLS FDBs. Similar output applies for the **ping** commands in the VPRN 11 context on PE-2 and for the **ping** commands in the VPRN 11 context on PE-3.

```
*A:PE-1# show service id 12 fdb detail

================================================================================
Forwarding Database, Service 12
================================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
--------------------------------------------------------------------------------
12         00:00:5e:00:53:01 sap:1/1/c1/1:11         L/0      12/20/22 12:53:36
```

```
12         00:00:5e:00:53:02 srv6-1:                 Evpn     12/20/22 12:53:36
                             2001:db8::2:2
           2001:db8:aaaa:2:8000::
12         00:00:5e:00:53:03 srv6-1:                 Evpn     12/20/22 12:53:44
                             2001:db8::2:3
           2001:db8:aaaa:3:8000::
-------------------------------------------------------------------------------
No. of MAC Entries: 3
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

Next to the locally learned MAC address for the locally configured interface, VPRN 11 on PE-1 has two dynamically learned MAC addresses, one for each of the BGP EVPN learned MAC addresses. Similar output applies for VPRN 11 on PE-2 and for VPRN 11 on PE-3.

```
*A:PE-1# show router 11 arp summary

============================================================
ARP Table Summary (Service: 11)
============================================================
Local ARP Entries    : 1
---snip---
Dynamic ARP Entries  : 2
---snip---
------------------------------------------------------------
No. of ARP Entries   : 3
============================================================
```

The **show router 11 arp** command on PE-1 shows the association between the IP address and the MAC address, and the interface that the MAC address belongs to. The MAC address for the remote interface to the EVPN-enabled VPLS service corresponds with that of the SAP that is configured for it in VPRN 11 on PE-2 and VPRN 11 on PE-3. The association between the IP address and the MAC address for dynamically learned remote MAC addresses expires after 4 hours. Similar output applies for PE-2 and for PE-3.

```
*A:PE-1# show router 11 arp

===============================================================================
ARP Table (Service: 11)
===============================================================================
IP Address      MAC Address       Expiry    Type   Interface
-------------------------------------------------------------------------------
172.16.0.1      00:00:5e:00:53:01 00h00m00s Oth[I] local
172.16.0.2      00:00:5e:00:53:02 03h59m24s Dyn[I] local
172.16.0.3      00:00:5e:00:53:03 03h59m19s Dyn[I] local
-------------------------------------------------------------------------------
No. of ARP Entries: 3
===============================================================================
```

The **show router bgp routes evpn mac hunt** command shows a consolidated view on the MAC routes for the EVPN family. On PE-1, in the RIB In Entries section, it shows for each learned next hop how PE-1 must handle the IPv4 and IPv6 unicast data destined for it and where PE-1 must send it to. In the RIB Out Entries section, it shows for each local next hop how PE-1 expects the remote routers to handle the IPv4 and IPv6 unicast data destined for it and where PE-1 expects that data. Similar output applies for PE-2 and for PE-3.

```
*A:PE-1# show router bgp routes evpn mac hunt
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
---snip---
===============================================================================
BGP EVPN MAC Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : n/a
Nexthop        : 2001:db8::2:2
Path Id        : None
From           : 2001:db8::2:4
Res. Nexthop   : fe80::60e:1ff:fe01:1
Local Pref.    : 100                    Interface Name : int-PE-1-PE-2
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64500:1
Cluster        : 4.4.4.4
Originator Id  : 192.0.2.2              Peer Router Id : 192.0.2.4
Flags          : Used Valid Best IGP
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : MAC
ESI            : ESI-0
Tag            : 0
IP Address     : n/a
Route Dist.    : 192.0.2.2:1
Mac Address    : 00:00:5e:00:53:02
MPLS Label1    : LABEL 524288           MPLS Label2    : n/a
Route Tag      : 0
Neighbor-AS    : n/a
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h00m32s
SRv6 TLV Type  : SRv6 L2 Service TLV (6)
SRv6 SubTLV    : SRv6 SID Information (1)
Sid            : 2001:db8:aaaa:2::
Full Sid       : 2001:db8:aaaa:2:8000::
Behavior       : End.DT2U (23)
SRv6 SubSubTLV : SRv6 SID Structure (1)
Loc-Block-Len  : 48                     Loc-Node-Len   : 16
Func-Len       : 20                     Arg-Len        : 0
Tpose-Len      : 20                     Tpose-offset   : 64

Network        : n/a
Nexthop        : 2001:db8::2:3
---snip---

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
Network        : n/a
Nexthop        : 2001:db8::2:1
Path Id        : None
To             : 2001:db8::2:4
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : n/a
```

```
Connector      : None
Community      : target:64500:1
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.4
Origin         : IGP
AS-Path        : No As-Path
EVPN type      : MAC
ESI            : ESI-0
Tag            : 0
IP Address     : n/a
Route Dist.    : 192.0.2.1:1
Mac Address    : 00:00:5e:00:53:01
MPLS Label1    : 524288                 MPLS Label2    : n/a
Route Tag      : 0
Neighbor-AS    : n/a
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
SRv6 TLV Type  : SRv6 L2 Service TLV (6)
SRv6 SubTLV    : SRv6 SID Information (1)
Sid            : 2001:db8:aaaa:1::
Full Sid       : 2001:db8:aaaa:1:8000::
Behavior       : End.DT2U (23)
SRv6 SubSubTLV : SRv6 SID Structure (1)
Loc-Block-Len  : 48                     Loc-Node-Len   : 16
Func-Len       : 20                     Arg-Len        : 0
Tpose-Len      : 20                     Tpose-offset   : 64

-------------------------------------------------------------------------------
Routes : 3
===============================================================================
```

Conclusion

Distributed EVPN-enabled VPLS services can be transported over SRv6 tunnels that are automatically set up between PEs. PEs attached to the same EVPN-enabled VPLS service exchange EVPN IMET routes and MAC/IP advertisement routes that contain the SRv6 SIDs. Those SRv6 SIDs are required so that PEs can create SRv6 destinations to send unicast and BUM traffic to the other PEs in the service.

EVPN ESI Type 1

This chapter provides information about EVPN ESI Type 1.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

The information and configuration in this chapter are based on SR OS Release 22.5.R1.

Overview

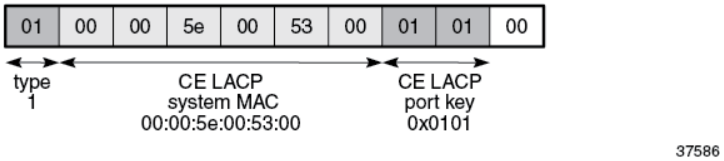
In SR OS releases earlier than 21.5.R1, the 10-byte Ethernet Segment Identifier (ESI) can only be configured manually; the auto-derived EVPN ESI type 1 (as per RFC 7432) is supported in SR OS Release 21.5.R1 and later. The **auto-esi** command is used to configure the ESI mode.

```
*A:PE-2>config>service>system>bgp-evpn>eth-seg$ auto-esi ?
  - auto-esi {none|type-1}
```

The default **auto-esi** value is **none**, which forces the user to configure the 10-byte ESI manually. When **type-1** is configured, a manual ESI cannot be configured and the ESI is auto-derived, as per RFC 7432.

ESI type 1 is auto-derived from the CE's Link Aggregation Control Protocol (LACP) system MAC address and port key. Figure 58: ESI type 1 example shows an example of ESI type 1 for LACP system MAC address 00:00:5e:00:53:00 and administrative key 257 (= 0x0101).

Figure 58: ESI type 1 example



RFC 7432, section "Ethernet Segment", defines ESI type 1 as follows:

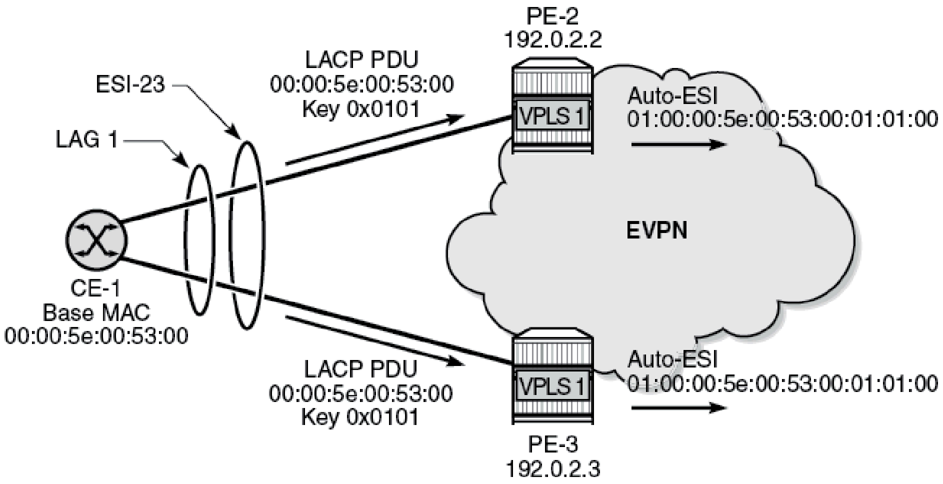
- Type 0x01 (byte 0)
- CE LACP system MAC address (bytes 1 through 6); for example, 00:00:5e:00:53:00
- CE LACP port key (bytes 7 and 8); for example, 0x0101

- 0x00 (byte 9 must be zero)

As per RFC 7432, this mechanism can only be used if the ESIs are unique, so the CE LACP system MAC and LACP port key combinations must be unique in the network.

Figure 59: ESI auto-configuration example shows the example where CE-1 has LACP system MAC address 00:00:5e:00:53:00 and LACP port key 257 (= 0x0101). CE-1 sends Link Aggregation Control Protocol Data Units (LACPDUs) to PE-2 and PE-3 with these values. Both PE-2 and PE-3 use ESI 01:00:00:5e:00:53:00:01:01:00 in ES "ESI-23". This applies both to all-active and to single-active ESs.

Figure 59: ESI auto-configuration example



The CE treats both PE-2 and PE-3 as the same switch. This allows the CE to aggregate links that are attached to different PEs in the same bundle.

When the ES LAG goes operationally down, due to the ports going down or LACP going down or standby, the previously auto-derived ESI is retained. However, when the LACP information on the CE is changed, such as a different LACP port key, the ES goes down and a new ESI will be generated.

The all-active ES "AA-ESI-23" with ESI type 1 is configured as follows:

```
# on PE-2, PE-3:
configure
    service
        system
            bgp-evpn
                ethernet-segment "AA-ESI-23" create
                    auto-esi type-1
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    ac-df-capability exclude
                    lag 1
                    no shutdown
                exit
```

The following restrictions apply for ESI type 1:

- ESI type 1 is only supported on non-virtual (regular) ESs. The following error message is raised when attempting to configure **auto-esi type-1** for a virtual ES:

```
*A:PE-2>config>service>system>bgp-evpn# ethernet-segment "vES-23" virtual create
*A:PE-2>config>service>system>bgp-evpn>eth-seg$ auto-esi type-1
MINOR: SVCMGR #8050 Ethernet segment config cannot be modified - auto-esi not supported with
 virtual ethernet-segment
```

- ESI type 1 is not supported in ESs with associations other than LAG:

```
*A:PE-2>config>service>system>bgp-evpn>eth-seg$ port 1/2/1
MINOR: SVCMGR #8048 Ethernet segment association is not valid - not allowed with auto-esi

*A:PE-2>config>service>system>bgp-evpn>eth-seg# sdp 24
MINOR: SVCMGR #8048 Ethernet segment association is not valid - not allowed with auto-esi
```

- An ES with ESI type 1 can only be enabled if the LAG has LACP enabled:

```
*A:PE-2>config>service>system>bgp-evpn>eth-seg$ lag 4
*A:PE-2>config>service>system>bgp-evpn>eth-seg$ no shutdown
MINOR: SVCMGR #8057 Ethernet segment cannot change admin state - LACP not enabled on LAG for
 auto-esi type 1 ethernet-segment
```

- ESI type 1 is allowed with all-active and single-active ESs. When used in single-active mode, the CE must use a single LAG to connect to the multi-homed PEs.
- It is not possible to manually configure an ESI when **auto-esi type-1** is configured:

```
*A:PE-2>config>service>system>bgp-evpn>eth-seg# esi 01:00:00:00:00:23:00:00:00:01
MINOR: SVCMGR #8050 Ethernet segment config cannot be modified - esi value and auto-esi type
 incompatible
```

- An ES with a manually configured ESI cannot be created with the same ESI value as the auto-derived ESI type 1 in another ES.

```
*A:PE-2>config>service>system>bgp-evpn>eth-seg# esi 01:00:00:5e:00:53:00:01:01:00
MINOR: SVCMGR #8047 Ethernet segment id is not valid - ESI already in use by another
 ethernet segment
```

- If an ES with manual ESI is active and another ES is configured with an auto-derived ESI with the same value as the manual ESI, the auto-ESI value is deleted, and a log event is added to log "99":

```
# in log "99":
97 2022/05/20 15:21:23.873 UTC MINOR: SVCMGR #2610 Base
"The Auto Ethernet segment identifier type-1 has been deleted for Ethernet Segment AA-ESI-23
 because the new ID 01:00:00:5e:00:53:00:01:01:00 conflicts with ES AA-ESI-23-5"
```

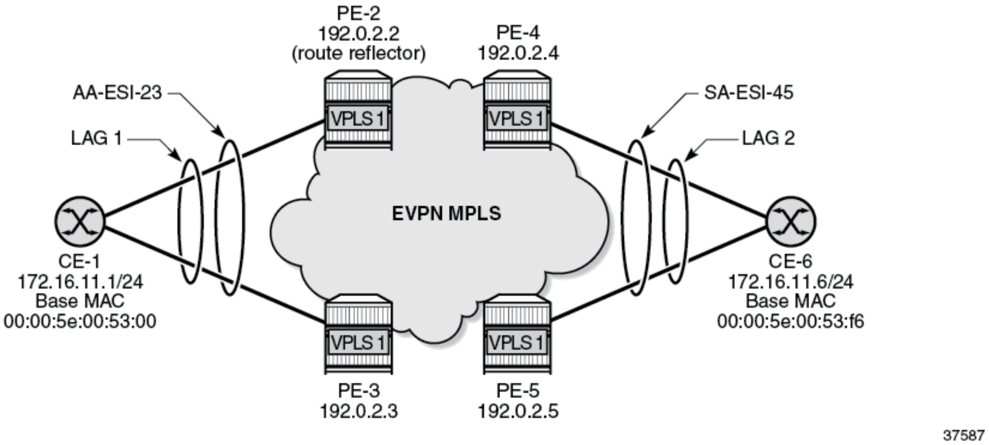
Configuration

In this section, ESI type 1 is configured in the following use cases:

- ESI type 1 in all-active ESs
- ESI type 1 in single-active ESs

Figure 60: Example topology shows the example topology with four PEs and two CEs. CE-1 is connected via LAG 1 to the all-active ES "AA-ESI-23" on PE-2 and PE-3; CE-6 is connected via LAG-2 to the single-active ES "SA-ESI-45" on PE-4 and PE-5. In this example, an EVPN-MPLS VPLS is configured, but other services are also supported.

Figure 60: Example topology



The initial configuration includes:

- cards, MDAs, ports
- on PEs: router interfaces, IS-IS, LDP

On the PEs, BGP is configured for the EVPN address family. PE-2 acts as the route reflector with the following configuration:

```
# on PE-2:
configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            rapid-update evpn
            group "internal"
                family evpn
                cluster 1.1.1.1
                peer-as 64500
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
```

On CE-1, LAG 1 is configured with LACP enabled and administrative key 257, as follows:

```
# on CE-1:
configure
```

```
    lag 1 name "lag-1"
        mode hybrid
        encap-type dot1q
        port 1/1/1
        port 1/1/2
        lacp active administrative-key 257
        no shutdown
```

The LACP system MAC address of CE-1 can be retrieved with the following command:

```
*A:CE-1# show chassis | match MAC
  Base MAC address                  : 00:00:5e:00:53:00
```

ESI type 1 in all-active ESs

On PE-2 and PE-3, the all-active ES "AA-ESI-23" is configured with **auto-esi type-1** and LAG 1:

```
# on PE-2, PE-3:
configure
    service
        system
            bgp-evpn
                ethernet-segment "AA-ESI-23" create
                    auto-esi type-1
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 1
                    no shutdown
                exit
```

The EVPN-MPLS VPLS 1 is configured as follows:

```
# on PE-2, PE-3:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap lag-1:1 create
                no shutdown
            exit
            no shutdown
        exit
```

The operational ESI on PE-2 is 01:00:00:5e:00:53:00:01:01:00 for CE LACP system MAC address 00:00:5e:00:53:00 and administrative key 0x0101, as can be verified with the following command:

```
*A:PE-2# show service system bgp-evpn ethernet-segment name "AA-ESI-23"

===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : AA-ESI-23
Eth Seg Type            : None
Admin State             : Enabled            Oper State         : Up
ESI                     : auto-esi
Oper ESI                : 01:00:00:5e:00:53:00:01:01:00
Auto-ESI Type           : Type 1
AC DF Capability        : Include
Multi-homing            : allActive          Oper Multi-homing  : allActive
ES SHG Label            : 524283
Source BMAC LSB         : None
Lag Id                  : 1
ES Activation Timer     : 3 secs (default)
Oper Group              : (Not Specified)
Svc Carving             : auto               Oper Svc Carving   : auto
Cfg Range Type          : primary
===============================================================================
```

This output is slightly different for a manually configured ES, as follows:

```
# on PE-2, PE-3:
configure
    service
        system
            bgp-evpn
                ethernet-segment "AA-ESI-23-5"
                    esi 01:00:00:00:00:23:05:00:00:01
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 5
                    no shutdown
                exit
```

```
*A:PE-2# show service system bgp-evpn ethernet-segment name "AA-ESI-23-5"

===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : AA-ESI-23-5
Eth Seg Type            : None
Admin State             : Enabled            Oper State         : Up
ESI                     : 01:00:00:00:00:23:05:00:00:01
Oper ESI                : 01:00:00:00:00:23:05:00:00:01
Auto-ESI Type           : None
AC DF Capability        : Include
Multi-homing            : allActive          Oper Multi-homing  : allActive
ES SHG Label            : 524282
Source BMAC LSB         : None
Lag Id                  : 5
ES Activation Timer     : 3 secs (default)
Oper Group              : (Not Specified)
Svc Carving             : auto               Oper Svc Carving   : auto
Cfg Range Type          : primary
```

```
===============================================================================
```

ESI type 1 in single-active ESs

CE-6 is connected via LAG 2 to the single-active ES "SA-ESI-45" on PE-4 and PE-5. An ES operational group and LAG monitor operational group is required in this use case.

On CE-6, LAG 2 is configured with LACP enabled and administrative key 32768 (= 0x8000), as follows:

```
# on CE-6:
configure
    lag 2 name "lag-2"
        mode hybrid
        encap-type dot1q
        port 1/1/1
        port 1/1/2
        lacp active administrative-key 32768
        no shutdown
```

The LACP system MAC address of CE-6 is the following:

```
*A:CE-6# show chassis | match MAC
  Base MAC address                  : 00:00:5e:00:53:f6
```

On PE-4 and PE-5, operational group "op-grp-2" is configured and assigned to single-active ES "SA-ESI-45".

Note: When an operational group is associated to an ES, the hold timers for the operational group must be zero (the default value for the group down timer).

LAG 2 monitors this operational group. The configuration is as follows:

```
# on PE-4:
configure
    service
        oper-group "op-grp-2" create
            hold-time
                group down 0    # default
                group up 0
            exit
       exit
    exit
    lag 2 name "lag-2"
        mode access
        encap-type dot1q
        monitor-oper-group "op-grp-2"
        port 1/1/1
        lacp active administrative-key 1 system-id 00:00:00:00:45:02
        no shutdown
    exit
    service
        system
            bgp-evpn
                ethernet-segment "SA-ESI-45" create
                    auto-esi type-1
                    service-carving
                        mode manual        # required for oper-group
                        manual
```

```
                            preference non-revertive create
                                value 200
                            exit
                        exit
                    exit
                    multi-homing single-active
                    ac-df-capability exclude
                    lag 2
                    oper-group "op-grp-2"
                    no shutdown
                exit
            exit
        exit
        vpls 1 name "VPLS 1" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap lag-2:1 create
                no shutdown
            exit
            no shutdown
        exit
```

The following command on Designated Forwarder (DF) PE-4 shows that the operational ESI is 01:00:00:5e:00:53:f6:80:00:00:

```
# on PE-4:
*A:PE-4# show service system bgp-evpn ethernet-segment name "SA-ESI-45" all

===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : SA-ESI-45
Eth Seg Type            : None
Admin State             : Enabled            Oper State         : Up
ESI                     : auto-esi
Oper ESI                : 01:00:00:5e:00:53:f6:80:00:00
Auto-ESI Type           : Type 1
AC DF Capability        : Exclude
Multi-homing            : singleActive       Oper Multi-homing  : singleActive
ES SHG Label            : 524283
Source BMAC LSB         : None
Lag Id                  : 2
ES Activation Timer     : 3 secs (default)
Oper Group              : op-grp-2
Svc Carving             : manual             Oper Svc Carving   : manual
Cfg Range Type          : lowest-pref

-------------------------------------------------------------------------------
DF Pref Election Information
```

```
-------------------------------------------------------------------------------
Preference     Preference     Last Admin Change        Oper Pref      Do No
Mode           Value                                   Value          Preempt
-------------------------------------------------------------------------------
non-revertive  200            06/08/2022 15:02:13      200            Enabled
-------------------------------------------------------------------------------
EVI Ranges: <none>
ISID Ranges: <none>
===============================================================================

===============================================================================
EVI Information
===============================================================================
EVI                 SvcId               Actv Timer Rem      DF
-------------------------------------------------------------------------------
1                   1                   0                   yes
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================

-------------------------------------------------------------------------------
DF Candidate list
-------------------------------------------------------------------------------
EVI                                     DF Address
-------------------------------------------------------------------------------
1                                       192.0.2.4
1                                       192.0.2.5
-------------------------------------------------------------------------------
Number of entries: 2
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
---snip---
```

The operational ESI on Non-Designated Forwarder (NDF) PE-5 is the same as for PE-4.

The operational status of the operational group "op-grp-2" on DF PE-4 is up, while it is down on NDF PE-5 where the ES is inactive, as follows:

```
*A:PE-4# show service oper-group "op-grp-2"

===============================================================================
Service Oper Group Information
===============================================================================
Oper Group       : op-grp-2
Creation Origin  : manual                          Oper Status: up
Hold DownTime    : 0 secs                          Hold UpTime: 0 secs
Members          : 1                               Monitoring : 1
===============================================================================


*A:PE-5# show service oper-group "op-grp-2" detail

===============================================================================
Service Oper Group Information
===============================================================================
Oper Group       : op-grp-2
Creation Origin  : manual                          Oper Status: down
Hold DownTime    : 0 secs                          Hold UpTime: 0 secs
Members          : 1                               Monitoring : 1
===============================================================================

===============================================================================
Member Ethernet-Segment for OperGroup: op-grp-2
```

```
===============================================================================
Ethernet-Segment                        Status
-------------------------------------------------------------------------------
SA-ESI-45                               Inactive
-------------------------------------------------------------------------------
Ethernet-Segment Entries found: 1
===============================================================================

===============================================================================
Monitoring LAG for OperGroup: op-grp-2
===============================================================================
Lag-id         Adm       Opr       Weighted  Threshold Up-Count  Act/Stdby
    name
-------------------------------------------------------------------------------
2              up        down      No        0         0         N/A
    lag-2
-------------------------------------------------------------------------------
LAG Entries found: 1
===============================================================================
port option not supported with monitoring
```

LAG 2 monitors the operational group "op-grp-2", so it follows the state of the ES "SA-ESI-45". On DF PE-4, LAG 2 is operationally up:

```
*A:PE-4# show lag "lag-2"

===============================================================================
Lag Data
===============================================================================
Lag-id         Adm     Opr     Weighted Threshold Up-Count MC Act/Stdby
    name
-------------------------------------------------------------------------------
2              up      up      No       0         1        N/A
    lag-2
===============================================================================
```

On NDF PE-5, LAG 2 is operationally down with reason operGroupDown:

```
*A:PE-5# show lag "lag-2" detail

===============================================================================
LAG Details
===============================================================================
Description        : N/A
-------------------------------------------------------------------------------
Details
-------------------------------------------------------------------------------
Lag-id              : 2                     Mode                 : access
Lag-name            : lag-2
Adm                 : up                    Opr                  : down
Reason Down         : operGroupDown
Thres. Last Cleared : 05/20/2022 14:57:23   Thres. Exceeded Cnt  : 0
Dynamic Cost        : false                 Encap Type           : dot1q
Configured Address  : 02:1f:ff:00:01:42     Lag-IfIndex          : 1342177282
Hardware Address    : 02:1f:ff:00:01:42     Adapt Qos (access)   : distribute
Hold-time Down      : 0.0 sec               Port Type            : standard
Per-Link-Hash       : disabled
Include-Egr-Hash-Cfg: disabled              Forced               : -
Per FP Ing Queuing  : disabled              Per FP Egr Queuing   : disabled
Per FP SAP Instance : disabled
Access Bandwidth    : N/A                   Access Booking Factor: 100
Access Available BW : 0
Access Booked BW    : 0
LACP                : enabled               Mode                 : active
LACP Transmit Intvl : fast                  LACP xmit stdby      : enabled
Selection Criteria  : highest-count         Slave-to-partner     : disabled
MUX control         : coupled
Subgrp hold time    : 0.0 sec               Remaining time       : 0.0 sec
Subgrp selected     : 1                     Subgrp candidate     : -
Subgrp count        : 1
System Id           : 00:00:00:00:45:02     System Priority      : 32768
Admin Key           : 1                     Oper Key             : 1
Prtr System Id      : 00:00:5e:00:53:f6     Prtr System Priority : 32768
Prtr Oper Key       : 32768
Standby Signaling   : lacp
Port hashing        : port-speed            Port weight speed    : 0 gbps
Ports Up            : 0
Weights Up          : 0                     Hash-Weights Up      : 0
Monitor oper group  : op-grp-2
Oper group status   : down
Adaptive loadbal.   : disabled              Tolerance            : N/A

-------------------------------------------------------------------------------
Port-id        Adm     Act/Stdby Opr     Primary   Sub-group     Forced  Prio
-------------------------------------------------------------------------------
1/1/2          up      active    down    yes       1             -       32768

-------------------------------------------------------------------------------
Port-id        Role      Exp   Def   Dist  Col   Syn   Aggr  Timeout  Activity
-------------------------------------------------------------------------------
1/1/2          actor     No    No    No    No    No    Yes   Yes      Yes
1/1/2          partner   No    No    No    No    Yes   Yes   Yes      Yes
===============================================================================
```

When the LAG is operationally down, the SAP is operationally down. On DF PE-4, the SAP is up:

```
*A:PE-4# show service id 1 sap

===============================================================================
SAP(Summary), Service 1
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
lag-2:1                         1          1     none    1     none   Up   Up
-------------------------------------------------------------------------------
Number of SAPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

On NDF PE-5, the SAP is operationally down:

```
*A:PE-5# show service id 1 sap lag-2:1

===============================================================================
Service Access Points(SAP)
===============================================================================
Service Id         : 1
SAP                : lag-2:1                Encap              : q-tag
Description        : (Not Specified)
Admin State        : Up                     Oper State         : Down
Flags              : PortOperDown StandByForMHProtocol
Multi Svc Site     : None
Last Status Change : 05/20/2022 15:02:07
Last Mgmt Change   : 05/20/2022 15:01:15
```

```
===============================================================================
```

Auto-derived ESI changes when LACP port key on CE is modified

When the LAG goes operationally down due to ports going down or LACP going down, the auto-derived ESI is preserved. However, when the CE LACP configuration is changed— for example, with a different LACP port key—a new ESI is auto-derived.

In this example, the initial operational ESI on PE-4 is 01:00:00:5e:00:53:f6:80:00:00, as follows:

```
*A:PE-4# show service system bgp-evpn ethernet-segment name "SA-ESI-45" | match ESI
Name                    : SA-ESI-45
ESI                     : auto-esi
Oper ESI                : 01:00:00:5e:00:53:f6:80:00:00
Auto-ESI Type           : Type 1
```

On CE-6, the initial configuration of LAG 2 has LACP active with administrative key 32768:

```
*A:CE-6>config>lag# info
----------------------------------------------
        mode hybrid
        encap-type dot1q
        port 1/1/1
        port 1/1/2
        lacp active administrative-key 32768
        no shutdown
----------------------------------------------
```

On CE-6, LAG 2 is reconfigured with administrative key 4095 (= 0x0fff), as follows:

```
# on CE-6:
configure
    lag 2 name "lag-2"
        mode hybrid
        encap-type dot1q
        port 1/1/1
        port 1/1/2
        lacp active administrative-key 4095
        no shutdown
```

As a result, the operational ESI on PE-4 is 01:00:00:5e:00:53:f6:0f:ff:00, as follows:

```
*A:PE-4# show service system bgp-evpn ethernet-segment name "SA-ESI-45" | match ESI
Name                    : SA-ESI-45
ESI                     : auto-esi
Oper ESI                : 01:00:00:5e:00:53:f6:0f:ff:00
Auto-ESI Type           : Type 1
```

When debugging is enabled for BGP updates, the following ES routes are seen: initially with ESI 01:00:00:5e:00:53:f6:**80:00**:00 and later with ESI 01:00:00:5e:00:53:f6:**0f:ff**:00, as follows:

```
39 2022/06/08 15:02:18.970 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 71
    Flag: 0x90 Type: 14 Len: 34 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.4
        Type: EVPN-ETH-SEG Len: 23 RD: 192.0.2.4:0 ESI: 01:00:00:5e:00:53:f6:80:00:00, IP-Len:
 4 Orig-IP-Addr: 192.0.2.4
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        df-election::DF-Type:Preference/DP:1/DF-Preference:200/AC:0
        target:00:00:5e:00:53:f6
"

---snip---

56 2022/06/08 15:10:53.605 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 71
    Flag: 0x90 Type: 14 Len: 34 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.4
        Type: EVPN-ETH-SEG Len: 23 RD: 192.0.2.4:0 ESI: 01:00:00:5e:00:53:f6:0f:ff:00, IP-Len:
 4 Orig-IP-Addr: 192.0.2.4
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        df-election::DF-Type:Preference/DP:1/DF-Preference:200/AC:0
        target:00:00:5e:00:53:f6
"
```

Conclusion

To simplify the configuration of single-active and all-active ESs with LAG association, ESI type 1 can be used to auto-derive the ESI from the CE's LACP system MAC address and LACP port key.

EVPN for MPLS Tunnels

This chapter provides information about EVPN for MPLS tunnels.

Topics in this chapter include:

Applicability

This chapter was initially written for SR OS Release 13.0.R6, but the CLI in the current edition corresponds to SR OS Release 21.2.R1. A prerequisite is to read the EVPN for VXLAN Tunnels (Layer 2) chapter.

Overview

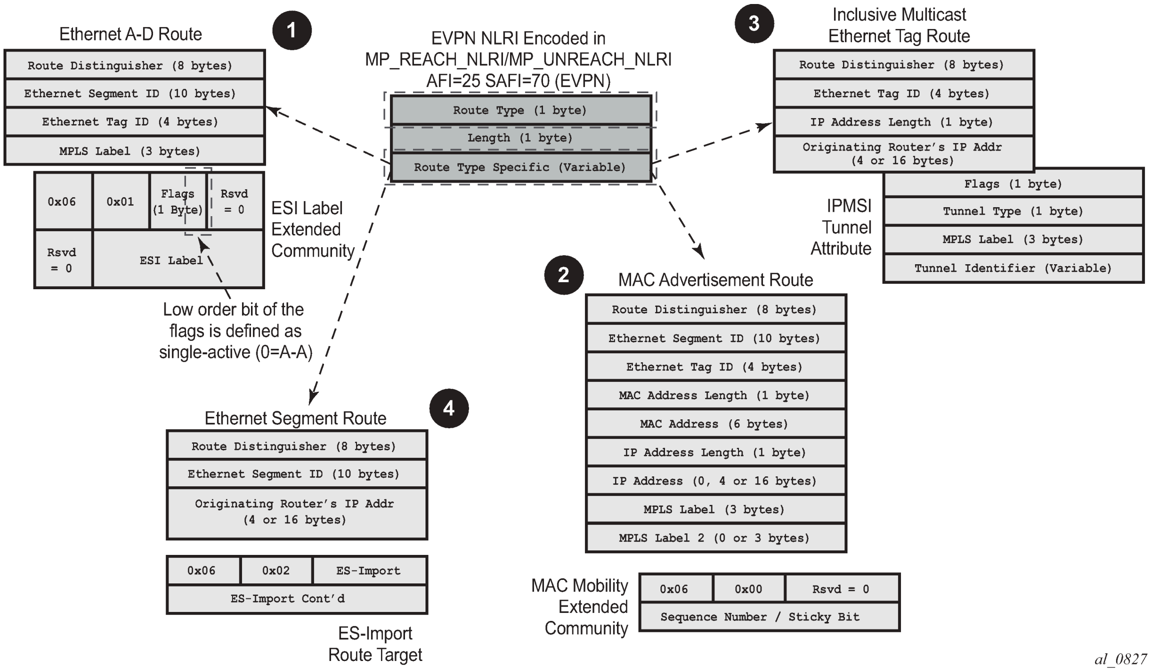
EVPN-MPLS is standardized in RFC 7432, *BGP MPLS-Based Ethernet VPN*, as a Layer 2 VPN technology that can supplement VPLS for E-LAN services. Besides the optimizations introduced by EVPN, a significant number of service providers offering E-LAN services today are requesting EVPN for their multi-homing capabilities. EVPN supports all-active multi-homing (per-flow load-balancing multi-homing) as well as single-active multi-homing (per-service load-balancing multi-homing). In addition to those superior multi-homing capabilities, EVPN also provides a number of significant benefits, such as:

- IP-VPN-like operation and control for E-LAN services.
- Reduction and (in some cases) suppression of the Broadcast, Unknown unicast, and Multicast (BUM) traffic in the network.
- Simple provisioning and management.
- New set of tools to control the distribution of MAC addresses and Address Resolution Protocol (ARP) entries in the network.

The EVPN for Virtual eXtensible Local Area Network (VXLAN) tunnels (Layer 2) chapter focuses on the use of EVPN as a control plane for VXLAN tunnels, whereas this chapter provides configuration guidelines for EVPN when used for MPLS tunnels. Similar to EVPN-VXLAN services, VPLS services with EVPN for MPLS tunnels are referred to as EVPN-MPLS services.

As a reference, the EVPN route types and NLRIs (Network Layer Reachability Information messages) used by the EVPN family in RFC 7432 are shown in Figure 61: EVPN route types and NLRIs.

Figure 61: EVPN route types and NLRIs


Ethernet A-D Route
1
Route Distinguisher (8 bytes)
Ethernet Segment ID (10 bytes)
Ethernet Tag ID (4 bytes)
MPLS Label (3 bytes)
0x06
0x01
Flags (1 Byte)
Rsvd = 0
Rsvd = 0
ESI Label
ESI Label Extended Community
Low order bit of the flags is defined as single-active (0=A-A)
EVPN NLRI Encoded in MP_REACH_NLRI/MP_UNREACH_NLRI AFI=25 SAFI=70 (EVPN)
Route Type (1 byte)
Length (1 byte)
Route Type Specific (Variable)
3
Inclusive Multicast Ethernet Tag Route
Route Distinguisher (8 bytes)
Ethernet Tag ID (4 bytes)
IP Address Length (1 byte)
Originating Router's IP Addr (4 or 16 bytes)
IPMSI Tunnel Attribute
Flags (1 byte)
Tunnel Type (1 byte)
MPLS Label (3 bytes)
Tunnel Identifier (Variable)
2
MAC Advertisement Route
Route Distinguisher (8 bytes)
Ethernet Segment ID (10 bytes)
Ethernet Tag ID (4 bytes)
MAC Address Length (1 byte)
MAC Address (6 bytes)
IP Address Length (1 byte)
IP Address (0, 4 or 16 bytes)
MPLS Label (3 bytes)
MPLS Label 2 (0 or 3 bytes)
MAC Mobility Extended Community
0x06
0x00
Rsvd = 0
Sequence Number / Sticky Bit
Ethernet Segment Route
4
Route Distinguisher (8 bytes)
Ethernet Segment ID (10 bytes)
Originating Router's IP Addr (4 or 16 bytes)
0x06
0x02
ES-Import
ES-Import Cont'd
ES-Import Route Target
al_0827

When no EVPN multi-homing is used in the network, only the base routes are used. Route types 2 and 3 are considered the base and mandatory routes:

- Route type 2 - MAC/IP route: This route advertises MAC addresses to be installed in the remote FDBs, or MAC/IP address pairs to be installed in the remote proxy-ARP/ND (Neighbor Discovery) tables.
- Route type 3 - Inclusive multicast route: This route advertises the multicast tree that the advertising PE intends to use for sending BUM traffic for an EVPN Instance (EVI). Ingress Replication, Point-to-multipoint multicast Label Distribution Protocol (P2MP mLDP), and composite tunnels are supported as tunnel types in route type 3 when BGP-EVPN MPLS is enabled. The ingress replication information, as well as the downstream MPLS label (for remote PEs to send BUM traffic to the advertising PE) are encoded in the Provider Multicast Service Interface Tunnel Attribute (PTA).

When EVPN multi-homing is used in an EVI, routes type 1 and 4 are used (where type 1 has two different purposes):

- Route type 1 - Auto-discovery per Ethernet segment (AD per ES) route: This route is advertised per ES from the PE, carries the Ethernet Segment Identifier (ESI) label (used for split-horizon) in multi-homing mode, and can affect procedures such as the Designated Forwarder (DF) election, as well as the aliasing/backup path/mass withdrawal on remote PEs.
- Route type 1 - Auto-discovery per EVPN instance (AD per-EVI) route: This route allows the remote PEs to provide aliasing and a backup path to the PEs part of the ES.
- Route type 4 - Ethernet Segment (ES) route: This route advertises a local configured ES. The exchange of this route can discover remote PEs that are part of the same ES and the DF election algorithm among them.

The AD per-EVI, MAC/IP, and inclusive multicast routes are considered service-level BGP-EVPN routes. Their RT/RD (Route-Target/Route-Distinguisher) are taken from the VPLS configuration.

The AD per-ES and the ES routes are considered base-level BGP-EVPN routes. However, their RT/RD are taken differently:

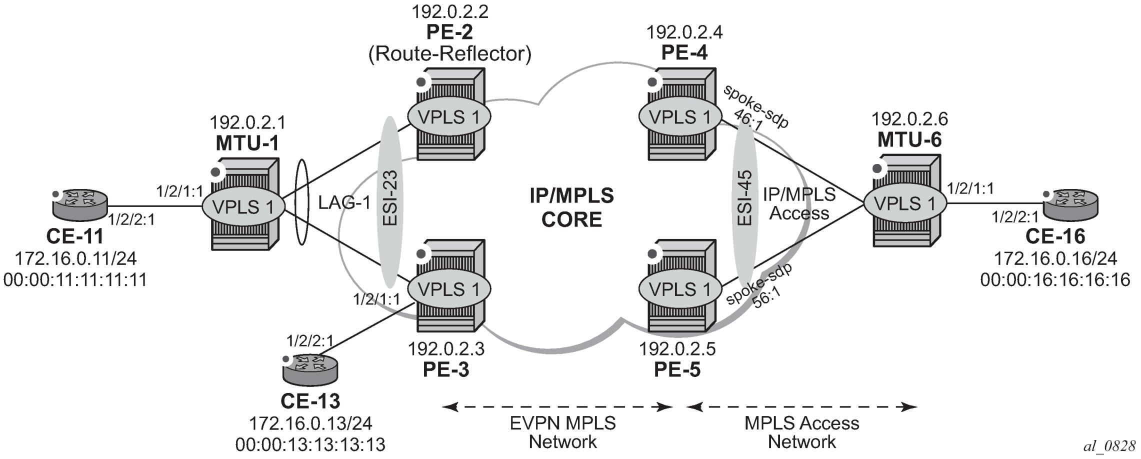
- The ES route RD is taken from the **service>system>bgp-evpn** configuration. The ES route RT is auto-derived from the Ethernet segment.
- The AD per-ES route RD is taken from the system level RD or service level RD. The RT extended community is taken from the service level RT or an RT set for the services defined on the Ethernet segment.

Configuration

This section describes the configuration of EVPN-MPLS for Layer 2 services on SR OS, as well as the available troubleshooting and show commands, and EVPN multi-homing.

Figure 62: EVPN-MPLS for VPLS services shows the topology used throughout this chapter. The network consists of a core with four EVPN PEs (PE-2, PE-3, PE-4, and PE-5) and two MTU devices that are dual-homed to the EVPN network. For MTU-1, all-active multi-homing is used, whereas MTU-6 is connected via single-active multi-homing to the EVPN network. Three CEs are connected to VPLS 1 in MTU-1, PE-3, and MTU-6 in order to test the connectivity.

Figure 62: EVPN-MPLS for VPLS services



As part of the network infrastructure configuration, the following settings and protocols must be added to the configuration before starting with the EVPN-specific configuration for the services:

- The ports interconnecting the four PEs in the core are configured as network ports (or hybrid) and will have router network interfaces defined in them. The ports on PE-2/PE-3 connected to MTU-1 can be access or hybrid ports, whereas the ports on PE-4/PE-5 connected to MTU-6 can be network or hybrid ports. In case of hybrid ports, no LACP can be configured.
- The four PEs in the core (as well as MTU-6 in the access MPLS network) are running IS-IS and establishing point-to-point adjacencies for the exchange of the system IP addresses.
- LDP is used as the MPLS protocol to signal transport tunnel labels among PE-2, PE-3, PE-4, PE-5, and MTU-6. There is no LDP running between MTU-1 and the rest of the network, that is, MTU-1 is a pure Ethernet aggregation device.

- EVPN uses MP-BGP for exchanging reachability at service level. Therefore, BGP peering sessions must be established among the core PEs for the EVPN family. Although typically a separate router is used, in this chapter, PE-2 is used as BGP RR (route reflector) for EVPN routes. For example, the following output shows the configuration of BGP in the RR and one of the BGP clients. The relevant commands for EVPN are shown in bold.

  The configuration on the route reflector PE-2 is as follows:

```
# on RR PE-2:
configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family evpn
                cluster 1.1.1.1
                peer-as 64500
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
```

  The BGP configuration on the clients PE-3, PE-4, and PE-5 is as follows:

```
# on RR clients PE-3, PE-4, PE-5:
configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family evpn
                peer-as 64500
                neighbor 192.0.2.2
                exit
            exit
```

Note:
The **def-recv-evpn-encap** command is not used in the preceding configuration because the default MPLS configuration is sufficient to have a correct interpretation of the received EVPN encapsulations.

The EVPN encapsulation type can be configured as MPLS or VXLAN, as follows:

```
*A:PE-3# configure router bgp group "internal" neighbor 192.0.2.2
                                                             def-recv-evpn-encap ?
```

```
 - no def-recv-evpn-encap
 - def-recv-evpn-encap <encap-type>

 <encap-type>         : mpls|vxlan
```

EVPN routes type 1 (auto-discovery per-EVI route), type 2 (MAC/IP route), type 3 (inclusive multicast route), and type 5 (IP-prefix route) are always sent with the RFC 5512, *the BGP Encapsulation Subsequent Address Family Identifier (SAFI) and the BGP Tunnel Encapsulation Attribute*, BGP encapsulation extended community that indicates the associated encapsulation of the route. Because the use of this extended community is not mandatory in RFC 7432, the **def-recv-evpn-encap** command indicates to the system what encapsulation is associated with routes received without any encapsulation. When interoperating with third-party EVPN vendors in mixed MPLS and EVPN-VXLAN networks, this command should be revised accordingly.

EVPN-MPLS configuration without multi-homing

After the base infrastructure (interfaces, IGP, LDP, BGP protocols) is configured, the service and EVPN can be enabled. When no multi-homing is used, the EVPN-MPLS configuration in a VPLS service looks similar to the configuration of EVPN-VXLAN for Layer 2, except for the commands related to the MPLS data plane. The following output shows the VPLS-1 configuration on PE-3 as an example:

```
# on PE-3:
configure
    service
        vpls 1 name "VPLS1" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/2/1:1 create
                no shutdown
            exit
            sap lag-1:1 create
                no shutdown
            exit
            no shutdown
```

Where the following commands are relevant for a basic EVPN configuration:

- **bgp** enables the context for the BGP configuration relevant to the service. If a manual (non-auto-derived) RD/RT, as well as import/export policies, are needed for the service, the commands in the **bgp** context must be configured. When **bgp-evpn** is enabled in a VPLS instance, other families are supported within the same service (bgp-ad, bgp-mh, bgp-vpls). This **bgp** context configures the common BGP parameters for all the BGP families in the service. Even if the general BGP parameters for the service are auto-derived (as in this example), the **bgp** context must be enabled.

```
*A:PE-3>config>service>vpls# bgp ?
  - bgp [<bgp-instance>]
  - no bgp [<bgp-instance>]

 <bgp-instance>       : [1..2]


 [no] pw-template-bi* + Configure pw-template bind policy
 [no] route-distingu* - Configure route distinguisher
 [no] route-target    - Configure route target
 [no] vsi-export      - VSI export route policies
 [no] vsi-import      - VSI import route policies
```

- **bgp-evpn evi <1..65535>** — The EVPN instance or EVI is a 2-byte identifier used for the auto-derivation of the service RD, service RT, and for the service-carving algorithm when multi-homing is used. The EVI can be used for both **bgp-evpn vxlan** and **bgp-evpn mpls** when the user needs to auto-derive the RD and RT for the service. The auto-derivation is always based on:
  – RD system-ip:evi
  – RT autonomous-system:evi

  The configured and operating RD/RT values can be checked with the following show command (in this example, the evi value is 1):

```
*A:PE-3# show service id 1 bgp

===============================================================================
BGP Information
===============================================================================
Bgp Instance         : 1
Vsi-Import           : None
Vsi-Export           : None
Route Dist           : None
Oper Route Dist      : 192.0.2.3:1
Oper RD Type         : derivedEvi
Rte-Target Import    : None                 Rte-Target Export: None
Oper RT Imp Origin   : derivedEvi           Oper RT Import   : 64500:1
Oper RT Exp Origin   : derivedEvi           Oper RT Export   : 64500:1

PW-Template Id       : None
-------------------------------------------------------------------------------
===============================================================================
```

Although not required for a basic BGP-EVPN MPLS configuration, some other parameters may be used at the **bgp-evpn** context level, when EVPN-MPLS services are deployed. Some examples are listed here:

- **bgp-evpn cfm-mac-advertisement** must be enabled when eth-cfm is used across an EVPN-MPLS service among different PEs. If a Maintenance Endpoint (MEP) or Maintenance domain Intermediate Point (MIP) is configured in any of the SAP/SDP bindings in the VPLS and has to exchange eth-cfm packets with a remote MEP/MIP across the EVPN-MPLS core, this command must be enabled. In that way, the MEP/MIP MAC address can be advertised in EVPN (otherwise, the MEP/MIP MAC address would not be learned on remote EVPN-MPLS PEs and eth-cfm would not work correctly).
- **bgp-evpn mac-advertisement** and **bgp-evpn mac-duplication** — See the EVPN for VXLAN Tunnels (Layer 2) chapter for a description of these two commands.

- **bgp-evpn mpls** must be enabled.

When two BGP instances are added to a VPLS service, both BGP-EVPN MPLS and BGP-EVPN VXLAN can be configured at the same time in the service. A maximum of two BGP instances are supported in the same VPLS. In this chapter, only one BGP instance will be used: BGP-EVPN MPLS uses the default BGP instance 1.

After the relevant **VPLS** parameters, **BGP** and **BGP-EVPN** attributes are added, the specific commands for **bgp-evpn mpls** can be configured as follows:

```
*A:PE-3>config>service>vpls>bgp-evpn>mpls# info
----------------------------------------------
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
----------------------------------------------
```

- **ingress-replication-bum-label** controls whether the system will advertise different service labels for unicast and BUM traffic. If no EVPN multi-homing is configured in the network, this command can be disabled (**no ingress-replication-bum-label**) and the same MPLS label will be advertised for the unicast and BUM traffic for the VPLS instance. If EVPN multi-homing is configured in the PE, this command is strongly recommended to avoid potential transient issues. See the EVPN-MPLS multi-homing section.
- **ecmp** controls the number of remote PEs to which the local PE can load balance the unicast traffic. See the EVPN multi-homing section.
- **auto-bind-tunnel** controls the resolution of EVPN destinations to MPLS transport tunnels. This command is also in VPRN services and works in the same way.
  - If the **auto-bind-tunnel resolution any** is configured, as in the example, EVPN destinations in the service are resolved based on the best tunnel in the Tunnel Table Manager (TTM). For instance, the following command shows the existing EVPN destinations for VPLS 1 in PE-3. The EVPN-MPLS destination (Termination Endpoint (TEP) 192.0.2.2, label 524282) is resolved to an LDP transport tunnel because the (best) LDP tunnel to 192.0.2.2 shown in the **show router tunnel-table** is LDP. If there was more than one tunnel type in the TTM to 192.0.2.2, the system would pick the lowest **Pref** (preference) tunnel.

```

*A:PE-3# show service id 1 evpn-mpls

===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address     Egr Label     Num. MACs   Mcast           Last Change
                Transport:Tnl                             Sup BCast Domain
-------------------------------------------------------------------------------
192.0.2.2       524282        0           bum             02/17/2021 14:06:13
                ldp:65537                                 No
192.0.2.4       524282        0           bum             02/17/2021 14:06:13
                ldp:65538                                 No
192.0.2.5       524282        0           bum             02/17/2021 14:06:13
                ldp:65539                                 No
-------------------------------------------------------------------------------
Number of entries : 3
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

```
*A:PE-3# show router tunnel-table

===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination           Owner     Encap TunnelId  Pref   Nexthop        Metric
   Color
-------------------------------------------------------------------------------
192.0.2.2/32          ldp       MPLS  65537     9      192.168.23.1   10
192.0.2.4/32          ldp       MPLS  65538     9      192.168.34.2   10
192.0.2.5/32          ldp       MPLS  65539     9      192.168.35.2   10
192.0.2.6/32          ldp       MPLS  65540     9      192.168.34.2   20
-------------------------------------------------------------------------------
Flags: B = BGP or MPLS backup hop available
       L = Loop-Free Alternate (LFA) hop available
       E = Inactive best-external BGP route
       k = RIB-API or Forwarding Policy backup hop
===============================================================================
```

- If resolution is set to **any**, the following tunnel types are selected in order of preference: RSVP, LDP, Segment Routing, and BGP. The user can configure the preference of the segment-routing tunnel type in the TTM for a specific IGP instance.
- If one or more explicit tunnel types are specified using the resolution-filter option, then only these tunnel types will be selected again following the TTM preference.
- The user must set the resolution to filter to activate the list of tunnel-types configured under resolution-filter.

Although not shown in the **bgp-evpn mpls** basic configuration for PE-3, there are other parameters that can be modified:

```
*A:PE-3>config>service>vpls>bgp-evpn# mpls ?
  - no mpls [bgp <bgp>]
  - mpls [bgp <bgp>]

 <bgp>                : [1..2]


      auto-bind-tunn* + Configure BGP EVPN mpls auto-bind-tunnel
 [no] control-word    - Enable/disable setting the CW bit in the label message
 [no] default-route-* - Configure default-route-tag to match against export policies
      ecmp            - Configure maximum ECMP routes information
 [no] entropy-label   - Enable/disable use of entropy-label
 [no] force-vlan-vc-* - Forces vlan-vc-type forwarding in the data-path
 [no] ingress-replic* - Use the same label as the one advertised for unicast traffic
 [no] oper-group      - Configure oper-group
 [no] restrict-prote* - Enable/disable protected src MAC restriction
      route-next-hop  - Configure route next-hop
 [no] send-tunnel-en* - Configure encapsulation for this service
 [no] shutdown        - Administratively Enable/Disable BGP-EVPN mpls
 [no] split-horizon-* - Configure a split-horizon-group
```

- **bgp** instance defines the BGP instance: default **bgp** or **bgp 1** can be used for either BGP-EVPN MPLS or BGP-EPVN VXLAN; **bgp 2** can only be used for BGP-EVPN MPLS.
- **control-word** enables/disables the insertion of the control-word in the data path. The control-word is disabled by default and is not signaled in EVPN (based on RFC 7432) and has to be consistently configured in all the PEs in the network. The use of the **control-word** prevents packet reordering from

happening in P routers that misinterpret the first nibble of the payload in the packets they receive. In some third-party EVPN vendors, the control-word is enabled by default, so it is recommended to enable it when interoperating with other vendors.

- **entropy-label** enables the use of entropy labels, as described in the *Entropy Label* chapter.
- **force-vlan-vc-forwarding** allows the system to preserve the VLAN ID and p-bits of the service-delimiting q-tag in a new tag added in the customer frame before sending it to the EVPN core. This command may be used with the **sap ingress vlan-translation** command: the configured translated VLAN ID will be sent to the EVPN binds, as opposed to the service-delimiting tag VLAN ID. If the ingress SAP/SDP-binding is null encapsulated, the output VLAN ID and p-bits will be zero.
- **restrict-protected-src** is by default disabled. When enabled, all packets entering the object will be verified not to contain a protected source MAC address. In combination with the parameter **discard-frame**, the packets that contain a protected MAC address will be discarded and an alarm is generated.
- **send-tunnel-encap** configures the encapsulation to be advertised with the EVPN routes for the service. The encapsulation is encoded in RFC 5512-based tunnel encapsulation extended communities. When configured in the **bgp-evpn>mpls** context, the supported options are none (no send-tunnel-encap), mpls, mplsoudp, or both.
- **shutdown** enables/disables the use of MPLS for EVPN. When **mpls no shutdown** is issued, a BGP route-refresh message is sent for the EVPN family.
- **split-horizon-group** <*group-name*> configures an explicit split-horizon-group (SHG) for all the EVPN destinations that can be shared with other SAP/SDP-bindings. See the VPLS to EVPN-MPLS integration section.

After **bgp-evpn mpls** is configured and enabled in the service, an inclusive multicast route is sent to the RR. The remote PEs receiving and importing that route will create an EVPN destination to the sending PE. An EVPN destination is identified by a TEP and MPLS label. Use the following show commands to view the service and the EVPN destinations created:

- **show service evpn-mpls**
- **show service id 1 evpn-mpls**
- **show service id 1 bgp-evpn**

An example of the output is shown for PE-2 when there is no traffic in the network. Therefore, only inclusive multicast routes have been exchanged among the four PEs.

```
*A:PE-2# show service evpn-mpls

===============================================================================
EVPN MPLS Tunnel Endpoints
===============================================================================
EvpnMplsTEP Address EVPN-MPLS Dest      ES Dest             ES BMac Dest
-------------------------------------------------------------------------------
192.0.2.3           1                   0                   0
192.0.2.4           1                   0                   0
192.0.2.5           1                   0                   0
-------------------------------------------------------------------------------
Number of EvpnMpls Tunnel Endpoints: 3
-------------------------------------------------------------------------------
===============================================================================
```

```
*A:PE-2# show service id 1 evpn-mpls

===============================================================================
BGP EVPN-MPLS Dest
```

```
===============================================================================
TEP Address     Egr Label     Num. MACs   Mcast           Last Change
                Transport:Tnl                             Sup BCast Domain
-------------------------------------------------------------------------------
192.0.2.3       524279        0           bum             02/17/2021 14:06:11
                ldp:65537                                 No
192.0.2.4       524282        0           bum             02/17/2021 14:01:56
                ldp:65538                                 No
192.0.2.5       524282        0           bum             02/17/2021 14:02:02
                ldp:65539                                 No
-------------------------------------------------------------------------------
Number of entries : 3
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                       Num. Macs               Last Change
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================

===============================================================================
BGP EVPN-MPLS ES BMAC Dest
===============================================================================
ES BMAC Addr                            Last Change
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

```
*A:PE-2# show service id 1 bgp-evpn

===============================================================================
BGP EVPN Table
===============================================================================
MAC Advertisement  : Enabled            Unknown MAC Route  : Disabled
CFM MAC Advertise  : Disabled
Creation Origin    : manual
MAC Dup Detn Moves : 5                  MAC Dup Detn Window: 3
MAC Dup Detn Retry : 9                  Number of Dup MACs : 0
MAC Dup Detn BH    : Disabled
IP Route Advert    : Disabled
Sel Mcast Advert   : Disabled

EVI                : 1
Ing Rep Inc McastAd: Enabled
Accept IVPLS Flush : Disabled

-------------------------------------------------------------------------------
Detected Duplicate MAC Addresses             Time Detected
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN MPLS Information
===============================================================================
Admin Status       : Enabled            Bgp Instance       : 1
Force Vlan Fwding  : Disabled
Route NextHop Type : system-ipv4
Control Word       : Disabled
Max Ecmp Routes    : 2
```

```
Entropy Label      : Disabled
Default Route Tag  : none
Split Horizon Group: (Not Specified)
Ingress Rep BUM Lbl: Enabled
Ingress Ucast Lbl  : 524282            Ingress Mcast Lbl  : 524281
RestProtSrcMacAct  : none
Evpn Mpls Encap    : Enabled           Evpn MplsoUdp      : Disabled
Oper Group         :
===============================================================================

===============================================================================
BGP EVPN MPLS Auto Bind Tunnel Information
===============================================================================
Allow-Flex-Algo-Fallback  : false
Resolution                : any               Strict Tnl Tag   : false
Max Ecmp Routes           : 1
Bgp Instance              : 1
Filter Tunnel Types       : (Not Specified)
===============================================================================
```

When traffic is generated, the PEs will start learning MAC addresses and advertising them in BGP so that the remote PEs learn those MAC addresses against EVPN destinations. For instance, when CE-13 sends traffic, PE-3 learns its MAC address and advertises it. The remote PEs (for instance, PE-2) will learn the MAC address and associate it with their EVPN destination to PE-3 (192.0.2.3:524280 in this example):

```
*A:PE-2# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier        Type     Last Change
            Transport:Tnl-Id                          Age
-------------------------------------------------------------------------------
1          00:00:11:11:11:11 sap:lag-1:1              L/0      02/17/21 14:16:39
1          00:00:13:13:13:13 mpls:                    Evpn     02/17/21 14:16:39
                             192.0.2.3:524280
           ldp:65537
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

When the **ingress-replication-bum-label** is enabled in the PEs, the advertisement of MAC addresses will create new EVPN destinations, because the label is different from the one previously sent by the inclusive multicast route that created an EVPN destination. In the preceding example, when PE-3 advertises the CE-13 MAC address, PE-2 will create a new binding (see in the following output in bold) that shows one MAC address that is not Mcast (multicast) capable:

```
*A:PE-2# show service id 1 evpn-mpls

===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address     Egr Label     Num. MACs   Mcast           Last Change
                Transport:Tnl                             Sup BCast Domain
-------------------------------------------------------------------------------
192.0.2.3       524278        0           bum             02/17/2021 14:17:42
                ldp:65537                                 No
192.0.2.3       524282        1           none            02/17/2021 14:17:42
                ldp:65537                                 No
```

```
192.0.2.4       524279        0           bum             02/17/2021 14:17:43
                ldp:65538                                 No
192.0.2.5       524279        0           bum             02/17/2021 14:17:45
                ldp:65539                                 No
-------------------------------------------------------------------------------
Number of entries : 4
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

When an EVPN-MPLS destination or MAC address is not created/installed correctly, the user may check the BGP-EVPN routes received and the routes kept in the RIB. The routes that the PE receives are shown when **debug router bgp update** is enabled. These routes are shown even before any BGP processing is carried out.

```
# on PE-2:
7 2021/02/17 14:01:47.146 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 81
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-MAC Len: 33 RD: 192.0.2.3:1 ESI: ESI-0, tag: 0, mac len: 48
              mac: 00:00:13:13:13:13, IP len: 0, IP: NULL, label1: 8388512
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:1
        bgp-tunnel-encap:MPLS
"
```

```
*A:PE-2# show router bgp routes evpn mac
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.3:1         00:00:13:13:13:13 ESI-0
      0                   Seq:0             LABEL 524282
                          n/a
                          192.0.2.3

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

If the route is successfully imported, it can be shown in the RIB (**show router bgp routes** commands). The route shown in the debug and the same route in a show command do not necessarily have the same label value. The reason for this expected mismatch is that the debug command shows the complete 24-bit field value because the route is shown before BGP can decide and decipher whether the label value is an MPLS label (high-order 20-bits of the label field) or a VNI (all 24 bits of the Label field for VXLAN). When the label in the debug command (8388512) is divided by 16 (2^4), the result is the MPLS label (524282), as follows: 8388512:16=524282.

VPLS to EVPN-MPLS integration

The SR OS EVPN implementation supports RFC 8560, *(PBB-)EVPN Seamless Integration with (PBB-)EVPN*, so that EVPN-MPLS and VPLS can be integrated into the same network and within the same service.

The following behavior enables the integration of EVPN and SDP-bindings in the same VPLS network:

- Systems with EVPN endpoints and SDP-bindings to the same far-end bring down the SDP-bindings.
  - SR OS will allow the establishment of an EVPN destination and an SDP-binding to the same far-end but the SDP-binding will be kept operationally down. Only the EVPN endpoint will be operationally up. This is true for spoke-SDPs (manual and BGP-AD) and mesh-SDPs. It is also true between VXLAN and SDP-bindings.
  - If there is an EVPN endpoint to a specified far-end and a spoke-SDP establishment is attempted, the spoke-SDP will be set up but kept down with an operational flag indicating that there is an EVPN route to the same far-end.
  - If there is a spoke-SDP and a valid/used EVPN route arrives, the EVPN endpoint will be set up and the spoke-SDP will be brought down with an operational flag indicating that there is an EVPN route to the same far-end.
  - In the case of an SDP-binding and EVPN endpoint to different far-end IP addresses on the same remote PE, both links will be up. This can happen if the SDP-binding is terminated in an IPv6 address or IPv4 address different from the system address where the EVPN endpoint is terminated.

  The following example illustrates the preceding description. A spoke-SDP is added to the VPLS 1 configuration on PE-2:

```
# on PE-2:
configure
    service
        sdp 24 mpls create
            far-end 192.0.2.4
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        vpls "VPLS1"
            spoke-sdp 24:1 create
                no shutdown
            exit
        exit
```

The service configuration on PE-4 is as follows:

```
# on PE-4:
configure
    service
        sdp 42 mpls create
            far-end 192.0.2.2
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        sdp 46 mpls create
            far-end 192.0.2.6
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        vpls 1 name "VPLS1" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            spoke-sdp 42:1 create
                no shutdown
            exit
            spoke-sdp 46:1 create
                no shutdown
            exit
            no shutdown
```

Spoke SDP 24:1 is operationally down, as can be verified as follows:

```
*A:PE-2# show service id 1 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
24:1             Spok     192.0.2.4       Up      Down      524282    524282
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

Spoke SDP 24:1 is down because of an EVPN route conflict, as indicated by the flags:

```
*A:PE-2# show service id 1 sdp 24 detail | match Flag context all
```

```
Flags                 : PWPeerFaultStatusBits
                        EvpnRouteConflict
```

- The user can add spoke-SDPs and all the EVPN-MPLS endpoints in the same SHG.
  - A CLI command exists in the **bgp-evpn>mpls** context so that the EVPN-MPLS endpoints can be added to an SHG.
  - The **bgp-evpn mpls split-horizon-group** must reference a user-configured SHG. User-configured SHGs can be configured within the service context.
  - The same group name can be associated with SAPs, spoke-SDPs, PW-templates, PW-template-bindings, and EVPN-MPLS endpoints.
  - If the **split-horizon-group** command in **bgp-evpn>mpls** is not used, the default SHG (in which all the EVPN endpoints are) is still used, but it will not be possible to refer to it on SAPs/spoke-SDPs.
- The system disables the advertisement of MAC addresses learned on spoke- SDPs/SAPs that are part of an EVPN SHG.
  - When the SAPs or spoke-SDPs (manual or BGP-AD-discovered) are configured within the same SHG as the EVPN endpoints, MAC addresses will still be learned on them, but will not be advertised in EVPN.
  - The preceding statement is also true if proxy-ARP/ND is enabled and an IP-->MAC address pair is learned on a SAP/SDP-binding that belongs to the EVPN SHG.
  - The SAPs and/or spoke-SDPs added to an EVPN SHG should not be part of any EVPN multi-homed ES. If that happened, the PE would still advertise the AD per-EVI route for the SAP and/or spoke-SDP, attracting EVPN traffic that could not be forwarded to that SAP and/or SDP-binding.
  - Similar to the preceding statement, an SHG composed of SAPs/SDP-bindings used in a BGP-MH site should not be configured under **bgp-evpn>mpls>split-horizon-group**. This misconfiguration would prevent traffic being forwarded from the EVPN to the BGP-MH site, regardless of the DF/Non-DF state.

An example of a shared SHG configuration on PE-2 is as follows. Because the SAP and EVPN-MPLS are in the same SHG, no MAC addresses learned over SAP 1/2/1:2 will be advertised in EVPN (not even static MACs).

```
# on PE-2:
configure
    service
        vpls 2 name "VPLS2" customer 1 create
            split-horizon-group "CORE" create
            exit
            bgp
            exit
            bgp-evpn
                evi 2
                mpls bgp 1
                    split-horizon-group "CORE"
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            sap 1/2/1:2 split-horizon-group "CORE" create
                no shutdown
            exit
            sap lag-1:2 create
                no shutdown
            exit
            no shutdown
```

EVPN-MPLS multi-homing

SR OS supports EVPN multi-homing as per RFC 7432.

The EVPN multi-homing implementation is based on the concept of the ES. An ES is a logical structure that can be defined in one or more PEs and identifies the CE (or access network) multi-homed to the EVPN PEs. An ES is associated with a port, LAG, or SDP object, and is shared by all the services defined on those objects.

Each ES has a unique identifier called ESI (Ethernet Segment Identifier) that is 10 bytes and is manually configured. The ESI is advertised in the control plane to all the PEs in an EVPN network; therefore, it is very important to ensure that the 10-byte ESI value is unique throughout the entire network. Single-homed CEs are assumed to be connected to an ES with ESI = 0 (single-homed ESs are not explicitly configured).

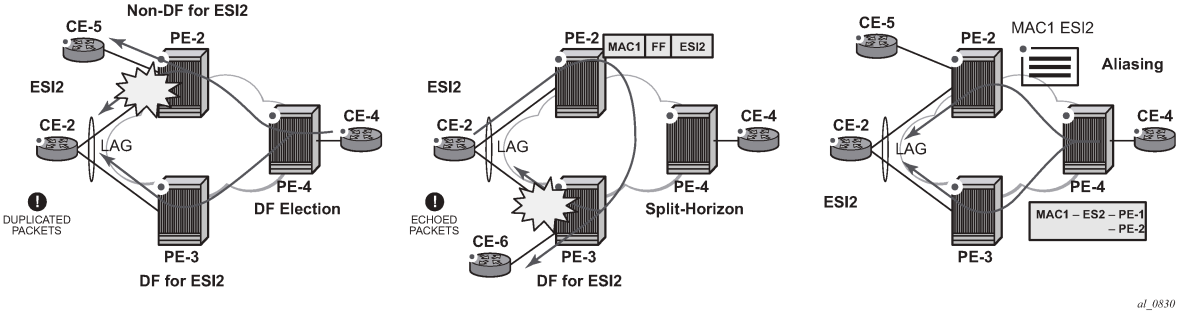
The ES is part of the base BGP-EVPN configuration and is not applied to any EVPN-MPLS service, by default. An ES can be shared by multiple services; the association of a specific SAP or spoke-SDP to an ES is automatically made when the SAP is defined in the same LAG or port configured in the ES, or when the spoke-SDP is defined in the same SDP configured in the ES. The following sections show the configuration of:

- an all-active multi-homing ES with a LAG associated with it
- a single-active multi-homing ES linked to an SDP

All-active multi-homing concepts

EVPN all-active multi-homing is built around three concepts: DF election, split-horizon (with an ESI-label), and aliasing, as shown in Figure 63: EVPN-MPLS all-active multi-homing concepts, from left to right.

Figure 63: EVPN-MPLS all-active multi-homing concepts



- With DF election, when PE-4 sends BUM traffic to the remote ES (CE-2), only one PE segment sends the BUM packets to the ES (PE-3 is the DF in the preceding example, and is elected to send BUM packets to CE-2). The non-DF, PE-2, removes the LAG SAP from the default multicast list (PE-2 does not bring CE-2 down, because it still needs to send upstream/downstream unicast traffic). PE-2 and PE-3 elect a DF for each service, based on the ES routes and the service-carving algorithm.

- With split-horizon, the PE part of the ES (PE-3 in the preceding example) identifies the BUM packets coming from the PE for the remote (PE-2), but within the same ES (ESI-2), and filters the packets so that they are not sent back to the ES, creating duplication. When PE-2 (non-DF) sends BUM traffic to PE-3 (DF), it uses a special MPLS label in the data path that PE-3 previously advertised for ESI-2 in an AD per-ES route. When PE-3 does an ingress lookup, it recognizes the ESI-label and filters the traffic (PE-3 still sends the BUM traffic to other SAPs/SDP-bindings).
- With aliasing, remote PEs that are not part of the ES can load-balance unicast traffic to all the PEs that are part of the ES, irrespective of from which PE a destination MAC address was learned. PE-4 will create an EVPN destination to ESI-2 that will be resolved to the two next-hops: PE-2 and PE-3. Unicast load-balancing will happen as long as ECMP > 1 is enabled in PE-4.

Nokia recommends the use of **ingress-replication-bum-label** on the PEs that are part of an all-active ES. In an all-active multi-homing scenario, if a specified MAC address (for example, the CE-2 MAC address in the left-hand-side diagram), is not learned yet in a remote PE (for example, PE-4), but is known in the two PEs of the ES (for example, PE-2 and PE-3), the latter PEs might send duplicated packets to the CE.

This issue is solved by the use of **ingress-replication-bum-label** in PE-2 and PE-3. If configured, PE-2/PE-3 will know that the received packet is an unknown unicast packet; therefore, the Non-DF (PE-2) will not send the packet to CE-2 and there will not be duplication.

All-active multi-homing configuration

The all-active multi-homing configuration example is based on Figure 62: EVPN-MPLS for VPLS services.

MTU-1 is connected to the EVPN network using all-active multi-homing. According to RFC 7432, MTU-1 will be able to send traffic to both PEs for VPLS-1. Regular LAG load-balancing is used in MTU-1. Remote PEs such as PE-4 or PE-5 will be able to load-balance the unicast traffic to PE-2 and PE-3. PE-2 and PE-3 will discover that both are part of ESI-23 (due to the exchange of ES routes) and will elect a DF for VPLS-1. The non-DF for VPLS-1, in this case PE-2, will remove lag-1:1 from the VPLS-1 default multicast list. Also, when PE-2 and PE-3 send BUM traffic to each other, they will insert an ESI-label so that they can identify that the source of the BUM packet is ESI-23.

The following output shows the configuration of ESI-23 in PE-2 and PE-3, as well as the LAG interfaces for all-active multi-homing (see Figure 62: EVPN-MPLS for VPLS services). The configuration of LAG-1 in MTU-1 is also shown. Per RFC 7432, only a CE/MTU with a LAG can be connected to an all-active multi-homing ES. No other configuration is permitted on the CE for all-active multi-homing.

LAG 1 is configured on MTU-1, PE-2, PE-3, as follows:

```
# on MTU-1:
configure
    lag 1
        mode access
        encap-type dot1q
        port 1/1/1
        port 1/1/2
        lacp active administrative-key 32768
        no shutdown
```

```
# on PE-2:
configure
    lag 1
        mode access
        encap-type dot1q
        port 1/1/2
        lacp active administrative-key 1 system-id 00:00:00:00:02:03
        no shutdown
```

```
# on PE-3:
configure
    lag 1
        mode access
        encap-type dot1q
        port 1/1/1
        lacp active administrative-key 1 system-id 00:00:00:00:02:03
        no shutdown
```

Ethernet segment “ESI-23” is configured in the service **system bgp-evpn** context on PE-2 and PE-3, as follows:

```
# on PE-2, PE-3:
configure
    service
        system
            bgp-evpn
                ethernet-segment "ESI-23" create
                    esi 01:00:00:00:00:23:00:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 1
                    no shutdown
                exit
```

When configuring an ES, the following must be considered:

- Any EVPN parameter that is not specific to any particular VPLS service, and is common to all the EVIs, is configured in a base BGP-EVPN instance located at **config>service>system>bgp-evpn**. In this base instance, the following attributes may be configured:
  - ethernet-segments
  - the base BGP-EVPN instance **route-distinguisher** that will be used for the ES routes. If this **route-distinguisher** is not configured, by default a type-1 RD will be derived as system-ip:0, as shown in the command help:

  ```
  *A:PE-2>config>service>system>bgp-evpn# route-distinguisher ?
    - no route-distinguisher
    - route-distinguisher <rd>

   <rd>                 : <ip-addr:comm-val>
                          ip-addr        - a.b.c.d
                          comm-val       - [0..65535]
                          default: system-ip:0
  ```

- The ES must be configured with a name and can contain the following parameters when configured for all-active multi-homing:
  - **esi** — 10-byte identifier that represents the ES in the BGP control plane. The same ESI must be configured in all the PEs connected to the same CE/MTU (using a unique value that cannot be associated with any other CE/MTU/access network). RFC 7432 defines five different types of ESI. In SR OS, the **type** byte, as well as the other 9 bytes can be arbitrarily configured.
  - **multi-homing all-active** — This command indicates that the ES is in all-active mode.

  - **lag** <*lag-id*> — The LAG connected to the CE/MTU must be added to the ES. In this example, lag-1 is added to ESI-23, on both PE-2 and PE-3. Although a different LAG-id may have been assigned to the same ES on PE-2 and PE-3, PE-2 and PE-3 must have the same configuration on the ES LAG; that is, encap-type. Also, if LACP is added (it is not mandatory), both PEs must have the same admin-key, system-id, and system-priority. MTU-1 will see PE-2 and PE-3 as a single LAG peer. For all-active multi-homing, only the **lag** option is accepted by the system; **port** or **sdp** are not accepted.
  - **[no] shutdown** — This command controls the administrative state of the ES.
- The preceding parameters are the minimum necessary so that the ES can be activated. In addition to those parameters, there are a few more that the user can configure if requiring values different from the default ones:
  - **es-activation-timer** *[0..100]* can be configured at **redundancy>bgp-evpn-multi-homing>es-activation-timer** or at **service>system>bgp-evpn>eth-seg>es-activation-timer** level (the most specific value is used).

    The **es-activation-timer** operation is as follows:

    - Upon reception of an ES, AD per-ES/EVI route update/withdrawal for a local ESI, the DF-candidate list of IPs is updated and the DF election algorithm is run without waiting for any timer.
    - If the result of the DF election requires the PE to be promoted from non-DF to DF, the **es-activation-timer** will start, and only after its expiration will the PE add the SAP to the default-multicast list. Transitions from non-DF to non-DF, or from DF to non-DF, are immediate and do not wait for any timer.
    - This use of an **es-activation-timer** value minimizes the risks of loops and packet duplication due to **transient** multiple DFs.
    - The same **es-activation-timer** must be configured in all the PEs that are part of the same ESI. The user must configure either a long timer to minimize the risks of loops/duplication, or **es-activation-timer** = 0 to speed up the convergence for NDF to DF transitions. The default value is 3 seconds.
  - **service-carving** — As defined in RFC 7432, service-carving controls the distribution of DF/non-DF roles across the different services defined in an ES.

    ```
    *A:PE-2>config>service>system>bgp-evpn>eth-seg>service-carving# mode ?
     - mode {auto|manual|off}

     <auto|manual|off>    : auto|manual|off
    ```

    ```
    *A:PE-2>config>service>system>bgp-evpn>eth-seg>service-carving# manual ?
      - manual

     [no] evi             - Configure EVI range (primary for non-preference based DF
                            election and lowest-preference for preference based DF
                            election)
     [no] isid            - Configure ISID range (primary for non-preference based DF
                            election and lowest-preference for preference based DF
                            election)
     [no] preference      + Configure DF preference election information
    ```

    As shown above, **service-carving** has three different modes:
  - **service-carving mode auto** (default) — The DF election algorithm will run the function [V(evi) mod N(peers) = i(ordinal)] to know who the DF for a specified service and ESI is. In this example, ESI-23 is configured with mode **auto**; therefore, for VPLS-1 (with EVI-1), PE-3 will be elected as DF

because evi(1) mod (2)peers = 1, and the ordinal 1 corresponds to the second lowest IP, PE-3. The algorithm takes the configured **evi** in the service; therefore, the **evi** is mandatory, and for the same service must match in all the PEs that are part of the ES. This guarantees that the election algorithm is consistent across all the PEs of the ESI.

- **service-carving mode manual** — The user can manually decide for which **evi** identifiers the PE is DF: **service-carving mode manual / manual evi** <*start*> **[to** <*to*>**]**. The PE will be non-DF for the non-specified EVIs. If **service-carving mode manual** is configured, but no range is defined, all the services are considered to be non-DF. If a range is configured, but the **service-carving** is not **mode manual**, the range has no effect. Only two PEs are supported when **service-carving mode manual** is configured.
- **service-carving mode off** — The lowest originator IP will win the election for a specified service and ES.
- Because the **evi** is used for the service-carving algorithm, it must always be configured in a service with SAPs/SDP bindings created in an ES, regardless of the service-carving mode (service-carving off, auto, or manual).

Although not configured as part of the ES, the **config>redundancy>bgp-evpn-multi-homing>boot-timer** allows the necessary time for the control plane protocols to come up after the PE has rebooted, and before bringing up the ESs and running the DF algorithm. Some considerations about the boot timer:

- The boot timer should use a value long enough to allow the IOMs and BGP sessions to come up before exchanging ES routes and run the DF election for each EVI (it is 10 s, by default).
- The boot timer runs per EVI on the ESs in the system. While **system-up-time** <*boot-timer*>, the system will not run the DF election for any EVI. When the boot timer expires, the DF election for the EVI is run and, if the system is elected DF for the EVI, the **es-activation-timer** will start.
- The system will not advertise ES routes until the boot timer expires. This guarantees that the peer ES PEs do not run the DF election either, until the PE is ready to become the DF, if needed.
- The following show command displays the configured boot timer, as well as the remaining timer if the system is still in boot stage.

```
*A:PE-2# show redundancy bgp-evpn-multi-homing

===============================================================================
Redundancy BGP EVPN Multi-homing Information
===============================================================================
Boot-Timer              : 10 secs
Boot-Timer Remaining    : 0 secs
ES Activation Timer     : 3 secs
===============================================================================
```

After ESI-23 is configured in PE-2 and PE-3, the lag-1 SAPs in both PEs can be added to the VPLS-1 service. Until the ESI-23 is successfully enabled, the LAG SAPs will be kept down with a *StandByForMHProtocol* flag. This is illustrated in the following example for PE-2.

```
# on PE-2:
configure
    service
        system
            bgp-evpn
                ethernet-segment "ESI-23"
                    shutdown
                exit
            exit
        exit
    exit
    service
        vpls "VPLS1"
            sap lag-1:1 create
                no shutdown
            exit
```

```
*A:PE-2# show service id 1 sap lag-1:1 detail | match "  Oper State"
Admin State        : Up                       Oper State        : Down
```

```
*A:PE-2# show service id 1 sap lag-1:1 detail | match Flag
Flags              : StandByForMHProtocol
```

```
# on PE-2:
configure
    service
        system
            bgp-evpn
                ethernet-segment "ESI-23"
                    no shutdown
                exit
```

```
*A:PE-2# show log log-id 99

===============================================================================
Event Log 99 log-name 99
===============================================================================
Description : Default System Log
Memory Log contents  [size=500   next event=107  (not wrapped)]

106 2021/02/17 14:38:48.980 UTC MINOR: SVCMGR #2203 Base
"Status of SAP lag-1:1 in service 1 (customer 1) changed to admin=up oper=up flags="
```

All-active multi-homing operation

To confirm that all-active multi-homing is working correctly for ESI-23, the user can use the following commands:

- **show service system bgp-evpn** — Shows the RD is used for the ES route.
- **show service system bgp-evpn ethernet-segment** — Shows all the ESs configured in the PE and their admin/operational status.
- **show service system bgp-evpn ethernet-segment name ESI-23 evi 1** — Shows the DF candidate PEs for EVI 1 and whether the system is DF for EVI.
- **show service system bgp-evpn ethernet-segment name ESI-23 all** — Shows all the information related to a specific ESI.

The base BGP-EVPN information includes the RD:

```
*A:PE-2# show service system bgp-evpn

===============================================================================
System BGP EVPN Information
===============================================================================
Eth Seg Route Dist.               : <none>
Eth Seg Oper Route Dist.          : 192.0.2.2:0
Eth Seg Oper Route Dist Type      : default
Ad Per ES Route Target            : evi-rt
Leaf Label                        : 0
Mcast Leave Sync Prop             : 5
Attribute Uniform Prop            : Disabled
BGP Path Selection                : Disabled
===============================================================================
```

The following command shows the configured ESs in the PE and their status:

```
*A:PE-2# show service system bgp-evpn ethernet-segment

===============================================================================
Service Ethernet Segment
===============================================================================
Name                             ESI                          Admin     Oper
-------------------------------------------------------------------------------
ESI-23                           01:00:00:00:00:23:00:00:00:01 Enabled   Up
-------------------------------------------------------------------------------
Entries found: 1
===============================================================================
```

The following command shows that PE-2 is not the DF and the DF candidate PEs for EVI 1 are PE-2 and PE-3:

```
*A:PE-2# show service system bgp-evpn ethernet-segment name "ESI-23" evi 1

===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
1             1             0                   no  02/17/2021 14:38:49
===============================================================================

===============================================================================
DF Candidates                          Time Added
-------------------------------------------------------------------------------
192.0.2.2                              02/17/2021 14:38:49
192.0.2.3                              02/17/2021 14:38:51
-------------------------------------------------------------------------------
Number of entries: 2
===============================================================================
```

The following command shows all information related to ESI-23 on PE-2:

```
*A:PE-2# show service system bgp-evpn ethernet-segment name "ESI-23" all

===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : ESI-23
Eth Seg Type            : None
Admin State             : Enabled           Oper State         : Up
ESI                     : 01:00:00:00:00:23:00:00:00:01
Multi-homing            : allActive         Oper Multi-homing  : allActive
ES SHG Label            : 524279
Source BMAC LSB         : <none>
Lag Id                  : 1
ES Activation Timer     : 3 secs
Oper Group              : (Not Specified)
```

```
Svc Carving             : auto              Oper Svc Carving   : auto
Cfg Range Type          : primary
===============================================================================

===============================================================================
EVI Information
===============================================================================
EVI                 SvcId               Actv Timer Rem      DF
-------------------------------------------------------------------------------
1                   1                   0                   no
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================

-------------------------------------------------------------------------------
DF Candidate list
-------------------------------------------------------------------------------
EVI                                     DF Address
-------------------------------------------------------------------------------
1                                       192.0.2.2
1                                       192.0.2.3
-------------------------------------------------------------------------------
Number of entries: 2
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
---snip---
```

The following command shows all information related to ESI-23 on PE-3:

```
*A:PE-3# show service system bgp-evpn ethernet-segment name "ESI-23" all

===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : ESI-23
Eth Seg Type            : None
Admin State             : Enabled           Oper State         : Up
ESI                     : 01:00:00:00:00:23:00:00:00:01
Multi-homing            : allActive         Oper Multi-homing  : allActive
ES SHG Label            : 524280
Source BMAC LSB         : <none>
Lag Id                  : 1
ES Activation Timer     : 3 secs
Oper Group              : (Not Specified)
Svc Carving             : auto              Oper Svc Carving   : auto
Cfg Range Type          : primary
===============================================================================
===============================================================================
EVI Information
===============================================================================
EVI                 SvcId               Actv Timer Rem      DF
-------------------------------------------------------------------------------
1                   1                   0                   yes
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
-------------------------------------------------------------------------------
DF Candidate list
-------------------------------------------------------------------------------
EVI                                     DF Address
-------------------------------------------------------------------------------
1                                       192.0.2.2
1                                       192.0.2.3
```

```
-------------------------------------------------------------------------------
Number of entries: 2
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
---snip---
```

The preceding commands show the ESI-23 configuration on both PEs and the result of the DF election for EVI 1.

The following output shows the ES route received on PE-2:

```
# on PE-2:
129 2021/02/17 14:31:29.996 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 71
    Flag: 0x90 Type: 14 Len: 34 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-ETH-SEG Len: 23 RD: 192.0.2.3:0 ESI: 01:00:00:00:00:23:00:00:00:01,
                          IP-Len: 4 Orig-IP-Addr: 192.0.2.3
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        df-election::DF-Type:Auto/DP:0/DF-Preference:0/AC:1
        target:00:00:00:00:23:00
"
```

The ES RT as shown as target:00:00:00:00:23:00 in the extended community is auto-derived from the ESI bytes 2 to 7 (with the type byte being byte 1). Only PE-2 and PE-3 generate this RT and therefore import each other’s ES route.

The following message in log 99 on PE-3 shows the result of the DF election for EVI 1.

```
86 2021/02/17 14:31:34.395 UTC MINOR: SVCMGR #2094 Base
"Ethernet Segment:ESI-23, EVI:1, Designated Forwarding state changed to:true"
```

The **show service system bgp-evpn ethernet-segment name ESI-23 all** command shows the ESI-label allocated to the PE: **ES SHG Label** *524280* in the CLI output for PE-3. In this example, this label is allocated by PE-3 for ESI-23 (a different one is allocated per ESI) and advertised in the AD per-ES route for ESI-23. The following output shows the AD per-ES and AD per-EVI (for evi 1) routes sent by PE-3 and received by PE-2.

- The AD per-ES route can be identified by the *MAX-ET* in the Ethernet-tag field (as per RFC 7432) and carries the ESI-label as well as the multi-homing mode (all-active in this case) in the ESI-label extended community (see Figure 61: EVPN route types and NLRIs).

  The user can enable the aggregation of AD per-ES routes by using the following command: **config service system bgp-evpn ad-per-es-route-target evi-rt-set route-distinguisher ip-address**. If enabled, a single AD per-ES route with the associated RD and a set of EVI route-targets will be advertised (to a maximum of 128). When there are more than 128 EVIs defined in the Ethernet-segment, more than one route will be sent by the system.
- The AD per-EVI route will have an eth-tag 0 and will carry the service label in the NLRI.

```
# AD per-ES route received on PE-2:
156 2021/02/24 08:39:37.907 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
```

```
Peer 1: 192.0.2.3 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 73
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-AD Len: 25 RD: 192.0.2.3:1 ESI: 01:00:00:00:00:23:00:00:00:01,
                      tag: MAX-ET Label: 0
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:1
        esi-label:524282/All-Active
"
```

```
# AD per-EVI route received on PE-2:
155 2021/02/24 08:39:37.906 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 73
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-AD Len: 25 RD: 192.0.2.3:1 ESI: 01:00:00:00:00:23:00:00:00:01,
                      tag: 0 Label: 8388464
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:1
        bgp-tunnel-encap:MPLS
"
```

```
*A:PE-2# show router bgp routes evpn auto-disc esi 01:00:00:00:00:23:00:00:00:01
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Flag  Route Dist.         ESI                           NextHop
      Tag                                               Label
-------------------------------------------------------------------------------
u*>i  192.0.2.3:1         01:00:00:00:00:23:00:00:00:01 192.0.2.3
      0                                                 LABEL 524282

u*>i  192.0.2.3:1         01:00:00:00:00:23:00:00:00:01 192.0.2.3
      MAX-ET                                            LABEL 0

-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

```
*A:PE-2# show router bgp routes evpn auto-disc esi 01:00:00:00:00:23:00:00:00:01 hunt
---snip---
```

```
===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : n/a
Nexthop        : 192.0.2.3
From           : 192.0.2.3
Res. Nexthop   : 192.168.23.2
---snip---
Community      : target:64500:1 bgp-tunnel-encap:MPLS
---snip---
EVPN type      : AUTO-DISC
ESI            : 01:00:00:00:00:23:00:00:00:01
Tag            : 0
Route Dist.    : 192.0.2.3:1
MPLS Label     : LABEL 524282

---snip---

Network        : n/a
Nexthop        : 192.0.2.3
From           : 192.0.2.3
Res. Nexthop   : 192.168.23.2
---snip---
Community      : target:64500:1 esi-label:524280/All-Active
---snip---
EVPN type      : AUTO-DISC
ESI            : 01:00:00:00:00:23:00:00:00:01
Tag            : MAX-ET
Route Dist.    : 192.0.2.3:1
MPLS Label     : LABEL 0
---snip---
```

From a service perspective, as soon as CE-11 sends some traffic, the PE learning the CE-11 MAC address will advertise it to the network. The remote PEs (PE-4 and PE-5) will create a new EVPN-MPLS ES destination to ESI-23, with two next-hops: PE-2 and PE-3. The following outputs on PE-4 show the following information:

- PE-4 has learned AD per-EVI/ES routes for ESI-23 from PE-2 and PE-3, as well as the CE-11 MAC address from PE-3 (because MTU-1 picked up its link to PE-3 to send CE-11 frames).

```

*A:PE-4# show router bgp routes evpn auto-disc esi 01:00:00:00:00:23:00:00:00:01
===============================================================================
 BGP Router ID:192.0.2.4        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Flag  Route Dist.         ESI                           NextHop
      Tag                                               Label
-------------------------------------------------------------------------------
u*>i  192.0.2.2:1         01:00:00:00:00:23:00:00:00:01 192.0.2.2
      0                                                 LABEL 524281

u*>i  192.0.2.2:1         01:00:00:00:00:23:00:00:00:01 192.0.2.2
```

```
      MAX-ET                                            LABEL 0

u*>i  192.0.2.3:1         01:00:00:00:00:23:00:00:00:01 192.0.2.3
      0                                                 LABEL 524282

u*>i  192.0.2.3:1         01:00:00:00:00:23:00:00:00:01 192.0.2.3
      MAX-ET                                            LABEL 0

-------------------------------------------------------------------------------
Routes : 4
===============================================================================
```

PE-4 has learned MAC address 00:00:11:11:11:11 of CE-11 in ESI-23. The BGP EVPN MAC route has PE-3 as next hop:

```
*A:PE-4# show router bgp routes evpn mac rd 192.0.2.3:1
===============================================================================
 BGP Router ID:192.0.2.4        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.3:1         00:00:11:11:11:11 01:00:00:00:00:23:00:00:00:01
      0                   Seq:0             LABEL 524282
                          n/a
                          192.0.2.3

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

- In the FDB for VPLS-1, PE-4 has learned the CE-11 MAC address associated with a newly created EVPN-MPLS ES destination:

```

*A:PE-4# show service id 1 fdb mac 00:00:11:11:11:11

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:11:11:11:11 eES:                    Evpn     02/17/21 14:54:42
                             01:00:00:00:00:23:00:00:00:01
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

- Due to the aliasing function, the newly created EVPN-MPLS ES destination to ESI-23 has two next-hops (PE-2 and PE-3), to which PE-4 can load-balance the unicast traffic because **ecmp 2** is configured in the VPLS-1 of PE-4.

```
*A:PE-4# show service id 1 evpn-mpls

===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address     Egr Label     Num. MACs   Mcast           Last Change
                Transport:Tnl                             Sup BCast Domain
-------------------------------------------------------------------------------
192.0.2.2       524280        0           bum             02/17/2021 14:17:45
                ldp:65538                                 No
192.0.2.3       524278        0           bum             02/17/2021 14:17:45
                ldp:65537                                 No
192.0.2.5       524279        0           bum             02/17/2021 14:17:45
                ldp:65539                                 No
-------------------------------------------------------------------------------
Number of entries : 3
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                       Num. Macs              Last Change
-------------------------------------------------------------------------------
01:00:00:00:00:23:00:00:00:01   1                      02/17/2021 14:54:42
-------------------------------------------------------------------------------
Number of entries: 1
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

The **show service id 1 evpn-mpls esi 01:00:00:00:00:23:00:00:00:01** command shows the next-hops that the EVPN-MPLS ES destination is resolved to.

```
*A:PE-4# show service id 1 evpn-mpls esi 01:00:00:00:00:23:00:00:00:01

===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                       Num. Macs              Last Change
-------------------------------------------------------------------------------
01:00:00:00:00:23:00:00:00:01   1                      02/17/2021 14:54:42
===============================================================================

===============================================================================
BGP EVPN-MPLS Dest TEP Info
===============================================================================
TEP Address              Egr Label                Last Change
                         Transport:Tnl-Id
-------------------------------------------------------------------------------
192.0.2.2                524281                   02/17/2021 14:54:42
                         ldp:65538
192.0.2.3                524282                   02/17/2021 14:54:42
                         ldp:65537
-------------------------------------------------------------------------------
Number of entries : 2
```

```
-------------------------------------------------------------------------------
===============================================================================
```

- PE-2 will show the CE-11 MAC address as learned locally in SAP lag-1:1 (because the data plane learning of the CE-11 MAC address happened in PE-2). For PE-3, even though it learned the MAC address from EVPN, it will install it as associated with SAP lag-1:1 because the EVPN route came with ESI-23, which is a local ESI. Because of this, whenever PE-3 receives a frame with MAC DA equal to the CE-11 MAC address, it will be able to forward the frame locally to the SAP lag-1:1. The following output shows the CE-11 MAC address as it is installed in PE-2 and PE-3:

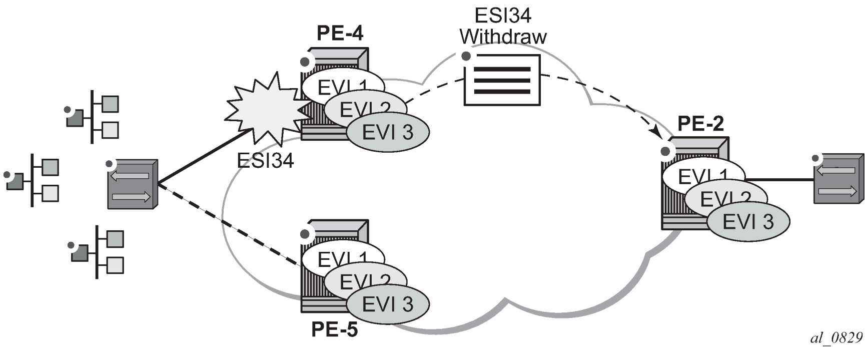
```
*A:PE-2# show service id 1 fdb mac 00:00:11:11:11:11

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:11:11:11:11 sap:lag-1:1             L/0      02/17/21 15:02:44
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

```
*A:PE-3# show service id 1 fdb mac 00:00:11:11:11:11

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:11:11:11:11 sap:lag-1:1             Evpn     02/17/21 15:02:44
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

Single-active multi-homing concepts

Figure 64: EVPN-MPLS single-active multi-homing: mass-withdraw, backup path illustrates two concepts in EVPN single-active multi-homing: mass-withdraw and backup path.

Figure 64: EVPN-MPLS single-active multi-homing: mass-withdraw, backup path



- With mass-withdraw, when ESI-45 goes down, PE-2 does not have to wait for all the MAC routes to be withdrawn to converge all the services. Instead, PE-4 will withdraw the AD per-ES routes (also the AD per-EVI and MAC routes) and that will be used at PE-2 as a notification to stop sending traffic to PE-4 for any MAC address associated with ESI-45.
- With backup path, when PE-2 is notified of the ESI-45 failure due to the withdrawn AD routes, it will not flush any MAC address associated with ESI-45. Instead, it will change the next-hop of the EVPN-MPLS ES destination to the remaining PE in the ESI-45. Backup path only works when there are two PEs in the same ES. If there were more than two PEs in ESI-45, PE-2 would flush all the MAC addresses upon receiving a mass-withdraw notification, because it would not know who the new active PE is.

Single-active multi-homing configuration

The single-active multi-homing configuration example is based on Figure 62: EVPN-MPLS for VPLS services:

MTU-6 is connected to the EVPN network using single-active multi-homing. With the MTU-6 configuration, a VPLS service with active-standby spoke-sdp to PE-4 and PE-5 is configured. In PE-4 and PE-5, the SDP connected to MTU-6 is linked to ESI-45. Both will run the DF election algorithm for EVI 1, and the non-DF PE (PE-4 in this example) will bring down the spoke-SDP and notify MTU-6.

The following output shows the configuration of ESI-45 in PE-4 and PE-5, as well as the SDPs. The configuration of MTU-6 is also shown for completeness. It is important to keep the default **no ignore-standby-signaling** command on MTU-6 spoke-SDPs because the PW switchover in MTU-6 will be triggered based on the PW status bits sent by PE-4 and PE-5.

SDP 46 with far-end MTU-6 is configured on PE-4:

```
# on PE-4:
configure
    service
        sdp 46 mpls create
            far-end 192.0.2.6
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
```

Ethernet segment “ESI-45” is configured on PE-4 as follows:

```
# on PE-4:
configure
    service
        system
            bgp-evpn
                ethernet-segment "ESI-45" create
                    esi 01:00:00:00:00:45:00:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 46
                    no shutdown
                exit
```

On PE-5, SDP 56 is configured as follows:

```
# on PE-5:
configure
    service
        sdp 56 mpls create
            far-end 192.0.2.6
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
```

Ethernet segment "ESI-45" is configured as follows on PE-5:

```
# on PE-5:
configure
    service
        system
            bgp-evpn
                ethernet-segment "ESI-45" create
                    esi 01:00:00:00:00:45:00:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 56
                    no shutdown
                exit
```

On MTU-6, the service configuration is as follows:

```
# on MTU-6:
configure
    service
        sdp 64 mpls create
            far-end 192.0.2.4
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        sdp 65 mpls create
            far-end 192.0.2.5
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        vpls 1 name "VPLS1" customer 1 create
            endpoint "CORE" create
            exit
            stp
                shutdown
            exit
            sap 1/2/1:1 create
                no shutdown
            exit
            spoke-sdp 64:1 endpoint "CORE" create
                stp
                    shutdown
                exit
                no shutdown
            exit
            spoke-sdp 65:1 endpoint "CORE" create
                stp
                    shutdown
                exit
                no shutdown
            exit
            no shutdown
        exit
```

For a detailed description of the base BGP-EVPN instance and ES configuration, see the All-active multi-homing configuration section. The **es-activation-timer**, **esi**, **service-carving**, **boot-timer**, and **shutdown** commands are used in the same way as for all-active multi-homing. Only the differences compared to all-active multi-homing are described here:

- **multi-homing single-active** must be configured so that the ES acts as single-active. Optionally, the **no-esi-label** attribute can be added to the **multi-homing single-active** command. This attribute controls the use of the ESI-label for single-active multi-homing. Although the ESI-label is always used in all-active multi-homing when sending BUM traffic between the PEs in the ES, it is configurable for single-active. However, Nokia recommends to use the default option (using ESI-label) to avoid potential transient issues when there is a DF switchover.
- **sdp** <*sdp-id*> is configured so that the ES can be associated with the SDP connected to MTU-6. Although all-active multi-homing only allows LAG associations to the ES, single-active allows LAG, port, and SDP. In this example, SDP is the option, because the access network is MPLS-based.

Similar to the all-active multi-homing case, when configuring the service in PE-4 and PE-5, the service objects are automatically associated with the ESI-45, because they are defined in the SDPs linked to the ESI. The configuration for VPLS 1 on PE-5 is as follows:

```
# on PE-5:
configure
    service
        vpls 1 name "VPLS1" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            spoke-sdp 56:1 create
                no shutdown
            exit
            no shutdown
```

In all-active multi-homing, the non-DF does not bring down the service SAP associated with the ES (it only removes it from the default-multicast-list). However, in single-active multi-homing, the service spoke-SDP (or SAP, if that was the object associated) is brought operationally down. The following output shows the spoke-SDP state in PE-4 (non-DF), as operationally down with the *StandbyForMHProtocol* flag and the **Local Pw Bits** that are signaled to MTU-6:

```
*A:PE-4# show service system bgp-evpn ethernet-segment name "ESI-45" evi 1

===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
1             1             0                   no  02/17/2021 15:04:43
===============================================================================

===============================================================================
DF Candidates                           Time Added
-------------------------------------------------------------------------------
192.0.2.4                               02/17/2021 15:04:28
192.0.2.5                               02/17/2021 15:04:43
-------------------------------------------------------------------------------
Number of entries: 2
===============================================================================
```

Spoke-SDP 46:1 is operationally down on PE-4:

```
*A:PE-4# show service id 1 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
46:1             Spok     192.0.2.6       Up      Down      524281    524281
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

Spoke-SDP 46:1 is operationally down with the StandbyForMHProtocol flag:

```
*A:PE-4# show service id 1 sdp 46:1 detail | match Flag
Flags              : StandbyForMHProtocol
```

The local PW bits (*pwFwdingStandby*) are sent to MTU-6:

```
*A:PE-4# show service id 1 sdp 46:1 detail | match Pw
Local Pw Bits      : pwFwdingStandby
Peer Pw Bits       : None
```

Single-active multi-homing operation

The same commands used in the All-active multi-homing operation section can be used for single-active; see that section.

The **show service system bgp-evpn ethernet-segment name "ESI-45"** command shows an Ethernet-segment **Oper Multi-homing** in addition to the configured **Multi-homing** mode. This occurs because, in spite of configuring the ES as all-active, it may operate as single-active if there is a mismatch between the modes advertised by PE-4 and PE-5 in the AD per-ES routes (per RFC 7432). In this example, the configured and the operational value are the same:

```
*A:PE-4# show service system bgp-evpn ethernet-segment name "ESI-45"

===============================================================================
Service Ethernet Segment
===============================================================================
Name                     : ESI-45
Eth Seg Type             : None
Admin State              : Enabled            Oper State         : Up
ESI                      : 01:00:00:00:00:45:00:00:00:01
Multi-homing             : singleActive       Oper Multi-homing  : singleActive
ES SHG Label             : 524278
Source BMAC LSB          : <none>
Sdp Id                   : 46
ES Activation Timer      : 3 secs
Oper Group               : (Not Specified)
Svc Carving              : auto               Oper Svc Carving   : auto
Cfg Range Type           : primary
===============================================================================
```

As soon as CE-16 sends some traffic, the DF PE (PE-5) will learn the CE-16 MAC address and will advertise it to the network. The remote PEs (PE-2 and PE-3) will create a new EVPN-MPLS ES destination to ESI-45, but this time with only one next-hop, PE-5, because this is single-active multi-homing. The following outputs show the following information:

- PE-2 has learned AD per-EVI/ES routes for ESI-45 from PE-4 and PE-5, as well as the CE-16 MAC address from an ES EVPN-MPLS destination, which is resolved to PE-5 (the DF for ESI-45).

```

*A:PE-2# show router bgp routes evpn auto-disc esi 01:00:00:00:00:45:00:00:00:01
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Flag  Route Dist.         ESI                           NextHop
      Tag                                               Label
-------------------------------------------------------------------------------
u*>i  192.0.2.4:1         01:00:00:00:00:45:00:00:00:01 192.0.2.4
      0                                                 LABEL 524280

u*>i  192.0.2.4:1         01:00:00:00:00:45:00:00:00:01 192.0.2.4
      MAX-ET                                            LABEL 0

u*>i  192.0.2.5:1         01:00:00:00:00:45:00:00:00:01 192.0.2.5
      0                                                 LABEL 524280

u*>i  192.0.2.5:1         01:00:00:00:00:45:00:00:00:01 192.0.2.5
      MAX-ET                                            LABEL 0

-------------------------------------------------------------------------------
Routes : 4
===============================================================================
```

PE-2 has learned the CE-16 MAC address from an ES EVPN-MPLS destination:

```
*A:PE-2# show service id 1 fdb mac 00:00:16:16:16:16

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:16:16:16:16 eES:                    Evpn     02/17/21 15:10:23
                             01:00:00:00:00:45:00:00:00:01
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

On PE-2, the ES EVPN-MPLS destination is resolved to DF PE-5:

```
*A:PE-2# show service id 1 evpn-mpls esi 01:00:00:00:00:45:00:00:00:01

===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                       Num. Macs               Last Change
-------------------------------------------------------------------------------
01:00:00:00:00:45:00:00:00:01   1                       02/17/2021 15:10:23
===============================================================================

===============================================================================
BGP EVPN-MPLS Dest TEP Info
===============================================================================
TEP Address              Egr Label                Last Change
                         Transport:Tnl-Id
-------------------------------------------------------------------------------
192.0.2.5                524280                   02/17/2021 15:10:23
                         ldp:65539
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

- In this case, the local PEs, PE-4 and PE-5, will learn the CE MAC address from an EVPN-MPLS destination and a local spoke-SDP, respectively.

```
*A:PE-4# show service id 1 fdb mac 00:00:16:16:16:16

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:16:16:16:16 eES:                    Evpn     02/17/21 15:10:23
                             01:00:00:00:00:45:00:00:00:01
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The ES EVPN-MPLS destination is resolved to DF PE-5:

```
*A:PE-4# show service id 1 evpn-mpls esi 01:00:00:00:00:45:00:00:00:01

===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                       Num. Macs               Last Change
-------------------------------------------------------------------------------
01:00:00:00:00:45:00:00:00:01   1                       02/17/2021 15:10:23
===============================================================================

===============================================================================
BGP EVPN-MPLS Dest TEP Info
===============================================================================
TEP Address              Egr Label                Last Change
                         Transport:Tnl-Id
-------------------------------------------------------------------------------
192.0.2.5                524280                   02/17/2021 15:10:23
                         ldp:65539
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

DF PE-5 learns the CE-16 MAC address from a local spoke SDP:

```
*A:PE-5# show service id 1 fdb mac 00:00:16:16:16:16

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:16:16:16:16 sdp:56:1                L/210    02/17/21 15:10:23
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

Ethernet segment failures

If either ES fails, a DF re-election will happen and the corresponding AD per-ES/EVI routes will be withdrawn, causing the remote PEs to modify the list of next-hops for the EVPN-MPLS ES destination. The following example illustrates a failure on the SDP between MTU-6 and PE-5 (the DF).

1. A failure occurs in the LSP between MTU-6 and PE-5. This can be any event that brings the SDP down.

```
# log 99 on PE-5:
85 2021/02/17 15:14:45.918 UTC MINOR: SVCMGR #2303 Base
"Status of SDP 56 changed to admin=up oper=down"
```

2. Immediately, PE-5 gives up the DF role and withdraws the ES route, as well as the AD routes and MAC routes. As soon as PE-4 receives any ES or AD withdraw, it will re-run the DF algorithm and, when the es-activation-timer expires, it will become the DF and activate its spoke-SDP.

```
# log 99 on PE-5:
87 2021/02/17 15:14:45.920 UTC MINOR: SVCMGR #2094 Base
```

```
"Ethernet Segment:ESI-45, EVI:1, Designated Forwarding state changed to:false"
```

The ES in PE-5 is operational down:

```
*A:PE-5# show service system bgp-evpn ethernet-segment name "ESI-45"

===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : ESI-45
Eth Seg Type            : None
Admin State             : Enabled            Oper State         : Down
ESI                     : 01:00:00:00:00:45:00:00:00:01
Multi-homing            : singleActive       Oper Multi-homing  : singleActive
ES SHG Label            : 524282
Source BMAC LSB         : <none>
Sdp Id                  : 56
ES Activation Timer     : 3 secs
Oper Group              : (Not Specified)
Svc Carving             : auto               Oper Svc Carving   : auto
Cfg Range Type          : primary
===============================================================================
```

PE-5 is no longer the DF and the only DF candidate is PE-4:

```
*A:PE-5# show service system bgp-evpn ethernet-segment name "ESI-45" evi 1

===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
1             1             0                   no  02/17/2021 15:14:46
===============================================================================

===============================================================================
DF Candidates                           Time Added
-------------------------------------------------------------------------------
192.0.2.4                               02/17/2021 15:04:45
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

PE-4 becomes the DF and the spoke-SDP 46:1 is brought up.

```
# log 99 on PE-4:
88 2021/02/17 15:14:48.951 UTC MINOR: SVCMGR #2094 Base
"Ethernet Segment:ESI-45, EVI:1, Designated Forwarding state changed to:true"

89 2021/02/17 15:14:48.951 UTC MINOR: SVCMGR #2326 Base
"Status of SDP Bind 46:1 in service 1 (customer 1) local PW status bits changed to none"

90 2021/02/17 15:14:48.951 UTC MINOR: SVCMGR #2306 Base
"Status of SDP Bind 46:1 in service 1 (customer 1) changed to admin=up oper=up flags="
```

The ES is up in PE-4:

```
*A:PE-4# show service system bgp-evpn ethernet-segment name "ESI-45"

===============================================================================
Service Ethernet Segment
```

```
===============================================================================
Name                    : ESI-45
Eth Seg Type            : None
Admin State             : Enabled            Oper State         : Up
ESI                     : 01:00:00:00:00:45:00:00:00:01
Multi-homing            : singleActive       Oper Multi-homing  : singleActive
ES SHG Label            : 524278
Source BMAC LSB         : <none>
Sdp Id                  : 46
ES Activation Timer     : 3 secs
Oper Group              : (Not Specified)
Svc Carving             : auto               Oper Svc Carving   : auto
Cfg Range Type          : primary
===============================================================================
```

PE-4 is the DF and there are no other DF candidates:

```
*A:PE-4# show service system bgp-evpn ethernet-segment name "ESI-45" evi 1

===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
1             1             0                   yes 02/17/2021 15:14:49
===============================================================================

===============================================================================
DF Candidates                           Time Added
-------------------------------------------------------------------------------
192.0.2.4                               02/17/2021 15:04:28
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

3. The remote PEs, PE-2 and PE-3, receive the BGP-EVPN routes withdrawal and modify the next-hop for the EVPN-MPLS ES destination.

```
241 2021/02/17 15:14:45.921 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 86
    Flag: 0x90 Type: 15 Len: 82 Multiprotocol Unreachable NLRI:
        Address Family EVPN
        Type: EVPN-AD Len: 25 RD: 192.0.2.5:1 ESI: 01:00:00:00:00:45:00:00:00:01,
                      tag: MAX-ET Label: 0
        Type: EVPN-AD Len: 25 RD: 192.0.2.5:1 ESI: 01:00:00:00:00:45:00:00:00:01,
                      tag: 0 Label: 0
        Type: EVPN-ETH-SEG Len: 23 RD: 192.0.2.5:0 ESI: 01:00:00:00:00:45:00:00:00:01,
                      IP-Len: 4 Orig-IP-Addr: 192.0.2.5
"
```

The ES EVPN-MPLS destination is resolved to the DF PE-4:

```
*A:PE-2# show service id 1 evpn-mpls esi 01:00:00:00:00:45:00:00:00:01

===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                       Num. Macs               Last Change
-------------------------------------------------------------------------------
```

```
01:00:00:00:00:45:00:00:00:01   1                       02/17/2021 15:15:06
===============================================================================

===============================================================================
BGP EVPN-MPLS Dest TEP Info
===============================================================================
TEP Address              Egr Label                Last Change
                         Transport:Tnl-Id
-------------------------------------------------------------------------------
192.0.2.4                524280                   02/17/2021 15:15:06
                         ldp:65538
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

The following must be considered:

- The DF election procedure is revertive, that is, when the failed SDP comes back up, PE-5 will take over again as DF and the network will re-converge.
- The DF election is triggered by the following events:
  - **configure service system bgp-evpn ethernet-segment ESI-45 no shutdown** triggers the DF election for all the services in the ES.
  - A new update/withdrawal of an ES route (containing an ESI configured locally) triggers the DF election for all the services in the ESI.
  - A new update/withdrawal of an AD per-ES route (containing an ESI configured locally) triggers the DF election for all the services associated with the list of RTs received along with the route.
  - A new update of an AD per-ES route with a change in the ESI-label extended community (single-active bit or MPLS label) triggers the DF election for all the services associated with the list of RTs received along with the route.
  - A new update/withdrawal of an AD route per-EVI (containing an ESI configured locally) triggers the DF election for that service.

BGP-EVPN route selection in EVPN networks

The selection of the best route for a MAC address is as follows:

- If a PE receives more than one route for the same MAC address, the best MAC route is chosen:
  - If the route key is equal in two or more routes (that is, the mac, mac-length, ip, ip-length, RD, eth-tag), then regular BGP selection applies:
    - If local-pref, AS-path, origin, and MED are equal, the lowest IGP distance to the BGP next-hop is chosen (unless **ignore-nh-metric** is configured). If the BGP next-hop is resolved by an LSP, the cost from the tunnel-table is used.
    - As a last resort tie-breaker, the route with the lowest originator ID, or received from the peer with the lowest BGP Identifier, is chosen (unless **ignore-router-id** is configured and the routes being compared are EBGP routes).
  - If the mac-length, mac, ip-length, ip, eth-tag are equal, and the RD is different, the EVPN selection process is applied in the following order:
    - Conditional static MAC addresses (local protected MAC addresses)
    - EVPN static MAC addresses (remote protected MAC addresses)

    - Data plane learned MAC addresses (regular learning on SAPs/SDP-bindings)
    - EVPN MAC addresses with higher sequence number
    - Lowest IP address (next-hop IP of the EVPN NLRI)
    - Lowest Ethernet tag (will be normally zero)
    - Lowest RD

- After a MAC route is selected, the system checks for an associated ES.
  - If it has an ES, the system uses the MAC address as the EVPN-MPLS ES destination. The ES destination is constructed based on the AD per-EVI routes received for that ES (regardless of MAC address priorities with the ES).
  - The system selects the first ECMP number of AD per-EVI routes arranged by the IP address of PEs (lower IPs are selected first).
  - If the same PE has advertised multiple RDs, the system selects the route with the lowest RD for that PE.

In the example of Figure 62: EVPN-MPLS for VPLS services, PE-4 resolves the next-hops for ESI-23 as described in the second choice above, that is, because ECMP=2, the two available next-hops are chosen. If ECMP is changed to 1, PE-4 will pick up the lower IP (in the BGP next-hop). This is illustrated in the following output:

```
*A:PE-4# show service id 1 evpn-mpls esi 01:00:00:00:00:23:00:00:00:01

===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                      Num. Macs               Last Change
-------------------------------------------------------------------------------
01:00:00:00:00:23:00:00:00:01   1                       02/17/2021 15:10:16
===============================================================================

===============================================================================
BGP EVPN-MPLS Dest TEP Info
===============================================================================
TEP Address               Egr Label                 Last Change
                          Transport:Tnl-Id
-------------------------------------------------------------------------------
192.0.2.2                 524281                    02/17/2021 15:10:16
                          ldp:65538
192.0.2.3                 524282                    02/17/2021 15:10:16
                          ldp:65537
-------------------------------------------------------------------------------
Number of entries : 2
-------------------------------------------------------------------------------
===============================================================================
```

When ECMP equals 1, only the BGP next hop with the lower IP is chosen:

```
# on PE-4:
configure
    service
        vpls "VPLS1"
            bgp-evpn
                mpls
                    ecmp 1
                exit
```

```
*A:PE-4# show service id 1 evpn-mpls esi 01:00:00:00:00:23:00:00:00:01

===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                       Num. Macs               Last Change
-------------------------------------------------------------------------------
01:00:00:00:00:23:00:00:00:01   1                       02/17/2021 15:22:05
===============================================================================

===============================================================================
BGP EVPN-MPLS Dest TEP Info
===============================================================================
TEP Address              Egr Label                Last Change
                         Transport:Tnl-Id
-------------------------------------------------------------------------------
192.0.2.2                524281                   02/17/2021 15:10:16
                         ldp:65538
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

Comparing EVPN multi-homing and BGP multi-homing

EVPN-MPLS services support EVPN-MH (EVPN multi-homing) and also BGP-MH as in chapter BGP Multi-Homing for VPLS Networks. While EVPN-MH is the standard way of providing access resiliency in RFC 7432, BGP-MH is also a standard mechanism supported in VPLS or EVPN networks. The following table provides some comparison between both technologies.

Table 3: Comparing EVPN multi-homing and BGP multi-homing

VPN Requirements	EVPN-MH	BGP-MH	Comments
All-active MH (flow-based load-balancing)	Yes	No	EVPN-MH provides better bandwidth utilization
Single-active MH (service-based load-balancing)	Yes	Yes	
DF PE election - automatic service balancing	Yes Service-carving	No Requires vsi policies and LP manipulation	EVPN-MH provides better automation
DF PE election – manual configuration per service	Yes	No	EVPN-MH allows for manual DF config for EVIs and ISIDs (2 PEs)
Split-horizon indication in the data plane	Yes ESI-label	No	Prevents transient loops when dual-active DFs show up
DF indication in the control plane	No	Yes	BGP MH guarantees one DF at a time. EVPN relies on Timers to ensure one DF at a time

VPN Requirements	EVPN-MH	BGP-MH	Comments
Allows multiple SAPs or SDP-bindings per service on the same site	No	Yes Through the use of SHGs	
Boot timer and site(es)-activation-timers	Yes	Yes	BGP-MH supports more granular configuration (service level)
Support for oper-groups	No	Yes	
Non-DF notification to the CE (MPLS and CFM)	Yes	Yes	Avoids blackholing

In addition to the preceding comparison, the following configuration excerpt compares EVPN-MH with BGP-MH on a bgp-evpn VPLS service and shows that, while EVPN-MH does not have any configuration at service level, BGP-MH is configured within the VPLS context, which gives a more granular control over the redundancy provided. See the BGP Multi-Homing for VPLS Networks chapter for more information about BGP-MH.

```
*A:PE-4>config>service>system>bgp-evpn# info
----------------------------------------------
                ethernet-segment "ESI-45"  create
                    esi 01:00:00:00:00:45:00:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 46
                    no shutdown
                exit
----------------------------------------------
```

```
*A:PE-4>config>service>vpls# info
----------------------------------------------
            bgp
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            spoke-sdp 46:1 create
                no shutdown
            exit
            no shutdown
----------------------------------------------
```

For BGP multi-homing, site "site-1" is configured, as follows. The RD needs to be configured in the **bgp** context.

```
*A:PE-4>config>service>vpls# info
----------------------------------------------
            bgp
                route-distinguisher 192.0.2.4:1
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            site "site-1" create
                site-id 1
                spoke-sdp 46:1
                site-activation-timer 3
                no shutdown
            exit
            spoke-sdp 46:1 create
                no shutdown
            exit
            no shutdown
----------------------------------------------
```

Proxy-ARP/ND configuration for EVPN-MPLS networks

Although not strictly a BGP-EVPN configuration, **vpls>proxy-arp** and **vpls>proxy-nd** functions are typically enabled along with EVPN-MPLS in order to reduce the amount of flooding in the network. The proxy-ARP/ND agent in the VPLS service will snoop ARP-requests and/or Neighbor Solicitation messages and will reply to those messages locally (if the information is known) without having to flood the requests to the network.

The configuration options for proxy-ARP are the following:

```
*A:PE-2>config>service>vpls# proxy-arp ?
  - no proxy-arp
  - proxy-arp

 [no] age-time        - Configure aging timer for proxy ARP entries
      dup-detect      - Configure anti-spoofing MAC address information
 [no] dynamic         + Configure dynamic entry information
 [no] dynamic-arp-po* - Configure population of dynamic proxy ARP entries
 [no] evpn-route-tag  - Configure EVPN Route Tag information
 [no] garp-flood-evpn - Configure to flood GARP request/replys into EVPN
 [no] send-refresh    - Configure send refresh time
 [no] shutdown        - Administratively enable/disable proxy ARP configuration
 [no] static          - Configure static IP address to MAC address associations
      table-size      - Configure the maximum number of entries in the proxy ARP table
 [no] unknown-arp-re* - Configure to flood unknown ARP request
```

The configuration options for proxy-ND are the following:

```
*A:PE-2>config>service>vpls# proxy-nd ?
  - no proxy-nd
  - proxy-nd

 [no] age-time        - Configure aging timer for proxy ND entries
      dup-detect      - Configure anti-spoofing MAC address information
 [no] dynamic         + Configure dynamic entry information
 [no] dynamic-nd-pop* - Configure population of dynamic proxy ND entries
      evpn-nd-advert* - Configure EVPN Neighbor Discovery advertisements
 [no] evpn-route-tag  - Configure EVPN Route Tag information
 [no] host-unsolicit* - Configure whether to flood evpn with host  neighbor
                        advertisement
 [no] router-unsolic* - Configure whether to flood evpn with router neighbor
                        advertisement
 [no] send-refresh    - Configure send refresh time
 [no] shutdown        - Administratively enable/disable proxy ND configuration
 [no] static          - Configure static IP address to MAC address associations
      table-size      - Configure the maximum number of entries in the proxy ND table
 [no] unknown-ns-flo* - Configure to flood unknown ND solicitation
```

When proxy-ARP/ND is enabled, the following configuration guidelines must be followed:

- **dynamic-arp-populate** or **dynamic-nd-populate** should be used only in networks with a consistent configuration of this command in all PEs.
- When using **dynamic-arp-populate**/**dynamic-nd-populate**, the **age-time** value should be configured to a value equal to three times the **send-refresh** value. This will help reduce the EVPN withdrawals and re-advertisements in the network.
- With large **age-time** values, it would be sufficient to configure the **send-refresh** value to half of the proxy-ARP/ND **age-time** or FDB **age-time**.
- In scaled environments (with thousands of services), it is not recommended to set the send-refresh value to less than 300 s. In such scenarios, Nokia recommends using a minimum proxy-ARP/ND **age-time** and FDB **age-time** of 900 s.
- The use of the following commands reduces or suppresses the ARP/ND flooding in an EVPN network, because EVPN MAC routes replace the function of the regular data plane ARP/ND messages:
  - **no garp-flood-evpn**
  - **no unknown-arp-request-flood-evpn**
  - **no unknown-ns-flood-evpn**
  - **no host-unsolicited-na-flood-evpn**
  - **no router-unsolicited-na-flood-evpn**
- Nokia recommends using the preceding commands only in EVPN networks where the CEs are routers directly connected to an SR OS node acting as the PE. Networks using aggregation switches between the host/routers and the PEs should flood GARP/ND messages in EVPN to make sure the remote caches are updated and BGP does not miss the advertisement of these entries.
- When the **anti-spoof-mac** is used with proxy-ARP/ND, ingress filters (in the access SAPs/SDP-bindings) should be configured to drop all traffic with destination anti-spoof-mac. The same MAC address should be configured in all PEs where dup-detect is active.
- When proxy-ND is used, the configuration of the following commands should be consistent in all the PEs in the network:

  - **router-unsolicited-na-flood-evpn**
  - **host-unsolicited-na-flood-evpn**
  - **evpn-nd-advertise**
- Because EVPN does not propagate the **router** flag in IPv6--> MAC address advertisements, in a mixed network with hosts and routers where **evpn-nd-advertise router** is configured, unsolicited host NA messages should be flooded so that the entire network gets to learn all of the host and router ND entries. In the same way, **evpn-nd-advertise host** should be configured so that unsolicited router NA messages are flooded.

Finally, along with proxy-ARP/ND, **vpls>discard-unknown** may be used in some EVPN-MPLS deployments where all the CEs are routers and they announce themselves to the network by sending GARPs or NAs (Neighbor Solicitation messages). According to RFC 7432, whether or not to flood packets to unknown destination MAC addresses should be an administrative choice, depending on how learning happens between CEs and PEs. **Discard-unknown** provides that administrative choice in case all the MAC addresses in an EVI can be learned even before any traffic is exchanged.

Proxy-ARP/ND along with **discard-unknown** helps reduce the BUM traffic in an EVPN network significantly; however, their use must be analyzed and considered, depending on the type of CEs in the EVI.

An example of proxy-ARP configuration is as follows. This configuration should be added to all PEs. When a new ARP message is received on any of the PEs, they will learn the IP-MAC address pair and will advertise it to the network.

```
# on PE-2, PE-3, PE-4, PE-5:
configure
    service
        vpls "VPLS1"
            proxy-arp
                age-time 900
                send-refresh 300
                dynamic-arp-populate
                no shutdown
            exit
```

Enabling proxy-ARP increases the number of MAC/IP routes being sent by the PEs. This is due to the following reasons:

- An additional MAC/IP route will be advertised per new learned IP-MAC address pair, regardless of having advertised the same MAC address already.
- A MAC per VPLS service will be advertised with a system MAC address. That MAC address will be used as MAC SA for proxy-ARP confirm messages when an IP moves to a different PE.

The following output shows the MAC/IP routes on PE-2 when proxy-ARP is enabled in the network.

```
*A:PE-2# show router bgp routes evpn mac
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
```

```
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.3:1         02:17:ff:00:03:3a ESI-0
      0                   Static            LABEL 524282
                          n/a
                          192.0.2.3

u*>i  192.0.2.4:1         02:1b:ff:00:03:3a ESI-0
      0                   Static            LABEL 524280
                          n/a
                          192.0.2.4

u*>i  192.0.2.5:1         00:00:16:16:16:16 01:00:00:00:00:45:00:00:00:01
      0                   Seq:0             LABEL 524280
                          n/a
                          192.0.2.5

u*>i  192.0.2.5:1         02:1f:ff:00:03:3a ESI-0
      0                   Static            LABEL 524280
                          n/a
                          192.0.2.5

-------------------------------------------------------------------------------
Routes : 4
===============================================================================
```

Troubleshooting and debug commands

When troubleshooting an EVPN-MPLS network, the following show commands and debug commands are recommended, as already discussed throughout this chapter:

- **show redundancy bgp-evpn-multi-homing**
- **show router bgp routes evpn** (and filters)
- **show service evpn-mpls [**<*TEP ip-address*>**]**
- **show service id bgp-evpn**
- **show service id evpn-mpls** (and modifiers)
- **show service id fdb** (and modifiers)
- **show service system bgp-evpn**
- **show service system bgp-evpn ethernet-segment** (and modifiers)
- **debug router bgp update**
- **log-id 99**

In addition to the preceding commands, the following tools dump commands may also help:

- **tools dump service evpn usage** — This command shows the amount of EVPN-MPLS (and EVPN-VXLAN) destinations consumed in the system.
- **tools dump service system bgp-evpn ethernet-segment** <*name*> **evi <[1..65535]> df** — This command computes the DF election for a specific ESI and EVI. Note: The **show service system bgp-evpn ethernet-segment** commands shows whether the local PE is DF or non-DF for a specific EVI,

but it does not show who the DF is if it is not the local PE. In case of more than 2 PEs in the ES, this command may be especially useful.

Some examples are provided below for PE-2. PE-2 is showing seven EVPN-MPLS destinations due to the following:

- Each remote PE consumes one EVPN-MPLS destination for unicast (if they advertise MAC/IP routes to PE-2 and the ingress-replication-bum-label is configured in all the PEs). PE-2 has three remote unicast EVPN-MPLS destinations.
- Each remote PE consumes one EVPN-MPLS destination for multicast (if they advertise inclusive multicast routes to PE-2). PE-2 has three remote multicast EVPN-MPLS destinations.
- Each remote ES consumes one EVPN-MPLS destination (it is only one per ES, regardless of the multi-homing mode and the number of PEs in the ES). PE-2 has one remote ES (ESI-45).

```
*A:PE-2# tools dump service evpn usage

vxlan-evpn-mpls usage statistics at 02/17/2021 15:38:31:

MPLS-TEP                                       :             3
VXLAN-TEP                                      :             0
Total-TEP                                      :      3/ 16383

Mpls Dests (TEP, Egress Label + ES + ES-BMAC)   :             7
Mpls Etree Leaf Dests                          :             0
Vxlan Dests (TEP, Egress VNI + ES)             :             0
Total-Dest                                     :      7/196607

Sdp Bind +  Evpn Dests                         :      8/245759
ES L2/L3 PBR                                   :      0/ 32767
Evpn Etree Remote BUM Leaf Labels              :             0
```

To compute the DF election for EVI 1:

```
*A:PE-2# tools dump service system bgp-evpn ethernet-segment "ESI-23" evi 1 df

[02/17/2021 15:39:51] Computed DF: 192.0.2.3 (Remote) (Boot Timer Expired: Yes)
```

Conclusion

SR OS has a full RFC 7432 EVPN-MPLS implementation including single-active and all-active multi-homing. This example has shown how to configure and operate EVPN-MPLS for a simple non multi-homing configuration as well as a multi-homing configuration. Other topics, such as the integration of VPLS objects with EVPN-MPLS and proxy-ARP/ND, have also been discussed.

EVPN for MPLS Tunnels in Epipe Services (EVPN-VPWS)

This chapter provides information about EVPN for MPLS tunnels in Epipe services (EVPN-VPWS).

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS Release 14.0.R4, but the CLI in the current edition is based on SR OS Release 22.10.R1. Ethernet Virtual Private Network - Virtual Private Wire Service (EVPN-VPWS) is supported in SR OS Release 14.0.R1 and later. EVPN-VPWS in multi-homing scenarios is supported in SR OS Release 14.0.R4 and later.

Chapter EVPN for MPLS Tunnels is prerequisite reading.

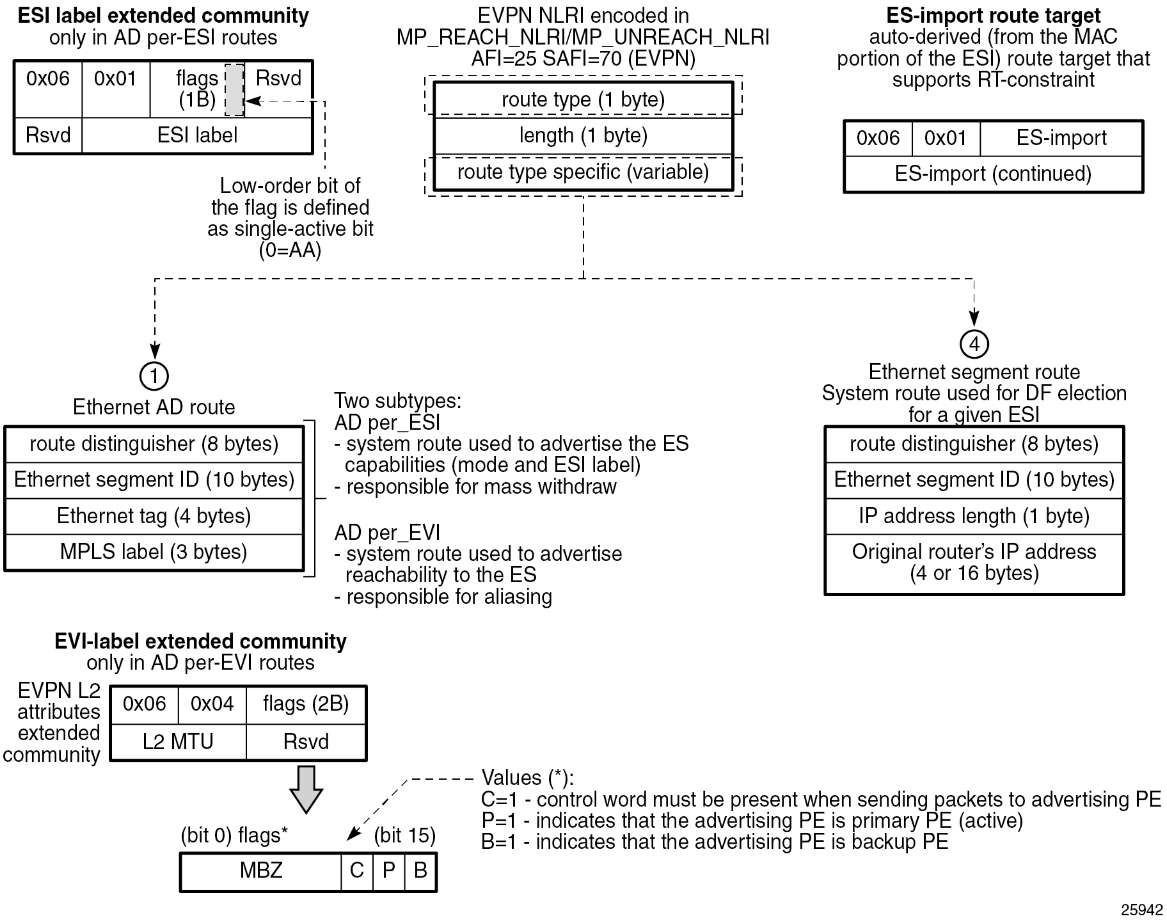
Overview

Service providers prefer an optimized, standardized, and unified control plane for VPNs. EVPN-VPWS is supported in MPLS networks that also run EVPN-MPLS in VPLS services. From a control plane perspective, EVPN-VPWS is a simplified point-to-point version of RFC 7432, *BGP MPLS-Based Ethernet VPN*, because there is no need to advertise MAC routes in VPWS. EVPN-VPWS is described in RFC 8214, *Virtual Private Wire Service Support in Ethernet VPN*.

EVPN-VPWS supports all-active multi-homing (per-flow load-balancing multi-homing) as well as single-active multi-homing (per-service load-balancing multi-homing), using the same Ethernet segments (ESs) used for EVPN-MPLS VPLS services. EVPN-VPWS uses route-type 1 and route-type 4; it does not use route-types 2, 3, or 5, because MAC/IP routes, inclusive multicast, or IP-prefix routes are not required.

Figure 65: Route types and NLRIs for EVPN-VPWS shows the encoding of the required extensions for the route-types 1 and 4 for EVPN-VPWS.

Figure 65: Route types and NLRIs for EVPN-VPWS


ESI label extended community
only in AD per-ESI routes
0x06
0x01
flags (1B)
Rsvd
Rsvd
ESI label
Low-order bit of the flag is defined as single-active bit (0=AA)
EVPN NLRI encoded in
MP_REACH_NLRI/MP_UNREACH_NLRI
AFI=25 SAFI=70 (EVPN)
route type (1 byte)
length (1 byte)
route type specific (variable)
ES-import route target
auto-derived (from the MAC portion of the ESI) route target that supports RT-constraint
0x06
0x01
ES-import
ES-import (continued)
1
Ethernet AD route
route distinguisher (8 bytes)
Ethernet segment ID (10 bytes)
Ethernet tag (4 bytes)
MPLS label (3 bytes)
Two subtypes:
AD per_ESI
- system route used to advertise the ES capabilities (mode and ESI label)
- responsible for mass withdraw
AD per_EVI
- system route used to advertise reachability to the ES
- responsible for aliasing
4
Ethernet segment route
System route used for DF election for a given ESI
route distinguisher (8 bytes)
Ethernet segment ID (10 bytes)
IP address length (1 byte)
Original router's IP address (4 or 16 bytes)
EVI-label extended community
only in AD per-EVI routes
EVPN L2 attributes extended community
0x06
0x04
flags (2B)
L2 MTU
Rsvd
(bit 0) flags*
(bit 15)
MBZ
C
P
B
Values (*):
C=1 - control word must be present when sending packets to advertising PE
P=1 - indicates that the advertising PE is primary PE (active)
B=1 - indicates that the advertising PE is backup PE
25942

Two sub-types are defined for route-type 1. Route-type 4 has no sub-types. The route types used for EVPN-VPWS have the following purposes:

- Route-type 1 - Auto-discovery per EVPN instance (AD per-EVI). This route type is used in all EVPN-VPWS scenarios, with or without multi-homing. For EVPN-VPWS, the Ethernet tag field is encoded with the local Attachment Circuit (AC) of the advertising PE. This value is configured using the **service epipe bgp-evpn local-attachment-circuit eth-tag** <*value*> command. The route distinguisher (RD), MPLS label, and the Ethernet segment ID (ESI) are encoded as for EVPN-MPLS. The MPLS label field is used as service label. In case of multi-homing, AD per-EVI routes containing the same ESI are used to provide aliasing and a backup path to the PEs part of the ES. The L2 MTU is encoded with the service MTU configured in the Epipe. The following flags are used for EVPN-VPWS:
  - Flag C is set if a control word is configured in the service.
  - Flag P is set if the advertising PE is primary PE.
    - If no multi-homing is used, there is no primary PE (P=0).
    - In all-active multi-homing, all PEs in the ES are primary (P=1).
    - In single-active multi-homing, only one PE per-EVI in the ES is primary (P=1).

- Flag B is set if the advertising PE is backup PE.
  - The B-flag is only set in case of single-active multi-homing and only for one PE, even if more than two PEs are present in the same single-active ES. The backup PE is the winner of the second Designated Forwarder (DF) election (excluding the DF). The remaining non-DF PEs send B=0.

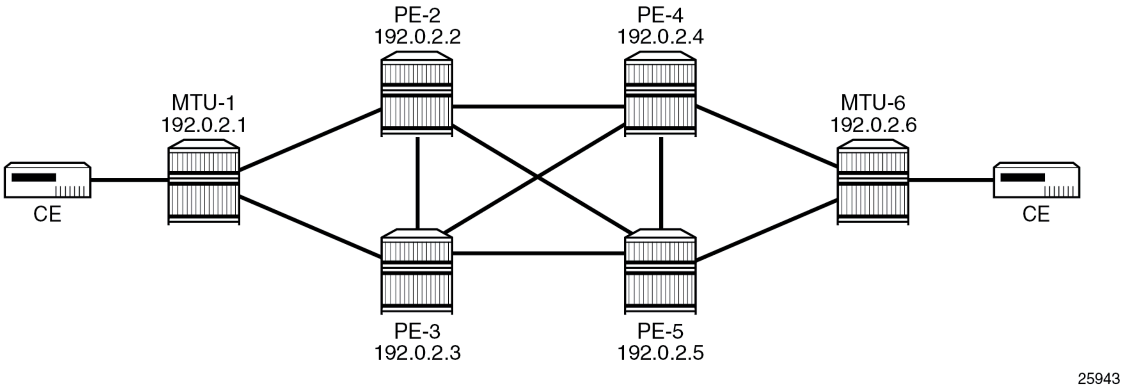
  If there is no multi-homing, the ESI, flag P, and flag B will be zero.

- Route-type 1 - AD per Ethernet segment (AD per-ES). Same encoding as for EVPN-MPLS. AD per-ES is only used in multi-homing scenarios where it is advertised per ES from the PE. It carries the ESI label (used for split-horizon, but only for VPLS services and not for Epipe services) and can affect procedures such as the DF election, as well as the aliasing on remote PEs.
- Route-type 4 - ES route. Same encoding as for EVPN-MPLS. Route-type 4 is only used in multi-homing scenarios. This route advertises a local configured ES. The exchange of this route can discover remote PEs that are part of the same ES and the DF election algorithm among them.

Configuration

Figure 66: EVPN-VPWS example topology shows the example topology that will be used throughout this chapter.

Figure 66: EVPN-VPWS example topology



The example topology consists of six SR OS nodes with the following initial configuration:

- Network (or hybrid) ports interconnect the core PEs with configured router interfaces.
- MTU-1 is a pure Ethernet aggregator. The ports toward the core PEs are access ports. Likewise, the ports on PE-2 and PE-3 toward MTU-1 are access ports.
- Core PEs and MTU-6 run IS-IS on all router interfaces. Point-to-point adjacencies are established for the exchange of system IP addresses.
- Link LDP is configured between all PEs, and toward/from MTU-6.
- EVPN uses BGP for exchanging reachability at service level. Therefore, BGP peering sessions must be established among the core PEs for the EVPN family. Although typically a separate router is used, in this chapter, PE-2 is used as route reflector with the following BGP configuration:

```
# on PE-2:
```

```
configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family evpn
                cluster 192.0.2.2
                peer-as 64500
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
        exit
```

The BGP configuration on the other PEs is as follows:

```
# on PE-3, PE-4, PE-5:
configure
    router
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family evpn
                peer-as 64500
                neighbor 192.0.2.2
                exit
            exit
        exit
```

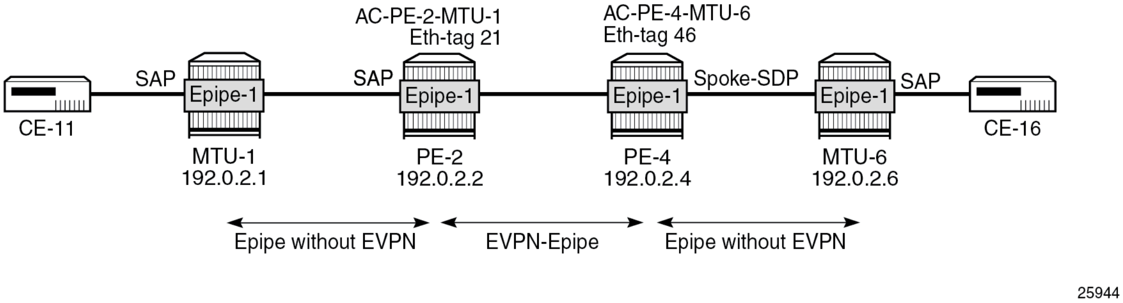
The following EVPN-VPWS scenarios are described in the following sections:

- EVPN for MPLS tunnels in Epipe services without multi-homing
- EVPN for MPLS tunnels in Epipe services with all-active multi-homing
- EVPN for MPLS tunnels in Epipe services with single-active multi-homing

EVPN for MPLS tunnels in Epipe services without multi-homing

BGP-EVPN can be enabled in Epipe services with either SAPs or spoke-SDPs at the access, as shown in Figure 67: Example topology for EVPN-VPWS without multi-homing.

Figure 67: Example topology for EVPN-VPWS without multi-homing



On PE-2, Epipe 1 is configured as follows:

```
# on PE-2:
configure
    service
        epipe 1 name "Epipe-1" customer 1 create
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-PE-2-MTU-1 create
                    eth-tag 21
                exit
                remote-attachment-circuit AC-PE-4-MTU-6 create
                    eth-tag 46
                exit
                evi 1
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            sap 1/1/c11/1:1 create
                no shutdown
            exit
            no shutdown
```

On PE-4, the service configuration is as follows:

```
# on PE-4:
configure
    service
        sdp 460 create
            far-end 192.0.2.6
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        epipe 1 name "Epipe-1" customer 1 create
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-PE-4-MTU-6 create
                    eth-tag 46
                exit
                remote-attachment-circuit AC-PE-2-MTU-1 create
                    eth-tag 21
                exit
                evi 1
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            spoke-sdp 460:1 create
                no shutdown
            exit
            no shutdown
```

Where the following commands are relevant for the EVPN-VPWS configuration:

- **bgp** enables the context for the BGP configuration relevant to the service. The **bgp** context configures the common BGP parameters for all BGP families in the service, such as route distinguisher and route target. Even if the general BGP parameters for the service are auto-derived, the **bgp** context must be enabled.

```
*A:PE-2>config>service>epipe# bgp ?
  - bgp
  - no bgp

 [no] adv-service-mtu - Configure service-mtu to be advertised
 [no] pw-template-bi* + Configure pw-template bind policy
 [no] route-distingu* - Configure route distinguisher
 [no] route-target    - Configure route target
 [no] vsi-export      - VSI export route policies
 [no] vsi-import      - VSI import route policies
```

- The following parameters can be configured in the **bgp-evpn** context:

```
*A:PE-2>config>service>epipe# bgp-evpn ?
  - bgp-evpn
  - no bgp-evpn

 [no] evi             - EVPN Identifier
 [no] local-attachme* + Configure local attachment circuit information
 [no] mpls            + Configure BGP EVPN mpls
 [no] remote-attachm* + Configure remote attachment circuit information
 [no] segment-routin* + Configure SRv6 instance
 [no] vxlan           + Configure BGP EVPN vxlan
```

  – The **evi** is a two-byte or three-byte identifier used for auto-deriving the service RD (only for two-byte EVI), service RT, and for the DF election in multi-homing. The auto-derivation of RD and RT for a two-byte EVI is as follows:

    - RD <system IP address>:<evi>
    - RT <autonomous system number>:<evi>

    The EVI values must be unique in the system, regardless of the type of service they are assigned to (Epipe or VPLS).

Note: Three-byte EVI values are supported in SR OS Release 21.10.R1 and later. For auto-derived RT as per RFC 8365, the **evi-three-byte-auto-rt** command must be configured, as described in the Three-byte EVI in EVPN Services chapter.

- The **local-attachment-circuit** and **remote-attachment-circuit** identify the two attachment circuits connected by the EVPN-VPWS service. The configured Ethernet tag for the local AC is advertised in the Ethernet tag field of the AD per-EVI route for the Epipe, along with the corresponding RD, RT, and MPLS label. Both local and remote Ethernet tags are mandatory to bring up the Epipe service. If the received Ethernet tag for the Epipe service matches the configured remote AC Ethernet tag, it will create an EVPN-MPLS destination to the next hop.

  The local Ethernet tag cannot be modified without disabling **bgp-evpn mpls** in the Epipe, as shown in the following output:

  ```
  *A:PE-2>config>service>epipe>bgp-evpn>local-att-cir# eth-tag 221
  MINOR: SVCMGR #8036 evpn-vpws ac eth-tag not allowed - cannot change while evpn mpls/
  vxlan/srv6 is enabled
  ```

  Unlike local Ethernet tags, remote Ethernet tags can be modified without disabling **bgp-evpn**.

- The following configuration options are available for Epipes in the **bgp-evpn>mpls** context:

  ```
  *A:PE-2>config>service>epipe>bgp-evpn# mpls ?
    - mpls [bgp <bgp>]
    - no mpls [bgp <bgp>]

   <bgp>                : [1..1]


       auto-bind-tunn* + Configure BGP EVPN mpls auto-bind-tunnel
   [no] control-word    - Enable/disable setting the CW bit in the label message
   [no] default-route-* - Configure default-route-tag to match against export policies
   [no] dynamic-egress* - Enable/disable Dynamic Egress Label Limit
        ecmp            - Configure maximum ECMP routes information
   [no] entropy-label   - Enable/disable use of entropy-label
   [no] evi-three-byte* - Enable/Disable evi-three-byte-auto-rt
   [no] force-qinq-vc-* - Forces qinq-vc-type forwarding in the data-path
   [no] force-vlan-vc-* - Forces vlan-vc-type forwarding in the data-path
   [no] oper-group      - Configure oper-group
        route-next-hop  - Configure route next-hop
   [no] send-tunnel-en* - Configure encapsulation for this service
   [no] shutdown        - Administratively Enable/Disable BGP-EVPN mpls
  ```

  This is a subset of the options for VPLS services; see chapter EVPN for MPLS Tunnels.

When the local AC (SAP 1/1/c11/1:1) is up, PE-2 sends a BGP EVPN AD per-EVI route that contains Ethernet tag 21 for the local AC:

```
# on PE-2:
2 2022/11/29 09:33:44.668 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 81
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-AD Len: 25 RD: 192.0.2.2:1 ESI: ESI-0, tag: 21 Label: 8388512 (Raw Label:
 0x7fffa0) PathId:
```

```
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:1
        l2-attribute:MTU: 1514 C: 0 P: 0 B: 0
        bgp-tunnel-encap:MPLS
"
```

The auto-derived RD for EVI 1 is 192.0.2.2:1 and the RT is 64500:1.

When the remote AC on PE-4 (spoke-SDP 460:1) is up, PE-2 receives the following EVPN-AD per-EVI route with Ethernet tag 46 from PE-4:

```
# on PE-2:
4 2022/11/29 09:33:54.253 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 81
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.4
        Type: EVPN-AD Len: 25 RD: 192.0.2.4:1 ESI: ESI-0, tag: 46 Label: 8388512 (Raw Label:
 0x7fffa0) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:1
        l2-attribute:MTU: 1514 C: 0 P: 0 B: 0
        bgp-tunnel-encap:MPLS
"
```

When the received RT 64500:1 matches and the received Ethernet tag 46 matches the configured remote AC on PE-2, the following EVPN-MPLS destination (comprised of a termination endpoint (TEP) 192.0.2.4 and egress label 524282) is created on PE-2. In a similar way, an EVPN-MPLS destination is created on PE-4.

```
*A:PE-2# show service id 1 evpn-mpls

===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address                         Egr Label               Last Change
                                     Transport:Tnl-id
-------------------------------------------------------------------------------
192.0.2.4                           524282                  11/29/2022 09:33:54
                                    ldp:65538
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                               Last Change
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

The MPLS label in the debug message is not the same as in the service, because the router will strip the extra four lowest bits to get the 20-bit MPLS label. The egress label for the EVPN-MPLS destination on PE-4 is 524282. The 24-bit label value in the BGP update debug is 16 (2^4) times as high: 524282*16 = 8388512. This is because the debug message is shown before the router can parse the label field and see if it corresponds to a 20-bit MPLS label or a 24-bit VXLAN VNI.

The BGP AD per-EVI routes for Ethernet tag 46 can be shown with the following command:

```
*A:PE-2# show router bgp routes evpn auto-disc tag 46
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Flag  Route Dist.         ESI                            NextHop
      Tag                                                Label
-------------------------------------------------------------------------------
u*>i  192.0.2.4:1         ESI-0                          192.0.2.4
      46                                                 LABEL 524282

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

The following command shows the BGP EVPN information for Epipe 1:

```
*A:PE-2# show service id 1 bgp-evpn

===============================================================================
BGP EVPN Table
===============================================================================
EVI               : 1                 Creation Origin    : manual

-------------------------------------------------------------------------------
Local AC Name              Eth Tag  Endpoint                    Ingress Label
-------------------------------------------------------------------------------
AC-PE-2-MTU-1              21                                   0
-------------------------------------------------------------------------------
Number of local ACs : 1

-------------------------------------------------------------------------------
Remote AC Name                     Eth Tag  Endpoint
-------------------------------------------------------------------------------
AC-PE-4-MTU-6                      46
-------------------------------------------------------------------------------
Number of Remote ACs : 1
===============================================================================

===============================================================================
BGP EVPN MPLS Information
===============================================================================
Admin Status       : Enabled            Bgp Instance       : 1
Force Vlan Fwding  : Disabled
Force Qinq Fwding  : none
Route NextHop Type : system-ipv4
Control Word       : Disabled
```

```
Max Ecmp Routes    : 1
Entropy Label      : Disabled
Default Route Tag  : none
Oper Group         :
Evi 3-byte Auto-RT : Disabled
Dyn Egr Lbl Limit  : Disabled
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN MPLS Auto Bind Tunnel Information
===============================================================================
Allow-Flex-Algo-Fallback  : false
Resolution                : any               Strict Tnl Tag   : false
Max Ecmp Routes           : 1
Bgp Instance              : 1
Filter Tunnel Types       : (Not Specified)
Weighted Ecmp             : false
-------------------------------------------------------------------------------
===============================================================================
```

Note: Each PE will send its service MTU into the L2 MTU field in the L2-attribute in the AD per-EVI route for the Epipe service. The received L2 MTU will be checked. In case of a mismatch between the received MTU and the configured service MTU, the router will not set up the EVPN destination and, therefore, the service will not come up.

EVPN for MPLS tunnels in Epipe services with multi-homing

SR OS supports EVPN multi-homing as per RFC 8214.

The EVPN multi-homing implementation is based on the concept of the Ethernet segment (ES). An ES is a logical structure that can be defined in one or more PEs and identifies the CE (or access network) multi-homed to the EVPN PEs. An ES is associated with a port, LAG, or SDP object, and is shared by all the services defined on those objects. It can also be shared between Epipe and VPLS services.

Each ES has a unique Ethernet segment Identifier (ESI) that is 10 bytes and is manually configured.

Note: Auto-derived EVPN ESI type 1 as per RFC 7432 is supported in SR OS Release 21.5.R1 and later, as described in the EVPN ESI Type 1 chapter.

The ESI is advertised in the control plane to all the PEs in an EVPN network; therefore, it is very important to ensure that the 10-byte ESI value is unique throughout the entire network. Single-homed CEs are assumed to be connected to an ES with ESI = 0 (single-homed ESs are not explicitly configured).

The ES is part of the base BGP-EVPN configuration and is not applied to any EVPN-MPLS service, by default. An ES can be shared by multiple services; the association of a specific SAP or spoke-SDP to an ES is automatically made when the SAP is defined in the same LAG or port configured in the ES, or when the spoke-SDP is defined in the same SDP configured in the ES.

Regardless of the multi-homing mode, the local Ethernet tag values must match on all the PEs that are part of the same ES. The PEs in the ES will use the AD per-EVI routes from the peer PEs to validate the PEs as DF election candidates for an EVI. The DF election is only relevant for single-active multi-homing ESs. For Epipes defined in an all-active multi-homing ES, there is no DF election required, because all PEs are forwarding traffic and all traffic is treated as unicast.

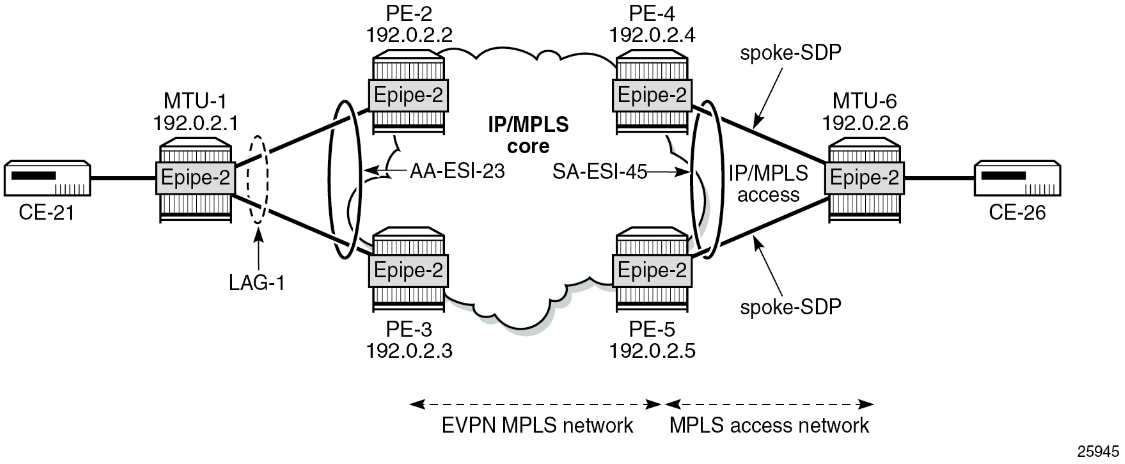
Aliasing is supported when sending traffic to an ES destination. Assuming ECMP is enabled on the ingress PE (and shared queuing or ingress policing), per-flow load-balancing will be performed among all the PEs that advertised P=1. PEs advertising P=0 are not considered as next hops for an ES destination.

The following sections show the configuration of:

- an all-active multi-homing ES with a LAG associated with it
- a single-active multi-homing ES linked to an SDP

Figure 68: Example topology EVPN-VPWS with multi-homing shows an all-active ES and a single-active ES. The all-active multi-homing ES "AA-ESI-23" in PE-2 and PE-3 has a LAG associated to it; the single-active multi-homing ES "SA-ESI-45" in PE-4 and PE-5 has an SDP associated to it.

Figure 68: Example topology EVPN-VPWS with multi-homing



EVPN for MPLS tunnels in Epipe services with all-active multi-homing

All-active multi-homing allows for per-flow load-balancing. Unlike EVPN-MPLS in VPLS services, EVPN-VPWS has no DF election in all-active multi-homing. All PEs in the ES are active and the remote PE will do per-flow load-balancing. AA-ESI-23 is configured on PE-2 and PE-3 in all-active multi-homing with LAG 1 associated to it. This LAG is used as a SAP in Epipe 2 on both PE-2 and PE-3. The configuration of the ES and Epipe 2 is identical on PE-2 and PE-3, including the local AC and remote AC names and Ethernet tags:

```
# on PE-2, PE-3:
configure
    service
        system
            bgp-evpn
                ethernet-segment "AA-ESI-23" create
                    esi 01:00:00:00:00:23:00:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 1
                    no shutdown
                exit
            exit
        exit
        epipe 2 name "Epipe 2" customer 1 create
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-AA-ESI-23-MTU-1 create
                    eth-tag 231
                exit
                remote-attachment-circuit AC-SA-ESI-45-MTU-6 create
                    eth-tag 456
                exit
                evi 2
                mpls bgp 1
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            sap lag-1:2 create
                no shutdown
            exit
            no shutdown
```

See chapter EVPN for MPLS Tunnels for a detailed explanation of the configuration parameters of the ES.

In EVPN-VPWS multi-homing scenarios, three route types are exchanged: AD per-EVI, AD per-ES, and ES routes. The following ES route (route-type 4) for ESI 01:00:00:00:00:23:00:00:00:01 sent by PE-2 is imported at PE-3:

```
# on PE-3:
5 2022/11/29 09:42:41.537 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 71
    Flag: 0x90 Type: 14 Len: 34 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-ETH-SEG Len: 23 RD: 192.0.2.2:0 ESI: 01:00:00:00:00:23:00:00:00:01, IP-Len:
 4 Orig-IP-Addr: 192.0.2.2
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        df-election::DF-Type:Auto/DP:0/DF-Preference:0/AC:1
        target:00:00:00:00:23:00
"
```

The target 00:00:00:00:23:00 in the extended community is derived from the ESI (bytes 2 to 7) and is only imported by the PEs that are part of the same ES; that is, PE-2 and PE-3 in this example.

At the same time, the following AD per-ES route (route-type 1) with maximum Ethernet tag (MAX-ET, all Fs) and label 0 is sent by route reflector (RR) PE-2 and imported by the rest of the PEs. The following two BGP updates with MAX-ET are received by PE-4:

```
# PE-4 receives EVPN AD per-ES (MAX-ET) from PE-2:
3 2022/11/29 09:42:41.491 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 81
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-AD Len: 25 RD: 192.0.2.2:2 ESI: 01:00:00:00:00:23:00:00:00:01, tag: MAX-ET
 Label: 0 (Raw Label: 0x0) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:2
        esi-label:524281/All-Active
        bgp-tunnel-encap:MPLS
"
```

```
# PE-4 receives EVPN AD per-ES (MAX-ET)(originator PE-3):
6 2022/11/29 09:42:43.033 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 95
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-AD Len: 25 RD: 192.0.2.3:2 ESI: 01:00:00:00:00:23:00:00:00:01, tag: MAX-ET
 Label: 0 (Raw Label: 0x0) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.3
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        192.0.2.2
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:2
        esi-label:524282/All-Active
        bgp-tunnel-encap:MPLS
"
```

The ESI label is in the extended community, as well as the indication that the multi-homing is all-active. Epipe services do not require ESI labels because BUM traffic is not recognized as such in EVPN-VPWS services. However, because the ES can be shared by Epipe and VPLS services, the AD per-ES route still includes a non-zero ESI label.

The following AD per-EVI routes (route-type 1) with Ethernet tag 231 sent by RR PE-2 are received and imported on PE-4:

```
# PE-4 receives EVPN AD per-ES with Ethernet tag 231 (originator PE-2):
4 2022/11/29 09:42:41.494 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 81
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-AD Len: 25 RD: 192.0.2.2:2 ESI: 01:00:00:00:00:23:00:00:00:01, tag: 231
 Label: 8388480 (Raw Label: 0x7fff80) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:2
        l2-attribute:MTU: 1514 C: 0 P: 1 B: 0
        bgp-tunnel-encap:MPLS
"
```

```
## PE-4 receives EVPN AD per-ES with Ethernet tag 231 (originator PE-3):
7 2022/11/29 09:42:43.047 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 95
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-AD Len: 25 RD: 192.0.2.3:2 ESI: 01:00:00:00:00:23:00:00:00:01, tag: 231
 Label: 8388496 (Raw Label: 0x7fff90) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.3
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        192.0.2.2
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:2
        l2-attribute:MTU: 1514 C: 0 P: 1 B: 0
        bgp-tunnel-encap:MPLS
"
```

This route contains the flags for control word (C), primary (P), and backup (B). In all-active multi-homing, all nodes are primary (P=1).

PE-4 has learned AD per-EVI/ES routes for AA-ESI-23 from PE-2 and PE-3, as shown in the following output:

```
*A:PE-4# show router bgp routes evpn auto-disc esi 01:00:00:00:00:23:00:00:00:01
===============================================================================
 BGP Router ID:192.0.2.4        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Flag  Route Dist.         ESI                           NextHop
      Tag                                               Label
-------------------------------------------------------------------------------
u*>i  192.0.2.2:2         01:00:00:00:00:23:00:00:00:01 192.0.2.2
      231                                               LABEL 524280

u*>i  192.0.2.2:2         01:00:00:00:00:23:00:00:00:01 192.0.2.2
```

```
      MAX-ET                                            LABEL 0

u*>i  192.0.2.3:2         01:00:00:00:00:23:00:00:00:01 192.0.2.3
      231                                               LABEL 524281

u*>i  192.0.2.3:2         01:00:00:00:00:23:00:00:00:01 192.0.2.3
      MAX-ET                                            LABEL 0

-------------------------------------------------------------------------------
Routes : 4
===============================================================================
```

For Epipe 2 on PE-4, the EVPN MPLS destination is not pointing at a specific TEP, but AA-ESI-23, as shown in the following output:

```
*A:PE-4# show service id 2 evpn-mpls

===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address                         Egr Label               Last Change
                                     Transport:Tnl-id
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================

===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                               Last Change
-------------------------------------------------------------------------------
01:00:00:00:00:23:00:00:00:01           11/29/2022 09:43:02
-------------------------------------------------------------------------------
Number of entries: 1
-------------------------------------------------------------------------------
===============================================================================
```

When ECMP > 1 on the ingress PE, multiple TEPs can correspond to a specific ESI (aliasing). In this case, ECMP=2 and PE-4 and PE-5 have two TEP addresses and egress labels for ESI 01:00:00:00:00:23:00:00:00:01, as shown for PE-4:

```
*A:PE-4# show service id 2 evpn-mpls esi 01:00:00:00:00:23:00:00:00:01

===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                               Last Change
-------------------------------------------------------------------------------
01:00:00:00:00:23:00:00:00:01           11/29/2022 09:43:02
===============================================================================

===============================================================================
BGP EVPN-MPLS Dest TEP Info
===============================================================================
TEP Address              Egr Label                Last Change
                         Transport:Tnl-Id
-------------------------------------------------------------------------------
192.0.2.2                524280                   11/29/2022 09:43:02
                         ldp:65537
192.0.2.3                524281                   11/29/2022 09:43:02
                         ldp:65538
-------------------------------------------------------------------------------
```

```
Number of entries : 2
-------------------------------------------------------------------------------
===============================================================================
```

Note: Even if ECMP is configured, the ingress router will not load-balance the traffic unless shared queuing or ingress policing is configured. This is not specific to EVPN, but generic to the way Epipes forward traffic.

In all-active multi-homing for EVPN-VPWS, there is no DF election and all PEs in the ES are active. For AA-ESI-23, both PE-2 and PE-3 are active/primary/DF, but there are no DF candidates, because there is no DF election:

```
*A:PE-2# show service system bgp-evpn ethernet-segment name "AA-ESI-23" evi 2

===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
2             2             0                   yes 11/29/2022 09:42:41
===============================================================================

===============================================================================
DF Candidates                           Time Added           Oper Pref  Do Not
                                                               Value    Preempt
-------------------------------------------------------------------------------
No entries found
===============================================================================
```

Similarly, on PE-3:

```
*A:PE-3# show service system bgp-evpn ethernet-segment name "AA-ESI-23" evi 2

===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
2             2             0                   yes 11/29/2022 09:42:43
===============================================================================

===============================================================================
DF Candidates                           Time Added           Oper Pref  Do Not
                                                               Value    Preempt
-------------------------------------------------------------------------------
No entries found
===============================================================================
```

To confirm that all-active multi-homing is working correctly, the following command shows all information related to a specific ESI; in this case, AA-ESI-23 on PE-2:

```
*A:PE-2# show service system bgp-evpn ethernet-segment name "AA-ESI-23" all

===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : AA-ESI-23
Eth Seg Type            : None
Admin State             : Enabled            Oper State         : Up
ESI                     : 01:00:00:00:00:23:00:00:00:01
```

```
Oper ESI                : 01:00:00:00:00:23:00:00:00:01
Auto-ESI Type           : None
AC DF Capability        : Include
Multi-homing            : allActive          Oper Multi-homing  : allActive
ES SHG Label            : 524281
Source BMAC LSB         : None
Lag Id                  : 1
ES Activation Timer     : 3 secs
Oper Group              : (Not Specified)
Svc Carving             : auto               Oper Svc Carving   : auto
Cfg Range Type          : primary
Vprn NextHop EVI Ranges : <none>
===============================================================================

===============================================================================
EVI Information
===============================================================================
EVI                 SvcId               Actv Timer Rem      DF
-------------------------------------------------------------------------------
2                   2                   0                   yes
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
---snip---
```

EVPN for MPLS tunnels in Epipe services with single-active multi-homing

Single-active multi-homing allows for per-service load-balancing. Single-active multi-homing is configured on PE-4 and PE-5 with ES "SA-ESI-45". Both PEs have an SDP to MTU-6, which is associated with the ES and to the Epipe service. The configuration of the local and remote AC names and Ethernet tags is identical on PE-4 and PE-5.

On PE-4, the service configuration is as follows:

```
# on PE-4:
configure
    service
        sdp 46 mpls create
            far-end 192.0.2.6
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        system
            bgp-evpn
                ethernet-segment "SA-ESI-45" create
                    esi 01:00:00:00:00:45:00:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 46
                    no shutdown
                exit
            exit
        exit
        epipe 2 name "Epipe 2" customer 1 create
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-SA-ESI-45-MTU-6 create
                    eth-tag 456
                exit
                remote-attachment-circuit AC-AA-ESI-23-MTU-1 create
                    eth-tag 231
                exit
                evi 2
                mpls bgp 1
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            spoke-sdp 46:2 create
                no shutdown
            exit
            no shutdown
```

On PE-5, the configuration is similar, but with a different SDP:

```
# on PE-5:
configure
    service
        sdp 56 mpls create
            far-end 192.0.2.6
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        system
            bgp-evpn
                ethernet-segment "SA-ESI-45" create
                    esi 01:00:00:00:00:45:00:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 56
                    no shutdown
                exit
            exit
        exit
        epipe 2 name "Epipe 2" customer 1 create
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-SA-ESI-45-MTU-6 create
                    eth-tag 456
                exit
                remote-attachment-circuit AC-AA-ESI-23-MTU-1 create
                    eth-tag 231
                exit
                evi 2
                mpls bgp 1
                    ecmp 2
                    auto-bind-tunnel
```

```
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            spoke-sdp 56:2 create
                no shutdown
            exit
            no shutdown
        exit
```

Three route types will be exchanged between the core PEs: AD per-EVI, AD per-ES, and ES routes.

PE-4 and PE-5 advertise ES routes that are only imported by them. As an example, the following is the ES route with originator PE-4 sent by RR PE-2 to PE-5. It contains a target 00:00:00:00:45:00 in the extended community that is derived from the ESI:

```
# on PE-2:
64 2022/11/29 09:43:18.845 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 85
    Flag: 0x90 Type: 14 Len: 34 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.4
        Type: EVPN-ETH-SEG Len: 23 RD: 192.0.2.4:0 ESI: 01:00:00:00:00:45:00:00:00:01, IP-Len:
 4 Orig-IP-Addr: 192.0.2.4
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.4
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        192.0.2.2
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        df-election::DF-Type:Auto/DP:0/DF-Preference:0/AC:1
        target:00:00:00:00:45:00
"
```

The AD per-ES route has a maximum Ethernet tag (MAX-ET) and an ESI label in the extended community. The multi-homing mode is single-active. As in the case of all-active multi-homing, the ESI label is not used in Epipe services. The following BGP update with originator PE-5 is sent by RR PE-2 to its client PE-4:

```
# on PE-2:
67 2022/11/29 09:43:18.970 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 95
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.5
        Type: EVPN-AD Len: 25 RD: 192.0.2.5:2 ESI: 01:00:00:00:00:45:00:00:00:01, tag: MAX-ET
 Label: 0 (Raw Label: 0x0) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.5
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        192.0.2.2
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:2
        esi-label:524282/Single-Active
        bgp-tunnel-encap:MPLS
"
```

The AD per-EVI route contains flags for primary and backup, which will be different for routes received from PE-4 and PE-5. In this case, PE-4 is primary in the single-active multi-homing ES (P=1):

```
# on PE-2:
70 2022/11/29 09:43:21.801 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 95
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.4
        Type: EVPN-AD Len: 25 RD: 192.0.2.4:2 ESI: 01:00:00:00:00:45:00:00:00:01, tag:
 456 Label: 8388464 (Raw Label: 0x7fff70) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.4
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        192.0.2.2
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:2
        l2-attribute:MTU: 1514 C: 0 P: 1 B: 0
        bgp-tunnel-encap:MPLS
"
```

PE-5 is backup in the single-active multi-homing ES (B=1):

```
# on PE-2:
78 2022/11/29 09:43:25.369 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 81
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.5
        Type: EVPN-AD Len: 25 RD: 192.0.2.5:2 ESI: 01:00:00:00:00:45:00:00:00:01, tag:
 456 Label: 8388496 (Raw Label: 0x7fff90) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:2
        l2-attribute:MTU: 1514 C: 0 P: 0 B: 1
        bgp-tunnel-encap:MPLS
"
```

The BGP EVPN AD routes can be shown with the following command:

```
*A:PE-2# show router bgp routes evpn auto-disc esi 01:00:00:00:00:45:00:00:00:01
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
```

```
===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Flag  Route Dist.         ESI                           NextHop
      Tag                                               Label
-------------------------------------------------------------------------------
u*>i  192.0.2.4:2         01:00:00:00:00:45:00:00:00:01 192.0.2.4
      456                                               LABEL 524279

u*>i  192.0.2.4:2         01:00:00:00:00:45:00:00:00:01 192.0.2.4
      MAX-ET                                            LABEL 0

u*>i  192.0.2.5:2         01:00:00:00:00:45:00:00:00:01 192.0.2.5
      456                                               LABEL 524281

u*>i  192.0.2.5:2         01:00:00:00:00:45:00:00:00:01 192.0.2.5
      MAX-ET                                            LABEL 0

-------------------------------------------------------------------------------
Routes : 4
===============================================================================
```

For each PE in the single-active ES, there are two AD routes: the routes with MAX-ET are AD per-ES routes and the routes with a configured Ethernet tag are AD per-EVI routes.

The EVPN MPLS destination for Epipe 2 on PE-2 is SA-ESI-45, as shown in the following output:

```
*A:PE-2# show service id 2 evpn-mpls

===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address                          Egr Label              Last Change
                                      Transport:Tnl-id
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================

===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                                Last Change
-------------------------------------------------------------------------------
01:00:00:00:00:45:00:00:00:01            11/29/2022 09:43:22
-------------------------------------------------------------------------------
Number of entries: 1
-------------------------------------------------------------------------------
===============================================================================
```

The ESI is resolved to the TEP address of the primary (DF) PE-4, as follows:

```
*A:PE-2# show service id 2 evpn-mpls esi 01:00:00:00:00:45:00:00:00:01

===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                                Last Change
-------------------------------------------------------------------------------
01:00:00:00:00:45:00:00:00:01            11/29/2022 09:43:22
===============================================================================

===============================================================================
BGP EVPN-MPLS Dest TEP Info
===============================================================================
TEP Address              Egr Label                Last Change
                         Transport:Tnl-Id
-------------------------------------------------------------------------------
192.0.2.4                524279                   11/29/2022 09:43:22
                         ldp:65538
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

The DF election is key for the forwarding and backup functions in single-active multi-homing ESs. The PE elected as DF will be the primary for the ES in the Epipe and will unblock the SAP/spoke-SDP for upstream and downstream traffic. The rest of the PEs in the ES will bring their ES SAPs or spoke-SDPs operationally down.

PE-5 is a non-DF, as follows:

```
*A:PE-5# show service system bgp-evpn ethernet-segment name "SA-ESI-45" evi 2

===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
2             2             0                   no  11/29/2022 09:43:09
===============================================================================

===============================================================================
DF Candidates                           Time Added           Oper Pref  Do Not
                                                               Value    Preempt
-------------------------------------------------------------------------------
192.0.2.4                               11/29/2022 09:43:19  0          Disabl*
192.0.2.5                               11/29/2022 09:43:22  0          Disabl*
-------------------------------------------------------------------------------
Number of entries: 2
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

In single-active multi-homing, the service spoke-SDP (or SAP) is brought operationally down on the non-DF, as shown in the following output:

```
*A:PE-5# show service id 2 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
56:2             Spok     192.0.2.6       Up      Down      524280    524280
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

The spoke-SDP 56:2 is operationally down with a StandbyForMHProtocol flag:

```
*A:PE-5# show service id 2 sdp 56:2 detail | match Flag
Flags              : StandbyForMHProtocol
```

Two consecutive DF elections take place: the first DF election includes all PEs in the ES for that Epipe and determines which PE is the primary PE (flags P=1, B=0). The second DF election excludes this DF and determines which PE is the backup (P=0, B=1). All other PEs signal flags P=0 and B=0.

When the primary PE fails, AD per-ES/EVI withdrawal messages are sent to the remote PE, which will update its next hop to the backup. The backup PE takes over immediately without waiting for the **es-activation-timer** to bring up its SAP/spoke-SDP.

Ethernet segment failures

When the SDP toward the primary (DF) fails, the backup PE needs to take over. An SDP failure is emulated and log 99 on PE-4 shows that SDP 46 is operational down and PE-4 is no longer the DF:

```
140 2022/11/29 10:05:00.118 UTC MINOR: SVCMGR #2303 Base
"Status of SDP 46 changed to admin=up oper=down"

142 2022/11/29 10:05:00.119 UTC MINOR: SVCMGR #2094 Base
"Ethernet Segment:SA-ESI-45, EVI:2, Designated Forwarding state changed to:false"
```

Remote PEs receive route withdrawal updates (unreachable NLRI) from former DF PE-4, for example on PE-2:

```
# on PE-2:
82 2022/11/29 10:05:00.122 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 34
    Flag: 0x90 Type: 15 Len: 30 Multiprotocol Unreachable NLRI:
        Address Family EVPN
        Type: EVPN-AD Len: 25 RD: 192.0.2.4:2 ESI: 01:00:00:00:00:45:00:00:00:01, tag: MAX-ET
 Label: 0 (Raw Label: 0x0) PathId:
"

81 2022/11/29 10:05:00.122 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 59
    Flag: 0x90 Type: 15 Len: 55 Multiprotocol Unreachable NLRI:
        Address Family EVPN
        Type: EVPN-AD Len: 25 RD: 192.0.2.4:2 ESI: 01:00:00:00:00:45:00:00:00:01, tag: 456
 Label: 0 (Raw Label: 0x0) PathId:
        Type: EVPN-ETH-SEG Len: 23 RD: 192.0.2.4:0 ESI: 01:00:00:00:00:45:00:00:00:01, IP-Len:
 4 Orig-IP-Addr: 192.0.2.4
"
```

The backup PE-5 is promoted to primary (P=1, B=0) and sends BGP updates accordingly. The following AD per-EVI is received on PE-2:

```
# on PE-2:
85 2022/11/29 10:05:00.124 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 81
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.5
        Type: EVPN-AD Len: 25 RD: 192.0.2.5:2 ESI: 01:00:00:00:00:45:00:00:00:01, tag: 456
 Label: 8388496 (Raw Label: 0x7fff90) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:2
        l2-attribute:MTU: 1514 C: 0 P: 1 B: 0
        bgp-tunnel-encap:MPLS
"
```

PE-5 brings up its spoke-SDP without waiting for the **es-activation-timer** and takes over immediately. It is now the only DF candidate, and therefore the DF, as follows:

```
*A:PE-5# show service system bgp-evpn ethernet-segment name "SA-ESI-45" evi 2

===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
2             2             0                   yes 11/29/2022 09:43:09
===============================================================================

===============================================================================
DF Candidates                           Time Added           Oper Pref  Do Not
                                                               Value    Preempt
-------------------------------------------------------------------------------
192.0.2.5                               11/29/2022 09:43:22  0          Disabl*
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

BGP updates are exchanged and the remote PEs will resolve the ESI to the TEP address 192.0.2.5. For example, on PE-2:

```
*A:PE-2# show service id 2 evpn-mpls esi 01:00:00:00:00:45:00:00:00:01

===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                               Last Change
-------------------------------------------------------------------------------
01:00:00:00:00:45:00:00:00:01           11/29/2022 10:05:00
===============================================================================

===============================================================================
BGP EVPN-MPLS Dest TEP Info
===============================================================================
TEP Address              Egr Label                Last Change
                         Transport:Tnl-Id
-------------------------------------------------------------------------------
192.0.2.5                524281                   11/29/2022 10:05:00
                         ldp:65539
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

This process is revertive; as soon as the SDP 46 is operationally up again, a new DF election is triggered with two DF candidates and PE-4 will be elected as DF.

Troubleshooting and debugging

The following show and debug commands can be used in EVPN-VPWS:

- **show redundancy bgp-evpn-multi-homing**
- **show router bgp routes evpn** (and filters)
- **show service evpn-mpls [<TEP ip-address>]**
- **show service id bgp-evpn**
- **show service id evpn-mpls** (and modifiers)
- **show service system bgp-evpn**
- **show service system bgp-evpn ethernet-segment** (and modifiers)
- **debug router bgp update**
- **show log log-id 99**

Most of these commands have been shown in the preceding sections; some commands are shown in this section.

Information about the configured boot timers (before DF election) and ES activation timer (after the system has been elected DF) can be shown as follows:

```
*A:PE-2# show redundancy bgp-evpn-multi-homing

===============================================================================
Redundancy BGP EVPN Multi-homing Information
===============================================================================
Boot-Timer              : 10 secs
Boot-Timer Remaining    : 0 secs
ES Activation Timer     : 3 secs
===============================================================================
```

See chapter EVPN for MPLS Tunnels for a description of these timers.

The following command shows that the BGP route-type 4 (ES route) messages are only imported by the PEs in the same ES; for example, on PE-3:

```
*A:PE-3# show router bgp routes evpn eth-seg
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Eth-Seg Routes
===============================================================================
Flag  Route Dist.         ESI                           NextHop
      OrigAddr
-------------------------------------------------------------------------------
u*>i  192.0.2.2:0         01:00:00:00:00:23:00:00:00:01 192.0.2.2
```

```
      192.0.2.2

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

On PE-4:

```
*A:PE-4# show router bgp routes evpn eth-seg
===============================================================================
 BGP Router ID:192.0.2.4        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Eth-Seg Routes
===============================================================================
Flag  Route Dist.         ESI                           NextHop
      OrigAddr
-------------------------------------------------------------------------------
u*>i  192.0.2.5:0         01:00:00:00:00:45:00:00:00:01 192.0.2.5
      192.0.2.5

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

The following command shows all the EVPN MPLS destinations toward TEP 192.0.2.4. Epipe 1 has an EVPN MPLS destination toward TEP 192.0.2.4 directly and Epipe 2 has an EVPN MPLS destination to SA-ESI-45, which can be resolved to TEP 192.0.2.4. This is shown in the following output:

```
*A:PE-2# show service evpn-mpls 192.0.2.4

===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
Service Id                              Egr Label                     Instance
-------------------------------------------------------------------------------
1                                       524282                        1
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Service Id          Eth Seg Id                              Egr Label
-------------------------------------------------------------------------------
2                   01:00:00:00:00:45:00:00:00:01           524279
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-MPLS ES BMac Dest
===============================================================================
Service Id               ES BMac                  Egr Label
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

The following command lists all configured ESs on the system:

```
*A:PE-2# show service system bgp-evpn ethernet-segment

===============================================================================
Service Ethernet Segment
===============================================================================
Name                             ESI                           Admin     Oper
-------------------------------------------------------------------------------
AA-ESI-23                        01:00:00:00:00:23:00:00:00:01 Enabled   Up
-------------------------------------------------------------------------------
Entries found: 1
===============================================================================
```

In addition to the preceding commands, the following tools dump commands may be useful:

- **tools dump service evpn usage** - This command shows the number of EVPN-MPLS (and EVPN-VXLAN) destinations in the system.
- **tools dump service system bgp-evpn ethernet-segment** <*name*> **evi <..> df** - This command computes the DF election for a specific ESI and EVI. For all-active, there is no DF election and all PEs forward traffic. For single-active, one PE will be active for a service while another PE will be backup. This command shows the DF (primary), even if it is not the local PE.

The usage of EVPN resources can be shown as follows:

```
*A:PE-2# tools dump service evpn usage

vxlan-srv6-evpn-mpls usage statistics at 11/29/2022 10:13:11:

MPLS-TEP                                        :             1
VXLAN-TEP                                       :             0
SRV6-TEP                                        :             0
Total-TEP                                       :      1/ 16383

Mpls Dests (TEP, Egress Label + ES + ES-BMAC)   :             2
Mpls Etree Leaf Dests                           :             0
Vxlan Dests (TEP, Egress VNI + ES)              :             0
Srv6 Dests (TEP, SID  + ES)                     :             0
Total-Dest                                      :      2/196607

Sdp Bind +  Evpn Dests                          :      2/245759
ES L2/L3 PBR                                    :      0/ 32767
Evpn Etree Remote BUM Leaf Labels               :             0
```

On PE-2, there is one MPLS-TEP (192.0.2.4 in Epipe 1 and Epipe 2) and there are two MPLS destinations: 192.0.2.4 and ESI 01:00:00:00:00:45:00:00:00:01. PE-5 is not an MPLS-TEP for PE-2, because it is not a primary and, therefore, not forwarding any traffic.

In all-active multi-homing, the DF election is not applicable:

```
*A:PE-2# tools dump service system bgp-evpn ethernet-segment "AA-ESI-23" evi 2 df

[11/29/2022 10:13:29] All Active VPWS or IP-ALIASING - DF N/A
```

In single-active multi-homing, the following command shows which PE is the DF and which PE is the backup:

```
*A:PE-5# tools dump service system bgp-evpn ethernet-segment "SA-ESI-45" evi 2 df

[11/29/2022 10:13:49] Computed DF: 192.0.2.4 (Remote) (Boot Timer Expired: Yes)
[11/29/2022 10:13:49] Computed Backup: 192.0.2.5 (This Node)
```

The command is launched on PE-5, which is a backup. The computed DF is PE-4 and the boot timer has expired, meaning there is no DF re-election pending.

Conclusion

EVPN-VPWS is a simplified point-to-point version of RFC 7432, *BGP MPLS-Based Ethernet VPN*. When used for Epipe and VPLS services, EVPN provides a unified control plane mechanism that simplifies the network deployment and operation. Single-active and all-active multi-homing can be used in Epipes; EVPN-VPWS is a differentiator of EVPN compared to traditional TLDP or BGP Epipe redundancy mechanisms. The Ethernet segments used for multi-homing can be shared between EVPN VPLS and EVPN Epipes.

EVPN for MPLS Tunnels in Routed VPLS

This chapter provides information about EVPN for MPLS tunnels in routed VPLS.

Topics in this chapter include:

Applicability

This chapter was initially written for SR OS Release 15.0.R4, but the CLI in the current edition is based on SR OS Release 21.10.R3. EVPN-MPLS and IP-prefix advertisement in routed VPLS (R-VPLS) without Multi-homing (MH) is supported in SR OS Release 14.0.R1, and later. EVPN-MPLS and IP-prefix advertisement in R-VPLS with all-active and single-active MH is supported in SR OS Release 14.0.R4, and later. Virtual Router Redundancy Protocol (VRRP) in passive mode is also supported in SR OS Release 14.0.R4, and later.

Chapter EVPN for VXLAN Tunnels (Layer 3) is prerequisite reading.

Overview

The EVPN-MPLS in R-VPLS feature matches the EVPN-VXLAN in R-VPLS feature, which is described in chapter EVPN for VXLAN Tunnels (Layer 3). The following capabilities are supported in an R-VPLS service where **bgp-evpn mpls** is enabled:

- R-VPLS with Virtual Router Redundancy Protocol (VRRP) support on the VPRN interfaces
- R-VPLS support including **ip-route-advertisement** (IP prefix routes—BGP-EVPN route type 5) with regular interfaces
- R-VPLS support including **ip-route-advertisement** with **evpn-tunnel** interfaces
- R-VPLS with IPv6 support on the VPRN IP interface

All-active and single-active MH Ethernet segments (ESs) are supported in R-VPLS. When Ethernet Segments (ESs) are used along with R-VPLS services in two or more PEs, passive VRRP provides an "anycast default gateway" that optimizes inter-subnet forwarding for hosts in the R-VPLS. Passive VRRP is described in the following section.

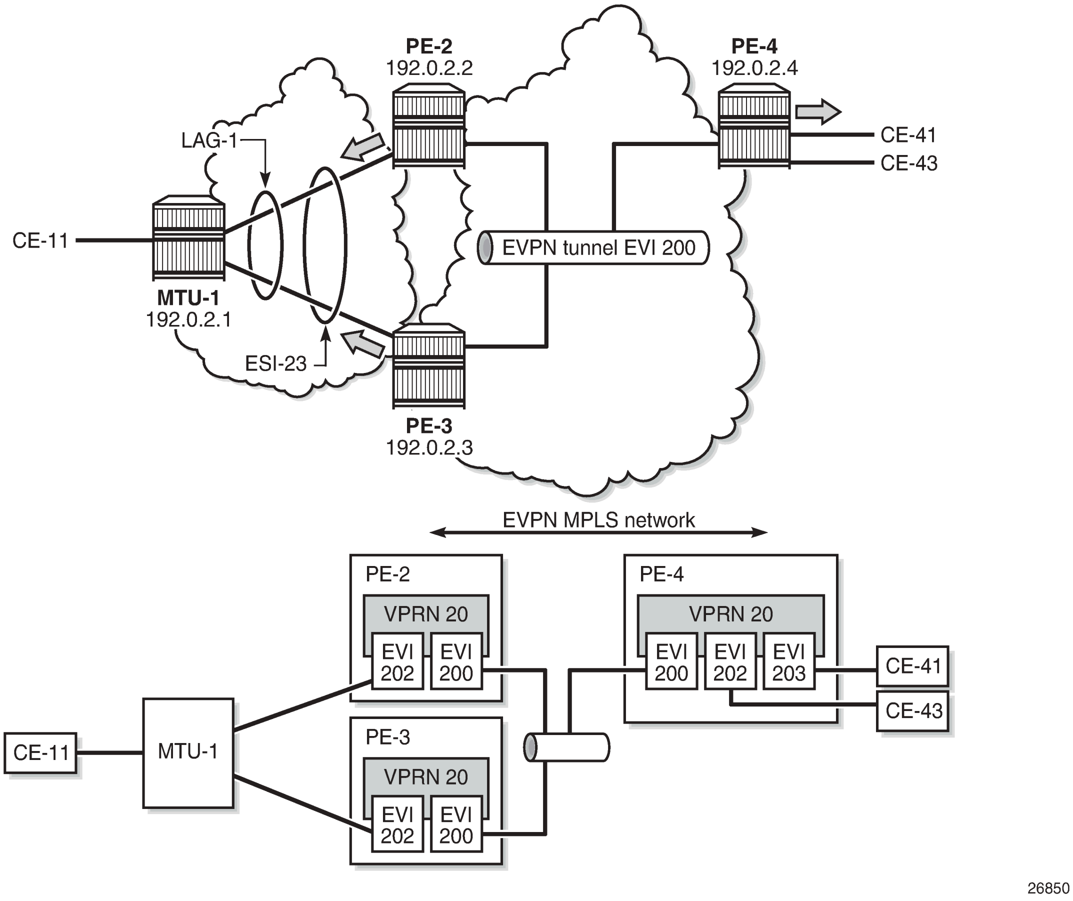
Passive VRRP

VRRP can be configured in passive mode, which suppresses the transmission and reception of keepalive messages. Passive mode can be enabled by adding the keyword **passive** at creation time. Passive mode cannot be enabled or disabled on the fly. Passive VRRP can be configured in the base router, in an IES, or in a VPRN, using the following commands:

```
*A:PE-2# tree flat detail | match vrrp | match passive
configure router interface ipv6 vrrp <virtual-router-id> [owner] [passive]
configure router interface vrrp <virtual-router-id> [owner] [passive]
configure service ies interface ipv6 vrrp <virtual-router-id> [owner] [passive]
configure service ies interface vrrp <virtual-router-id> [owner] [passive]
configure service vprn interface ipv6 vrrp <virtual-router-id> [owner] [passive]
configure service vprn interface vrrp <virtual-router-id> [owner] [passive]
```

All PEs configured with passive VRRP become VRRP master and take ownership of the virtual IP and MAC addresses. Figure 69: Passive VRRP - vMAC/vIP advertised by GARP shows the use of passive VRRP where the VRID and default gateway (GW) are identical for all nodes, and therefore, the vMAC/vIP are identical. Each PE sends Gratuitous Address Resolution Protocol (GARP) messages with the same vMAC/vIP.

Figure 69: Passive VRRP - vMAC/vIP advertised by GARP



Ethernet VPN instance (EVI) 202 is configured on all PEs as an R-VPLS with passive VRRP. Each individual R-VPLS interface has a unique MAC/IP, but they all have the same vMAC/vIP because they share the same VRID and backup IP address. The vMAC address is auto-derived out of 00:00:5e:00:00:<VRID>, as per RFC 3768.

The behavior is as follows:

- PEs advertise their real MAC/IP and their vMAC/vIP in EVPN for EVI 202.
- All hosts in EVI 202 have a unique configured default GW.
- When a CE sends upstream traffic to a remote subnet, the packets are routed by the closest PE because the vMAC address is local on each PE.
- In case of ES failure, or in case of single-active MH if the traffic arrives at the non-Designated Forwarder (NDF) PE, the traffic will not be discarded at the peer ES PE. Virtual MAC addresses bypass the R-VPLS interface protection, so traffic can be forwarded between the PEs without being dropped. Note that if passive VRRP was not used in this case and the same regular interface anycast MAC/IP was used instead, the peer PE would discard the traffic due to the MAC Source Address (SA).

Passive VRRP provides an efficient anycast default gateway solution, with the following advantages compared to regular VRRP:

- No need for multiple VRRP instances to achieve default GW load-balancing. Only one VRRP instance is in the R-VPLS, so only one default GW is needed for all hosts.
- Fast convergence because all the nodes in the VRID are master.
- Better scalability because there is no need for keepalive messages or BFD to detect failures.

Passive VRRP provides the following advantages compared to using the same anycast MAC/IP in all the Integrated Routing Bridging (IRB) interfaces:

- VRRP vMAC SA bypasses the protection in the receiving R-VPLS service; therefore, frames with MAC SA matching the local vMAC address are not discarded, and VRRP vMAC SAs can be used in combination with EVPN multi-homing.
- PEs will not show traps claiming duplicate IP addresses.
- vMAC addresses are auto-derived from the VRID, so no need to configure the same MAC address in all the IRB interfaces.
- PEs can still use their real (unique) IRB IP addresses when sending ICMP packets for troubleshooting purposes.

Configuration

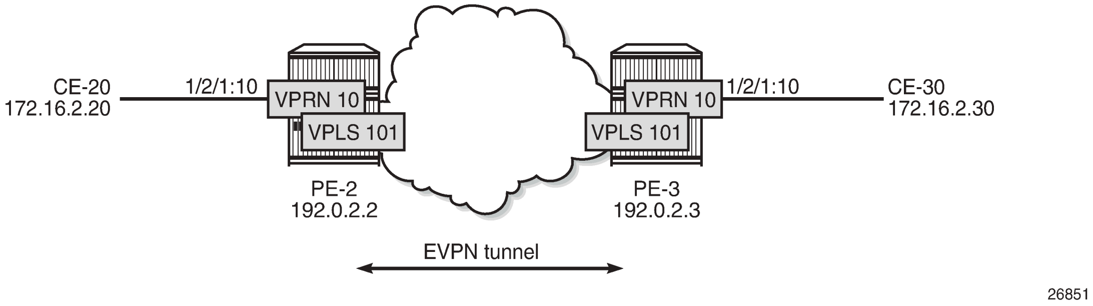
In this section, the following use cases are described:

- EVPN-MPLS R-VPLS without multi-homing
- EVPN-MPLS R-VPLS with all-active multi-homing ES
- EVPN-MPLS R-VPLS with single-active multi-homing ES

EVPN-MPLS R-VPLS without multi-homing

The first scenario describes R-VPLS support including IP route advertisement (BGP-EVPN route type 5) with EVPN tunnel interfaces, without multi-homing. VPLS 101 does not have any connected host, but the linked VPRN has SAP 1/2/1:10. Figure 70: R-VPLS with EVPN tunnel, without multi-homing shows the example topology used for R-VPLS with EVPN tunnel but without multi-homing. IP prefixes are advertised.

Figure 70: R-VPLS with EVPN tunnel, without multi-homing



The initial configuration includes the following:

- Cards, MDAs, ports
- Router interface between PE-2 and PE-3
- IS-IS (or OSPF)
- LDP enabled on the router interface between PE-2 and PE-3

BGP is configured for address family EVPN on PE-2 and PE-3. The BGP configuration on PE-2 is as follows. The BGP configuration on PE-3 is similar.

```
# on PE-2:
configure
    router Base
        autonomous-system 64500
        bgp
            family evpn
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            rapid-update evpn
            group "internal"
                peer-as 64500
                neighbor 192.0.2.3
                exit
            exit
        exit
```

The CEs are connected to SAP 1/2/1:10 in VPRN 10. R-VPLS 101 is bound to VPRN 10 and VPRN 10 has a dedicated interface "int-evi-101" for the EVPN tunnel. In general, if only one route-target (RT) is used for import and export in the EVPN-VPLS, it is good to add the EVI and have the route distinguisher (RD) and RT auto-derived from the EVI. It is simpler and avoids configuration mistakes. The service configuration on PE-2 is as follows:

```
# on PE-2:
configure
    service
        vprn 10 name "VPRN 10" customer 1 create
            interface "int-PE-2-CE-20" create
                address 172.16.2.1/24
                sap 1/2/1:10 create
                exit
            exit
            interface "int-evi-101" create
                vpls "evi-101"
                    evpn-tunnel
                exit
            exit
            no shutdown
        exit
        vpls 101 name "evi-101" customer 1 create
            allow-ip-int-bind
            exit
            bgp            # RD and RT are not manually configured in BGP context
            exit
            bgp-evpn
                ip-route-advertisement
                evi 101    # RD and RT will be auto-derived from the EVI
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

- The **allow-ip-int-binding** command is required so that R-VPLS 101 can be bound to VPRN 10.
- The service name is required and the configured name "evi-101" must match the name in the VPRN 10 VPLS interface. The service name is configured at service creation time.
- The VPRN 10 VPLS interface is configured with the keyword **evpn-tunnel**. This configuration has the advantage of not having to allocate IP addresses to the R-VPLS interfaces, however, it cannot be used when the R-VPLS has local SAPs.

The configuration is similar on PE-3. It is important that the RD is different on PE-2 and PE-3, but it is automatically the case when the RD is auto-derived from the configured EVI, as in the example. The RD on PE-2 is 192.0.2.2:101; on PE-3, the RD is 192.0.2.3:101.

PE-3 receives the following BGP-EVPN IP prefix route for prefix 172.16.2.0/24 from PE-2:

```
2 2022/02/24 11:00:28.145 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 90
    Flag: 0x90 Type: 14 Len: 45 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-IP-PREFIX Len: 34 RD: 192.0.2.2:101, tag: 0,
           ip_prefix: 172.16.2.0/24 gw_ip 0.0.0.0 Label: 8388496 (Raw Label: 0x7fff90)
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:101
        mac-nh:02:13:ff:ff:ff:a2
        bgp-tunnel-encap:MPLS
"
```

GW IP 0.0.0.0 is an indication that an EVPN tunnel is in use. With EVPN tunnels, no IRB IP address needs to be configured in the VPRN. EVPN tunnels make provisioning easier to automate and save IP addresses from the tenant IP space.

The BGP tunnel encapsulation is MPLS, but the MPLS label in the debug message is not the same as in the service, because the router will strip the extra four lowest bits to get the 20-bit MPLS label. In the debug message, the label is 8388496. This is because the debug message is shown before the router can parse the label field and see if it corresponds to an MPLS label (20 bits) or a VXLAN VNI (24 bits). The MPLS label is calculated by dividing the label value by 24 (16), as follows: 8388496/16 = 524281.

The MAC next-hop extended community 02:13:ff:ff:ff:a2 is the MAC address of the interface "int-evi-101" in VPRN 10 on PE-2, as follows:

```
*A:PE-2# show router 10 interface "int-evi-101" detail | match "MAC Address"
MAC Address      : 02:13:ff:ff:ff:a2    Mac Accounting    : Disabled
```

The routing table for VPRN 10 on PE-3 contains the route for prefix 172.16.2.0/24 as the EVPN-IFF (IFF stands for Interface-ful) route with next-hop "int-evi-101" and interface name "ET-02:13:ff:ff:ff:a2" (ET stands for EVPN Tunnel), as follows:

```
*A:PE-3# show router 10 route-table

===============================================================================
Route Table (Service: 10)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age         Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
172.16.2.0/24                                 Remote  EVPN-IFF  01h43m58s  169
       int-evi-101 (ET-02:13:ff:ff:ff:a2)                           0
172.16.3.0/24                                 Local   Local     01h43m59s  0
       int-PE-3-CE-30                                               0
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The forwarding database (FDB) for VPLS 101 on PE-3 shows an entry for MAC address 02:13:ff:ff:ff:a2 that is learned via EVPN. The MAC address is static (S) and protected (P). The MPLS label is 524281.

```
*A:PE-3# show service id 101 fdb detail

===============================================================================
Forwarding Database, Service 101
===============================================================================
ServId     MAC               Source-Identifier        Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
101        02:13:ff:ff:ff:a2 mpls-1:                  EvpnS:P  02/24/22 11:00:35
                             192.0.2.2:524281
           ldp:65538
101        02:17:ff:ff:ff:a2 cpm                      Intf     02/24/22 11:00:34
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

When the CEs have IPv6 addresses, the VPRN configuration is similar on the PEs, but the **ipv6** context must be enabled in the EVPN tunnel interface, so that the router can advertise and process BGP-EVPN routes type 5 with IPv6 prefixes. The configuration of the VPLS is identical for IPv4 and IPv6.

```
# on PE-2:
configure
    service
        vprn 16 name "VPRN 16" customer 1 create
            interface "int-PE-2-CE-26" create
                ipv6
                    address 2001:db8:16::2:1/120
                exit
                sap 1/2/1:16 create
                exit
            exit
            interface "int-evi-106" create
                ipv6
                exit
                vpls "evi-106"
                    evpn-tunnel
                exit
            exit
            no shutdown
        exit
        vpls 106 name "evi-106" customer 1 create
            allow-ip-int-bind
            exit
            bgp
            exit
            bgp-evpn
                ip-route-advertisement
                evi 106
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

When advertising IPv6 prefixes, the GW IP field in the route type 5 is always populated with the IPv6 address of the R-VPLS interface. In this example, because no specific IPv6 global address is configured, the GW IP will be populated with the auto-created link local address. The following BGP update is received by PE-3 for IPv6 prefix 2001:db8:16::2:0/120:

```
# on PE-3:
9 2022/02/24 11:00:35.338 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 106
    Flag: 0x90 Type: 14 Len: 69 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-IP-PREFIX Len: 58 RD: 192.0.2.2:106, tag: 0,
               ip_prefix: 2001:db8:16::2:0/120 gw_ip fe80::14:1ff:fe02:1
               Label: 8388480 (Raw Label: 0x7fff80)
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:106
        bgp-tunnel-encap:MPLS
"
```

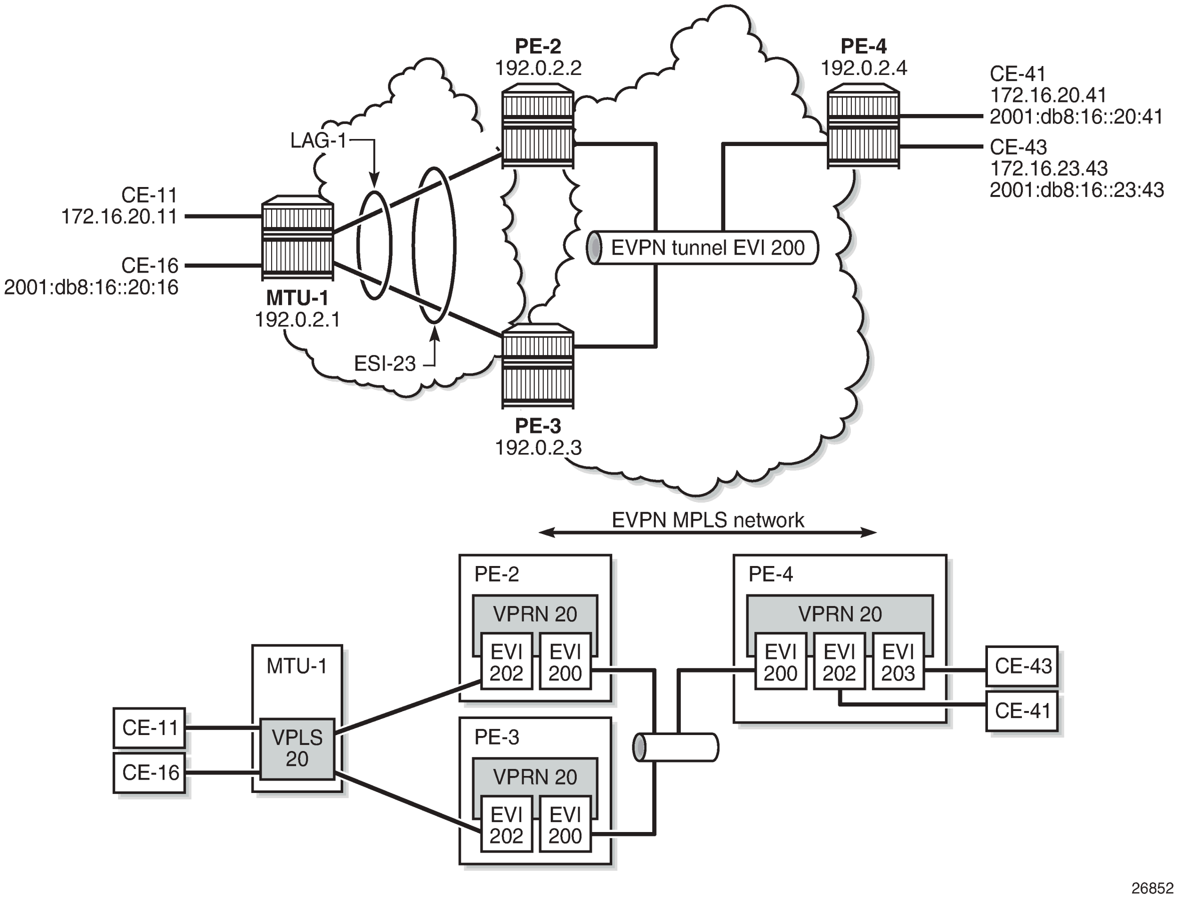
The IPv6 route-table on PE-3 is as follows:

```
*A:PE-3# show router 16 route-table ipv6

===============================================================================
IPv6 Route Table (Service: 16)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
2001:db8:16::2:0/120                          Remote  EVPN-IFF  01h50m01s  169
       fe80::14:1ff:fe02:1-"int-evi-106"                            0
2001:db8:16::3:0/120                          Local   Local     01h50m01s  0
       int-PE-3-CE-36                                               0
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

EVPN-MPLS R-VPLS with all-active MH

Figure 71: EVPN-MPLS R-VPLS with all-active MH ES shows the example topology with all-active multi-homing ES "AA-ESI-23".

Figure 71: EVPN-MPLS R-VPLS with all-active MH ES



BGP is configured between PE-2, PE-3, and PE-4 for address family EVPN. The configuration on PE-2 is as follows:

```
# on PE-2:
configure
    router Base
        autonomous-system 64500
        bgp
            family evpn
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            rapid-update evpn
            group "internal"
                peer-as 64500
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
            exit
        exit
```

All-active multi-homing Ethernet segment "AA-ESI-23" is configured on PE-2 and PE-3, as follows:

```
# on PE-2, PE-3:
configure
    service
        system
            bgp-evpn
                ethernet-segment "AA-ESI-23" create
                    esi 01:00:00:00:00:23:00:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 1
                    no shutdown
                exit
```

The following services are configured on the PEs:

- VPRN 20 has interfaces bound to VPLS 200 and VPLS 202. On PE-4, VPRN 20 also has an interface bound to VPLS 203.
- VPLS 200 is configured as an EVPN tunnel that connects the PEs.
- VPLS 202 and VPLS 203 have attachment circuits to CEs.

The services are configured on PE-2 as follows. The configuration on PE-3 and PE-4 is similar.

```
# on PE-2:
configure
    service
        vprn 20 name "VPRN 20" customer 1 create
            interface "int-evi-202" create
                address 172.16.20.2/24
                mac 00:ca:fe:00:02:02
                vrrp 1 passive
                    backup 172.16.20.254
                    ping-reply
                    traceroute-reply
                exit
                ipv6
                    address 2001:db8:16::20:2/120
                    link-local-address fe80::16:20:2 dad-disable
                    vrrp 1 passive
                        backup fe80::16:20:fe
                        ping-reply
                        traceroute-reply
                    exit
                exit
                vpls "evi-202"
                exit
            exit
            interface "int-evi-200" create
                ipv6
                exit
                vpls "evi-200"
                    evpn-tunnel
                exit
            exit
            router-advertisement
                interface "int-evi-202"
                    use-virtual-mac
                    no shutdown
                exit
            exit
            no shutdown
        exit
        vpls 200 name "evi-200" customer 1 create
            allow-ip-int-bind
            exit
            bgp
            exit
            bgp-evpn
                ip-route-advertisement
                evi 200
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            no shutdown
        exit
        vpls 202 name "evi-202" customer 1 create
            allow-ip-int-bind
            exit
            bgp
            exit
            bgp-evpn
                evi 202
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap lag-1:20 create
            exit
            no shutdown
        exit
```

The IPv6 VRRP backup address is in the same subnet as the link local address of the interface "int-evi-202". The option **dad-disable** is configured on the link local address to disable Duplicate Address Detection (DAD) and set the IPv6 address as preferred. Also for IPv6, router advertisement must be enabled and configured to use the virtual MAC address.

Passive VRRP

EVI 202 is configured as an R-VPLS with passive VRRP. A passive-VRRP VRID instance suppresses the transmission and reception of keepalive messages. All PEs configured with passive VRRP become VRRP master and take ownership of the virtual IP and MAC address.

Each individual R-VPLS interface has a different MAC/IP on each PE. The MAC/IPs for "int-evi-202" on PE-2 are MAC 00:ca:fe:00:02:02 and IP 172.16.20.2/24 for IPv4 and the same MAC address with IPv6 2001:db8:16::20:2 and fe80::16:20:2. However, the R-VPLS interfaces on all PEs share the same VRID 1 and backup IP address 172.16.20.254, so the same vMAC/vIP 00:00:5e:00:01:01/172.16.20.254 and vMAC/vIP 00:00:5e:00:02:01/ fe80::16:20:fe are advertised by all PEs. PE-2 advertises the following EVPN MAC routes:

```
83 2022/02/24 15:09:15.841 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 285
    Flag: 0x90 Type: 14 Len: 240 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-MAC Len: 49 RD: 192.0.2.2:202 ESI: ESI-0, tag: 0, mac len: 48
            mac: 00:00:5e:00:02:01, IP len: 16, IP: fe80::16:20:fe, label1: 8388416
        Type: EVPN-MAC Len: 37 RD: 192.0.2.2:202 ESI: ESI-0, tag: 0, mac len: 48
            mac: 00:00:5e:00:01:01, IP len: 4, IP: 172.16.20.254, label1: 8388416
        Type: EVPN-MAC Len: 49 RD: 192.0.2.2:202 ESI: ESI-0, tag: 0, mac len: 48
            mac: 00:ca:fe:00:02:02, IP len: 16, IP: fe80::16:20:2, label1: 8388416
        Type: EVPN-MAC Len: 49 RD: 192.0.2.2:202 ESI: ESI-0, tag: 0, mac len: 48
            mac: 00:ca:fe:00:02:02, IP len: 16, IP: 2001:db8:16::20:2, label1: 8388416
        Type: EVPN-MAC Len: 37 RD: 192.0.2.2:202 ESI: ESI-0, tag: 0, mac len: 48
            mac: 00:ca:fe:00:02:02, IP len: 4, IP: 172.16.20.2, label1: 8388416
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:202
        bgp-tunnel-encap:MPLS
        mac-mobility:Seq:0/Static
"
```

The three PEs advertise the same (anycast) vMAC/vIP in EVI 202 as protected, but each PE keeps its own MAC entry in the FDB. The following FDB shows that the source identifier for vMAC 00:00:5e:00:01:01 and vMAC 00:00:5e:00:02:01 is the CPM. These two vMAC entries with source identifier CPM are seen on all PEs.

```
*A:PE-2# show service id 202 fdb detail

===============================================================================
Forwarding Database, Service 202
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
202        00:00:01:00:00:11 sap:lag-1:20            L/0      02/24/22 15:09:21
202        00:00:01:00:00:16 sap:lag-1:20            L/0      02/24/22 15:09:22
202        00:00:04:00:00:41 mpls-1:                 Evpn     02/24/22 15:09:14
                             192.0.2.4:524281
           ldp:65539
202        00:00:5e:00:01:01 cpm                     Intf     02/24/22 15:08:50
202        00:00:5e:00:02:01 cpm                     Intf     02/24/22 15:08:50
202        00:ca:fe:00:02:02 cpm                     Intf     02/24/22 15:08:50
202        00:ca:fe:00:02:03 mpls-1:                 EvpnS:P  02/24/22 15:09:03
                             192.0.2.3:524276
           ldp:65538
202        00:ca:fe:00:02:04 mpls-1:                 EvpnS:P  02/24/22 15:09:14
                             192.0.2.4:524281
           ldp:65539
-------------------------------------------------------------------------------
No. of MAC Entries: 8
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
```

```
===============================================================================
```

The interface MAC 00:ca:fe:00:02:02 is local, so it also has the CPM as source identifier. MAC 00:ca:fe:00:02:03 is the PE-3's R-VPLS interface MAC and it is learned via EVPN-MPLS (mpls-1) as static (S) and protected (P). MAC address 00:ca:fe:00:02:04 on PE-4 is also static and protected.

PE-4 sends the following IP prefix route (BGP-EVPN route type 5) for prefix 172.16.23.0/24 to the other PEs:

```
37 2022/02/24 15:09:13.665 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 90
    Flag: 0x90 Type: 14 Len: 45 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.4
        Type: EVPN-IP-PREFIX Len: 34 RD: 192.0.2.4:200, tag: 0,
                             ip_prefix: 172.16.23.0/24 gw_ip 0.0.0.0
                             Label: 8388512 (Raw Label: 0x7fffa0)
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:200
        mac-nh:02:1b:ff:00:00:05
        bgp-tunnel-encap:MPLS
"
```

The IP prefixes are advertised with next-hop equal to the EVPN-tunnel GW MAC "int-evi-200", as follows:

```
*A:PE-4# show router 20 interface "int-evi-200" detail | match "MAC Address"
MAC Address      : 02:1b:ff:00:00:05    Mac Accounting    : Disabled
```

The routing table for VPRN 20 on PE-2 contains IP-prefix 172.16.23.0/24 with next-hop 02:1b:ff:00:00:05, as follows:

```
*A:PE-2# show router 20 route-table

===============================================================================
Route Table (Service: 20)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
172.16.20.0/24                                Local   Local     00h17m12s  0
       int-evi-202                                                  0
172.16.23.0/24                                Remote  EVPN-IFF  00h16m48s  169
       int-evi-200 (ET-02:1b:ff:00:00:05)                           0
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The following IPv6 routing table for VPRN 20 on PE-2 contains prefix 2001:db8:16::23:0/120, which has also been advertised by PE-4. The next-hop is again "int-evi-200", only this time the link local ipv6 address is displayed (GW IP) instead of the MAC address. The next-hop is the GW IP value in the route type 5, as long as it is non-zero. When the GW IP address is zero, the route type 5 is expected to contain a mac-nh extended community. The MAC encoded in the extended community is used as next-hop in that case.

```
*A:PE-2# show router 20 route-table ipv6

===============================================================================
IPv6 Route Table (Service: 20)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
2001:db8:16::20:0/120                         Local   Local     00h17m10s  0
       int-evi-202                                                  0
2001:db8:16::23:0/120                         Remote  EVPN-IFF  00h16m46s  169
       fe80::a5:9124:c1ed:83ce-"int-evi-200"                        0
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The EVPN tunnel service VPLS 200 has all the MAC addresses of the EVPN interfaces within VPRN 20 as static (S) and protected (P), as follows:

```
*A:PE-2# show service id "evi-200" fdb detail

===============================================================================
Forwarding Database, Service 200
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
200        02:13:ff:00:00:05 cpm                     Intf     02/24/22 15:08:50
200        02:17:ff:00:00:05 mpls-1:                 EvpnS:P  02/24/22 15:09:03
                             192.0.2.3:524277
           ldp:65538
200        02:1b:ff:00:00:05 mpls-1:                 EvpnS:P  02/24/22 15:09:14
                             192.0.2.4:524282
           ldp:65539
-------------------------------------------------------------------------------
No. of MAC Entries: 3
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The VRRP instance in each PE is master, as follows:

```
*A:PE-2# show router 20 vrrp instance

===============================================================================
VRRP Instances
===============================================================================
Interface Name                   VR Id Own Adm  State       Base Pri   Msg Int
                                 IP        Opr  Pol Id      InUse Pri  Inh Int
-------------------------------------------------------------------------------
int-evi-202                      1     No  Up   Master       100       1
                                 IPv4      Up   n/a         100        No
  Backup Addr: 172.16.20.254
int-evi-202                      1     No  Up   Master       100       1
                                 IPv6      Up   n/a         100        Yes
  Backup Addr: fe80::16:20:fe
-------------------------------------------------------------------------------
Instances : 2
===============================================================================
```

```
*A:PE-3# show router 20 vrrp instance

===============================================================================
VRRP Instances
===============================================================================
Interface Name                   VR Id Own Adm  State       Base Pri   Msg Int
                                 IP        Opr  Pol Id      InUse Pri  Inh Int
-------------------------------------------------------------------------------
int-evi-202                      1     No  Up   Master       100       1
                                 IPv4      Up   n/a         100        No
  Backup Addr: 172.16.20.254
int-evi-202                      1     No  Up   Master       100       1
                                 IPv6      Up   n/a         100        Yes
  Backup Addr: fe80::16:20:fe
-------------------------------------------------------------------------------
Instances : 2
===============================================================================
```

```
*A:PE-4# show router 20 vrrp instance

===============================================================================
VRRP Instances
===============================================================================
Interface Name                   VR Id Own Adm  State       Base Pri   Msg Int
                                 IP        Opr  Pol Id      InUse Pri  Inh Int
-------------------------------------------------------------------------------
int-evi-202                      1     No  Up   Master       100       1
                                 IPv4      Up   n/a         100        No
  Backup Addr: 172.16.20.254
int-evi-203                      2     No  Up   Master       100       1
                                 IPv4      Up   n/a         100        No
  Backup Addr: 172.16.23.254
int-evi-202                      1     No  Up   Master       100       1
                                 IPv6      Up   n/a         100        Yes
  Backup Addr: fe80::16:20:fe
int-evi-203                      2     No  Up   Master       100       1
                                 IPv6      Up   n/a         100        Yes
  Backup Addr: fe80::16:23:fe
-------------------------------------------------------------------------------
Instances : 4
===============================================================================
```

Operation

On PE-4, VPRN 20 has one interface bound to VPLS 202 and another interface bound to VPLS 203. CE-41 is attached to VPLS 202, whereas CE-43 is attached to VPLS 203. When ping messages are sent from CE-41 to CE-43, or vice versa, the messages go via VPRN 20, which has routes to both CEs, as follows:

```
*A:PE-4# show router 20 route-table

===============================================================================
Route Table (Service: 20)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
172.16.20.0/24                                Local   Local     00h19m37s  0
       int-evi-202                                                  0
172.16.23.0/24                                Local   Local     00h19m37s  0
       int-evi-203                                                  0
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

```
*A:PE-4# show router 20 route-table ipv6

===============================================================================
IPv6 Route Table (Service: 20)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
2001:db8:16::20:0/120                         Local   Local     00h19m36s  0
       int-evi-202                                                  0
2001:db8:16::23:0/120                         Local   Local     00h19m36s  0
       int-evi-203                                                  0
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

When traffic is sent between CE-11 and CE-41, which are both associated with VPLS 202, the forwarding is done by the VPLS and not via the VPRN. The FDB for VPLS 202 on PE-3 is as follows:

```
*A:PE-3# show service id 202 fdb detail

===============================================================================
Forwarding Database, Service 202
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
202        00:00:01:00:00:11 sap:lag-1:20            L/0      02/24/22 15:28:41
202        00:00:01:00:00:16 sap:lag-1:20            L/0      02/24/22 15:28:45
202        00:00:04:00:00:41 mpls-1:                 Evpn     02/24/22 15:28:40
                             192.0.2.4:524281
           ldp:65539
202        00:00:5e:00:01:01 cpm                     Intf     02/24/22 15:09:03
202        00:00:5e:00:02:01 cpm                     Intf     02/24/22 15:09:03
202        00:ca:fe:00:02:02 mpls-1:                 EvpnS:P  02/24/22 15:09:04
                             192.0.2.2:524276
           ldp:65538
202        00:ca:fe:00:02:03 cpm                     Intf     02/24/22 15:09:03
202        00:ca:fe:00:02:04 mpls-1:                 EvpnS:P  02/24/22 15:09:14
                             192.0.2.4:524281
           ldp:65539
-------------------------------------------------------------------------------
No. of MAC Entries: 8
```

```
--------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
================================================================================
```

MAC 00:00:01:00:00:11 corresponds to CE-11 and is learned on SAP lag-1:20 on PE-3 and advertised via an EVPN MAC route to the BGP peers. MAC 00:00:04:00:00:41 corresponds to CE-41 and was advertised via an EVPN MAC route from PE-4, where the MAC was learned on SAP 1/2/1:41 of VPLS 202, as shown in the following FDB:

```
*A:PE-4# show service id 202 fdb detail

================================================================================
Forwarding Database, Service 202
================================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                          Age
--------------------------------------------------------------------------------
202        00:00:01:00:00:11 eES:                    Evpn     02/24/22 15:28:41
                             01:00:00:00:00:23:00:00:00:01
202        00:00:01:00:00:16 eES:                    Evpn     02/24/22 15:28:45
                             01:00:00:00:00:23:00:00:00:01
202        00:00:04:00:00:41 sap:1/2/1:41            L/90     02/24/22 15:28:40
202        00:00:5e:00:01:01 cpm                     Intf     02/24/22 15:09:14
202        00:00:5e:00:02:01 cpm                     Intf     02/24/22 15:09:14
202        00:ca:fe:00:02:02 mpls-1:                 EvpnS:P  02/24/22 15:09:16
                             192.0.2.2:524276
           ldp:65538
202        00:ca:fe:00:02:03 mpls-1:                 EvpnS:P  02/24/22 15:09:16
                             192.0.2.3:524276
           ldp:65539
202        00:ca:fe:00:02:04 cpm                     Intf     02/24/22 15:09:14
--------------------------------------------------------------------------------
No. of MAC Entries: 8
--------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
================================================================================
```

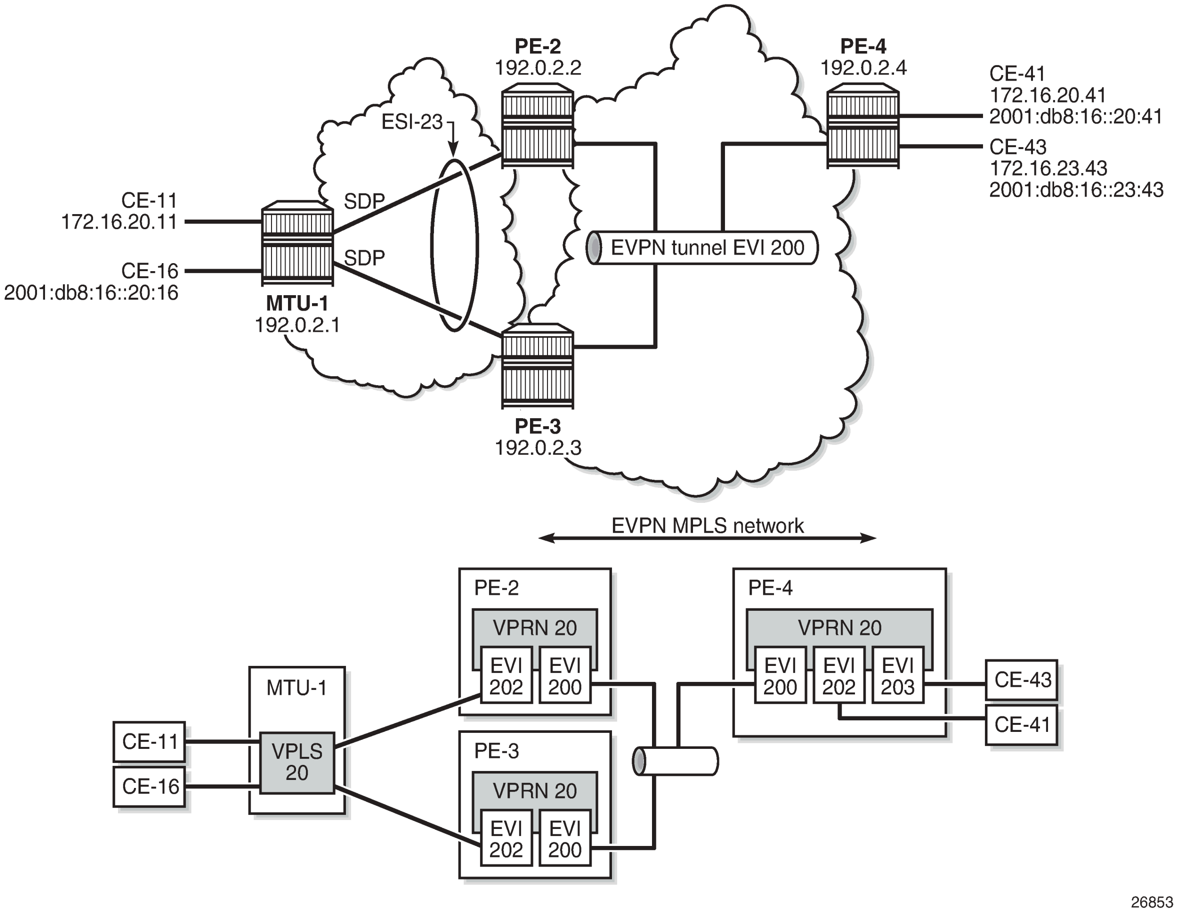
CE-43's MAC address is not present in VPLS 202's FDB. VPLS 203's FDB shows the CE-43's MAC address, but not CE-41's. Traffic between these two VPLS services goes via the VPRN and cannot use Layer 2 forwarding.

```
*A:PE-4# show service id 203 fdb detail

================================================================================
Forwarding Database, Service 203
================================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                          Age
--------------------------------------------------------------------------------
203        00:00:04:00:00:43 sap:1/2/1:43            L/90     02/24/22 15:28:40
203        00:00:5e:00:01:02 cpm                     Intf     02/24/22 15:09:14
203        00:00:5e:00:02:02 cpm                     Intf     02/24/22 15:09:14
203        00:ca:fe:00:23:04 cpm                     Intf     02/24/22 15:09:14
--------------------------------------------------------------------------------
No. of MAC Entries: 4
--------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
================================================================================
```

EVPN-MPLS R-VPLS with single-active MH

Figure 72: EVPN-MPLS R-VPLS with single-active multi-homing shows the example topology with single-active multi-homing ES "SA-ESI-23". The difference is that the ES is single-active and SDPs are used instead of a LAG.

Figure 72: EVPN-MPLS R-VPLS with single-active multi-homing



The configuration is modified as follows:

- LAG 1 is removed from MTU-1, PE-2, and PE-3.
- Network interfaces are configured between MTU-1 and PE-2/PE-3 with IS-IS and LDP enabled.
- SDPs are configured.
- Ethernet segment "SA-ESI-23" is defined as single-active multi-homing. The SDP is associated with this ES.
- VPLS 202 on PE-2 and PE-3 no longer has a SAP, but a spoke-SDP instead.
- No changes are required on VPRN 20 or VPLS 200.

The service configuration on PE-2 is as follows. The configuration on PE-3 is similar. No changes are required on PE-4.

```
# on PE-2:
configure
    service
        system
            bgp-evpn
                ethernet-segment "SA-ESI-23"  create
                    esi 01:00:00:00:00:23:00:00:00:02
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 21
                    no shutdown
                exit
            exit
        exit
---snip---
        sdp 21 mpls create
            far-end 192.0.2.1
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
---snip---
        vprn 20 name "VPRN 20" customer 1 create
            interface "int-evi-202" create
                address 172.16.20.2/24
                mac 00:ca:fe:00:02:02
                vrrp 1 passive
                    backup 172.16.20.254
                    ping-reply
                    traceroute-reply
                exit
                ipv6
                    address 2001:db8:16::20:2/120
                    link-local-address fe80::16:20:2 dad-disable
                    vrrp 1 passive
                        backup fe80::16:20:fe
                        ping-reply
                        traceroute-reply
                    exit
                exit
                vpls "evi-202"
                exit
            exit
            interface "int-evi-200" create
                ipv6
                exit
                vpls "evi-200"
                    evpn-tunnel
                exit
            exit
            router-advertisement
                interface "int-evi-202"
                    use-virtual-mac
                    no shutdown
                exit
            exit
            no shutdown
        exit
        vpls 200 name "evi-200" customer 1 create
            allow-ip-int-bind
            exit
            bgp
            exit
            bgp-evpn
                ip-route-advertisement
                evi 200
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
        vpls 202 name "evi-202" customer 1 create
            allow-ip-int-bind
            exit
            bgp
            exit
            bgp-evpn
                evi 202
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            spoke-sdp 21:20 create
                no shutdown
            exit
            no shutdown
        exit
```

PE-2 is the Designated Forwarder (DF) in the single-active ES, as shown in the following output:

```
*A:PE-2# show service id 202 ethernet-segment
No sap entries

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                           Status
-------------------------------------------------------------------------------
21:20                 SA-ESI-23                         DF
===============================================================================
No vxlan instance entries
```

```
*A:PE-3# show service id 202 ethernet-segment
No sap entries
```

```
===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
31:20                 SA-ESI-23                        NDF
===============================================================================
No vxlan instance entries
```

When traffic has been sent between CE-11 and CE-41, the FDB on PE-2 is as follows. MAC address 00:00:01:00:00:11 corresponds to CE-11 and has been learned on spoke-SDP 21:20; MAC address 00:00:04:00:00:41 corresponds to CE-41 and has been advertised by PE-4 in an EVPN-MAC route.

```
*A:PE-2# show service id 202 fdb detail

===============================================================================
Forwarding Database, Service 202
===============================================================================
ServId     MAC               Source-Identifier        Type     Last Change
            Transport:Tnl-Id                          Age
-------------------------------------------------------------------------------
202        00:00:01:00:00:11 sdp:21:20                L/30     02/24/22 15:36:52
202        00:00:01:00:00:16 sdp:21:20                L/30     02/24/22 15:37:00
202        00:00:04:00:00:41 mpls-1:                  Evpn     02/24/22 15:36:56
                             192.0.2.4:524281
           ldp:65539
202        00:00:5e:00:01:01 cpm                      Intf     02/24/22 15:08:50
202        00:00:5e:00:02:01 cpm                      Intf     02/24/22 15:08:50
202        00:ca:fe:00:02:02 cpm                      Intf     02/24/22 15:08:50
202        00:ca:fe:00:02:03 mpls-1:                  EvpnS:P  02/24/22 15:09:03
                             192.0.2.3:524276
           ldp:65538
202        00:ca:fe:00:02:04 mpls-1:                  EvpnS:P  02/24/22 15:09:14
                             192.0.2.4:524281
           ldp:65539
-------------------------------------------------------------------------------
No. of MAC Entries: 8
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

When the SDP between MTU-1 and DF PE-2 goes down, traffic from CE-41 to CE-11 is forwarded by PE-4 to DF PE-2. PE-2 cannot forward the packets to CE-11 directly, and will forward the packets to its ES peer PE-3. PE-3 will forward to CE-11 even if the MAC SA matches its own vMAC. Virtual MACs bypass the R-VPLS interface protection, so traffic can be forwarded between the PEs without being dropped.

Conclusion

EVPN can be used as the unified control plane VPN technology, not only for providing Layer 2 connectivity, but also Layer 3 (inter-subnet forwarding). EVPN for MPLS tunnels, along with multi-homing and passive VRRP, provides efficient layer-2/layer-3 connectivity to distributed hosts and routers.

EVPN for PBB over MPLS (PBB-EVPN)

This chapter provides information about EVPN for PBB over MPLS (PBB-EVPN).

Topics in this chapter include:

Applicability

This chapter was initially written for SR OS Release 13.0.R6. The CLI in the current edition is based on SR OS Release 21.2.R1.

Important note: A prerequisite is to read the EVPN for MPLS Tunnels chapter.

Overview

EVPN for Provider Backbone Bridging (PBB) over MPLS (hereafter called PBB-EVPN) is specified in RFC 7623, *Provider Backbone Bridging Combined with Ethernet VPN (PBB-EVPN)*. It provides a simplified version of EVPN-MPLS for cases where the network requires very high scalability and does not need all the advanced features supported by EVPN-MPLS (but still requires single-active and all-active multi-homing capabilities). Table 4: EVPN and PBB-EVPN SR OS feature comparison provides a comparison between the capabilities of EVPN and PBB-EVPN in SR OS, and may help to choose between them when designing a VPN service.

Table 4: EVPN and PBB-EVPN SR OS feature comparison

VPN requirements	EVPN	PBB-EVPN	Comments
All-active Multi-Homing (MH) (flow-based load-balancing)	Yes	Yes	Allows better bandwidth utilization
Single-active MH (service-based load-balancing)	Yes	Yes	
Ethernet Local Area Network (E-LAN) and point-to-point E-Line services	Yes	Yes	
Inter-subnet-forwarding	Yes	No	Allows combined Layer 2 / Layer 3 services. EVPN

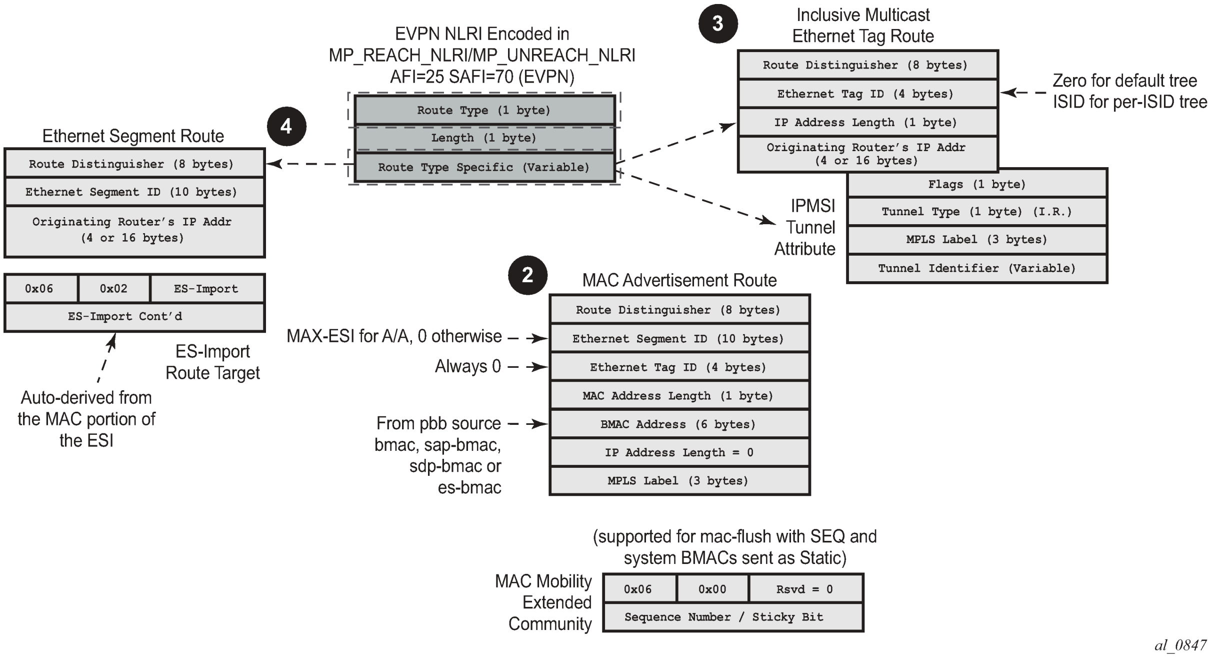
VPN requirements	EVPN	PBB-EVPN	Comments
Proxy-Address Resolution Protocol / Neighbor Discovery (Proxy-ARP/ND) and IP-duplication protection	Yes	No	Allows Broadcast, Unknown unicast and Multicast (BUM) traffic reduction and better security
Customer MAC (CMAC) protection	Yes	No	Allows protecting key static CMACs
Data Center integration	Yes	No	Integration with VXLAN and Nuage Virtualized Services Directory (VSD)
Control plane overhead	Medium	Low	PBB-EVPN only advertises Backbone MACs (BMACs) and no route type 1s
Confinement of CMAC learning	No	Yes	CMACs are only learned on PEs with flows using those CMACs
CMAC summarization	No	Yes	Aggregation of CMACs into BMACs

PBB-EVPN is a combination of 802.1ah PBB and RFC 7432, *BGP MPLS-Based Ethernet VPN* (EVPN-MPLS), and reuses the PBB-Virtual Private LAN Service (VPLS) service model, where Border Gateway Protocol BGP-EVPN is enabled in the backbone VPLS (B-VPLS) domain. EVPN is used as the control plane in the B-VPLS domain to control the distribution of BMACs and set up per-backbone service instance identifier (ISID) flooding trees for service instance VPLS (I-VPLS) services. The learning of the CMACs, either on local SAPs/SDP-bindings or associated with remote BMACs, is still performed in the data plane. Only the learning of BMACs in the B-VPLS is performed through BGP.

The SR OS PBB-EVPN implementation supports I-VPLS and PBB-Epipe services, including single-active and all-active multi-homing.

Because PBB-EVPN is based on the same control plane model as EVPN for MPLS, it is recommended to read the EVPN for MPLS Tunnels chapter before configuring PBB-EVPN. PBB-EVPN uses a subset of the BGP-EVPN routes described in EVPN for MPLS Tunnels as shown in Figure 73: EVPN route types.

Figure 73: EVPN route types


EVPN NLRI Encoded in
MP_REACH_NLRI/MP_UNREACH_NLRI
AFI=25 SAFI=70 (EVPN)
Route Type (1 byte)
Length (1 byte)
Route Type Specific (Variable)
4
Ethernet Segment Route
Route Distinguisher (8 bytes)
Ethernet Segment ID (10 bytes)
Originating Router's IP Addr (4 or 16 bytes)
0x06
0x02
ES-Import
ES-Import Cont'd
ES-Import Route Target
Auto-derived from the MAC portion of the ESI
3
Inclusive Multicast Ethernet Tag Route
Route Distinguisher (8 bytes)
Ethernet Tag ID (4 bytes)
IP Address Length (1 byte)
Originating Router's IP Addr (4 or 16 bytes)
Zero for default tree
ISID for per-ISID tree
IPMSI Tunnel Attribute
Flags (1 byte)
Tunnel Type (1 byte) (I.R.)
MPLS Label (3 bytes)
Tunnel Identifier (Variable)
2
MAC Advertisement Route
Route Distinguisher (8 bytes)
Ethernet Segment ID (10 bytes)
Ethernet Tag ID (4 bytes)
MAC Address Length (1 byte)
BMAC Address (6 bytes)
IP Address Length = 0
MPLS Label (3 bytes)
MAX-ESI for A/A, 0 otherwise
Always 0
From pbb source bmac, sap-bmac, sdp-bmac or es-bmac
(supported for mac-flush with SEQ and system BMACs sent as Static)
MAC Mobility Extended Community
0x06
0x00
Rsvd = 0
Sequence Number / Sticky Bit
al_0847

When no EVPN multi-homing is used in the network, only the base routes are used. Route types 2 and 3 are considered the base and mandatory routes:

- Route type 2 — (B) MAC route — In PBB-EVPN, this route type is used for the advertisement of BMACs that will be installed in the remote Forwarding Data Bases (FDBs). There are no IP addresses advertised in PBB-EVPN. The MAC mobility extended community is used for advertising system BMACs as **protected** (with the sticky bit set) and it is also used for CMAC flush in some single-homing scenarios that will be described later.
- Route type 3 — Inclusive Multicast route — This route type is used for the advertisement of the I-VPLS ISIDs (no Epipes) and the desired multicast tree for each of them. The ISIDs are encoded in the Ethernet-tag field of the Network Layer Reachability Information (NLRI). When the B-VPLS is created and enabled, an Inclusive Multicast route with ISID = 0 is advertised. This is for the creation of the default multicast tree.

When EVPN multi-homing is used in an ISID, route type 4 (Ethernet Segment (ES) route) is used. In PBB-EVPN, there is no route type 1 advertised when multi-homing is used on the ISID services (I-VPLS and Epipes). Only route type 4 is used, and in the same way as it is for EVPN-MPLS. See the EVPN for MPLS Tunnels example for more information about ES routes, how they are formed, and how their RT/RD values are populated.

Configuration

This example describes the basic PBB-EVPN configuration first (without multi-homing) and how the flood containment is handled in PBB-EVPN. Flood containment refers to the efficient distribution of the BUM traffic generated for an ISID.

Networks are not always greenfield, so a smooth migration of PBB-EVPN from PBB-VPLS is required to minimize the effect on existing services. This example also describes this migration, starting from a common PBB-VPLS configuration.

Finally, this example describes the configuration of PBB-EVPN multi-homing.

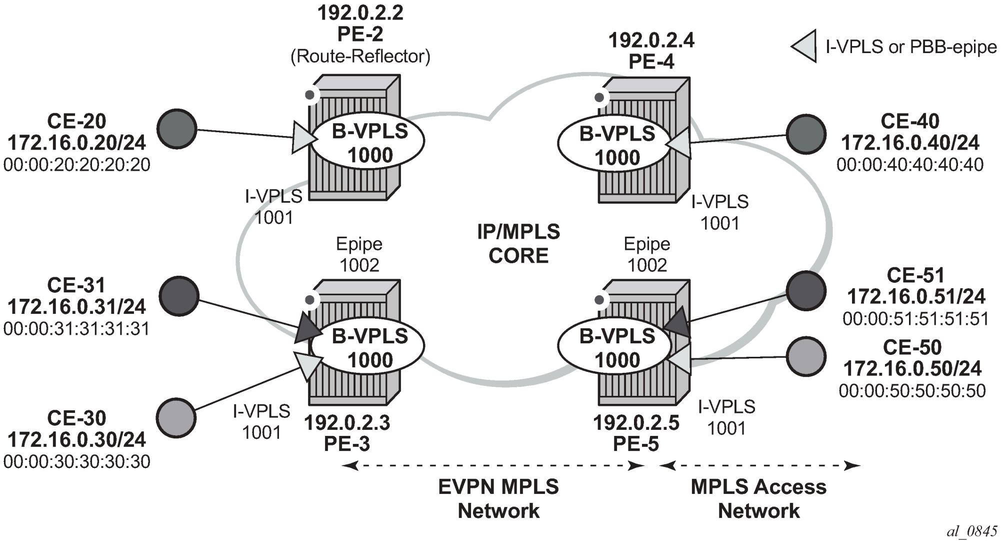
The same setup described in the EVPN for MPLS Tunnels example is used:

- Four PEs in the core (PE-2, PE-3, PE-4, and PE-5).
- The PEs are interconnected in the same way as explained in EVPN for MPLS Tunnels with the same IP addressing, IS-IS, transport LDP, and BGP peering configuration. There is not any difference with the basic infrastructure. See the EVPN for MPLS Tunnels chapter if more information is required.
- When configuring multi-homing, MTU-1 and MTU-6 are connected to the core.

PBB-EVPN configuration without multi-homing

Figure 74: PBB-EVPN network without multi-homing shows the example topology used in this chapter.

Figure 74: PBB-EVPN network without multi-homing



When configuring PBB-EVPN:

- There is no difference at the access side (I-VPLS and Epipe configuration) compared to other PBB technologies supported in SR OS, such as Shortest Path Bridging for MAC (SPBM) or PBB-VPLS.
- The B-VPLS becomes an EVPN-MPLS service, where bgp-evpn mpls is added.

The following output shows an example of a basic configuration in PE-3. B-VPLS 1000 is bgp-evpn enabled and I-VPLS 1001 and Epipe 1002 are linked to B-VPLS 1000.

```
# on PE-3:
configure
    service
        vpls 1000 name "B-VPLS 1000" customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:00:00:00:00:03
            exit
            bgp
            exit
            bgp-evpn
                evi 1000
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
        vpls 1001 name "I-VPLS 1001" customer 1 i-vpls create
            pbb
                backbone-vpls 1000
                exit
            exit
            stp
                shutdown
            exit
            sap 1/2/1:1001 create
                no shutdown
            exit
            no shutdown
        exit
        epipe 1002 name "Epipe 1002" customer 1 create
            pbb
                tunnel 1000 backbone-dest-mac 00:00:00:00:00:05 isid 1002
            exit
            sap 1/2/1:1002 create
                no shutdown
            exit
            no shutdown
        exit
```

In the preceding output, there is no new configuration needed for I-VPLS/Epipe services. As for the B-VPLS, the output shows the minimum configuration required. If needed, the following parameters can be modified under **bgp-evpn**:

```
*A:PE-2>config>service>vpls# bgp-evpn ?
  - bgp-evpn
  - no bgp-evpn

 [no] accept-ivpls-e* - Configure to accept non-zero ethernet-tag MAC routes and
                        process for CMAC flushing
 [no] cfm-mac-advert* - Enable/disable the advertisement of MEP, MIP, and VMEP MAC
                        addresses over the BGP EVPN
 [no] evi             - EVPN Identifier
 [no] incl-mcast-ori* - Configure originating IP address
 [no] ingress-repl-i* - Configure BGP EVPN IMET-IR route advertisement
 [no] ip-route-adver* - Configure BGP EVPN IP Route Advertisement
      isid-route-tar* + configure ISID route target information
 [no] mac-advertisem* - Configure BGP EVPN MAC Advertisement
      mac-duplication + Configure BGP EVPN MAC Duplication
 [no] mpls            + Configure BGP EVPN mpls
 [no] sel-mcast-adve* - Enable/disable selective multicast advertisements
 [no] unknown-mac-ro* - Configure BGP EVPN Unknown MAC Route
 [no] vxlan           + Configure BGP EVPN vxlan
```

```
*A:PE-2>config>service>vpls>bgp-evpn# mpls ?
  - no mpls [bgp <bgp>]
  - mpls [bgp <bgp>]

 <bgp>                : [1..2]


      auto-bind-tunn* + Configure BGP EVPN mpls auto-bind-tunnel
 [no] control-word    - Enable/disable setting the CW bit in the label message
 [no] default-route-* - Configure default-route-tag to match against export policies
      ecmp            - Configure maximum ECMP routes information
 [no] entropy-label   - Enable/disable use of entropy-label
 [no] force-vlan-vc-* - Forces vlan-vc-type forwarding in the data-path
 [no] ingress-replic* - Use the same label as the one advertised for unicast traffic
 [no] oper-group      - Configure oper-group
 [no] restrict-prote* - Enable/disable protected src MAC restriction
      route-next-hop  - Configure route next-hop
 [no] send-tunnel-en* - Configure encapsulation for this service
 [no] shutdown        - Administratively Enable/Disable BGP-EVPN mpls
 [no] split-horizon-* - Configure a split-horizon-group
```

A detailed description of these commands is included in the EVPN for MPLS Tunnels chapter. In addition to the preceding commands, the following **service (b-)vpls pbb** commands are relevant for PBB-EVPN in the B-VPLS service:

- **force-qtag-forwarding** allows the transparent transport of the customer 802.1p bits across the B-VPLS services.
- **source-bmac** can modify the source BMAC for all the PBB packets containing traffic from non-multi-homed I-VPLS and Epipe services.
- **use-es-bmac** instructs the system to use an ES-specific BMAC for traffic coming from an ES on an I-VPLS or Epipe.
- **use-sap-bmac** instructs the system to use a SAP-specific BMAC for traffic coming from an MC-LAG I-VPLS/Epipe SAP.

Flood containment for I-VPLS services

In general, PBB technologies in SR OS support a way to contain flooding for a specified I-VPLS ISID, so that BUM traffic for that ISID only reaches the PEs where the ISID is locally defined. Each PE creates a Multicast Forwarding Information Base (MFIB) per I-VPLS ISID on the B-VPLS instance. That MFIB supports SAP/SDP-binding endpoints that can be populated by:

- Multiple MAC Registration Protocol (MMRP) in regular PBB-VPLS
- IS-IS in SPBM

In PBB-EVPN, B-VPLS EVPN destinations can be added to the MFIBs using EVPN Inclusive Multicast Ethernet tag routes when they include the ISID in the Ethernet-tag. By default, when a B-VPLS is successfully enabled (**no shutdown**), the PE advertises:

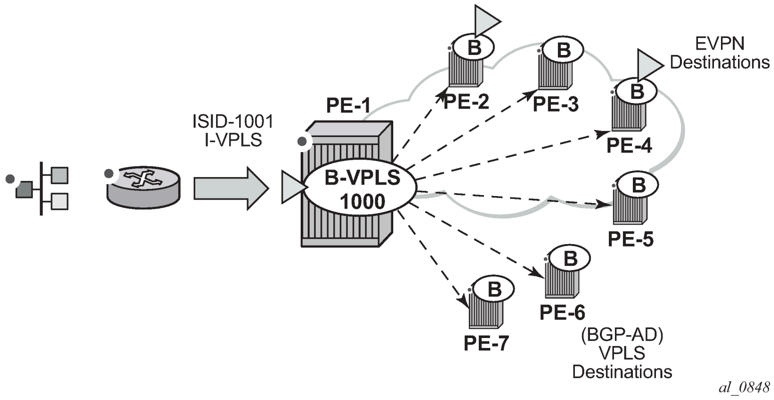
- An Inclusive Multicast route for ISID = 0 — This allows the remote PEs to add the advertising PE to the default-multicast-list for the B-VPLS.

- An Inclusive Multicast route for each local ISID defined in the system (a local ISID includes configured I-VPLS and static-ISIDs) — This allows the remote PEs to create MFIB entries in the B-VPLS for the received ISIDs.

Because EVPN destinations, B-SAPs, and B-spoke-SDPs can coexist in the same B-VPLS, be aware of the different flooding lists created and how they are used in a B-VPLS. Figure 75: PBB-EVPN — flooding lists illustrates this concept with an example for B-VPLS 1000 in PE-1. The assumptions are:

- I-VPLS 1001 is created in PE-1, PE-2, and PE-4 only.
- PE-1, PE-2, PE-3, PE-4, and PE-5 support BGP-EVPN in B-VPLS 1000.
- PE-6 and PE-7 only support spoke-SDPs.
- PE-1 is connected to all six PEs.

Figure 75: PBB-EVPN — flooding lists



In this situation, PE-1 creates two flooding lists in B-VPLS 1000:

- Default-multicast-list — composed of:
  - All the EVPN PEs that advertised ISID = 0 (PE-2, PE-3, PE-4, PE-5).
  - All the B-spoke-SDPs (or B-SAPs) (PE-6, PE-7).
  - All the EVPN PEs that advertised ISID 1001 and no ISID 0 (if an isid-policy is created in PE-1 stating **use-def-mcast** for ISID 1001). Note: third-party PEs may not advertise ISID = 0, but only non-zero ISIDs.
- MFIB for ISID 1001 is composed of:
  - All the EVPN PEs that advertised ISID 1001 (PE-2 and PE-4) unless there is an ISID-policy in PE-1 stating **use-def-mcast** for ISID 1001.
  - Static-ISIDs defined in manual B-spoke-SDPs and B-SAPs (static-ISIDs cannot be created on BGP-AD auto-discovered B-spoke-SDPs).

Based on the above, when BUM traffic is sent to I-VPLS 1001 on PE-1:

- The traffic is encapsulated in PBB with the group BMAC for ISID 1001 and sent (by default) to the MFIB created for ISID 1001 (PE-2 and PE-4).
- If an ISID-policy is added with **use-def-mcast** for ISID 1001, the BUM traffic is encapsulated in PBB with the group BMAC for ISID 1001 and sent to the default-multicast-list, that is, all six remote PEs.

Referring to Figure 74: PBB-EVPN network without multi-homing, the following output illustrates the use of the ISID-policy in PBB-EVPN. PE-2 does not have any ISID-policy configured; when it receives BUM traffic from the local I-VPLS 1001, it uses the MFIB for ISID 1001:

```
# on PE-2:
configure
    service
        vpls 1000 name "B-VPLS 1000" customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:00:00:00:00:02
            exit
            bgp
            exit
            bgp-evpn
                evi 1000
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
```

```
*A:PE-2# show service id 1000 mfib

===============================================================================
Multicast FIB, Service 1000
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
*               01:1e:83:00:03:e9     b-mpls:192.0.2.3:524279      Local    Fwd
                                      b-mpls:192.0.2.4:524280      Local    Fwd
                                      b-mpls:192.0.2.5:524279      Local    Fwd
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

An ISID-policy can be added to modify this behavior and allow PE-2 to use the default multicast list. If I-VPLS 1001 exists in all the remote PEs (as in this example), using the default multicast list is as efficient as using the MFIB and saves expensive MFIB resources. In the following output, as soon as the ISID-policy is added, the MFIB entries for ISID 1001 are removed and PE-2 starts using the default multicast list.

```
# on PE-2:
configure
    service
        vpls "B-VPLS 1000"
            isid-policy
                entry 10 create
                    use-def-mcast
                    range 1001 to 2000
                exit
            exit
```

The MFIB on PE-2 does not contain any entries for ISID 1001 anymore, as follows:

```
*A:PE-2# show service id 1000 mfib

===============================================================================
Multicast FIB, Service 1000
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Number of entries: 0
===============================================================================
```

PBB-VPLS to PBB-EVPN migration

The principles required for migrating a PBB-VPLS network to PBB-EVPN are explained in the *VPLS to EVPN-MPLS Integration* section of the EVPN for MPLS Tunnels chapter. Those principles are also applicable to EVPN destinations and spoke-SDPs in the B-VPLS and can be summarized in three points:

- Systems with an EVPN destination and SDP-binding to the same far-end IP bring down the SDP-binding. This avoids loops when both constructs exist in the same network.
- SDP-bindings and EVPN destinations can be placed in the same Split-Horizon Group (SHG). When traffic from an SDP-binding/EVPN destination belonging to that SHG is received on a PE, it is never forwarded to another SDP-binding or EVPN destination on the same SHG.
- MAC addresses learned on an SDP-binding or SAP, that belong to an SHG where EVPN destinations are also created, are not advertised in BGP-EVPN.

Based on those principles, this section describes how to migrate a PBB-VPLS network to PBB-EVPN. The network in Figure 74: PBB-EVPN network without multi-homing represents a regular PBB-VPLS network that needs to be migrated to PBB-EVPN.

In that network, the four PEs are running BGP-AD and TLDP for the discovery and setup of the pseudowires in the B-VPLS instance. The advantage of this configuration is that the migration can be done node by node and with minimum impact on customer service.

Initial configuration

Initially, the network is configured for PBB-VPLS with BGP-AD in B-VPLS 1000. The EVPN family is to be added. At the access, I-VPLS 1001 is connected to the CEs. As an example, the configuration in PE-3 is shown. An equivalent configuration exists in the other three PEs.

Note:
The EVPN family is added to the BGP configuration because PBB-EVPN uses this address family. Assuming there are redundant Route Reflectors (RRs), the addition of EVPN can be done without service impact. In this example, the assumption is that the PEs are already configured with the EVPN family.

```
*A:PE-3#
configure
    router Base
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family l2-vpn evpn
                peer-as 64500
                neighbor 192.0.2.2
                exit
            exit
```

```
# on PE-3:
configure
    service
        pw-template 1 name "PW1" create
            split-horizon-group "CORE"
            exit
        exit
        vpls 1000 name "B-VPLS 1000" customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:00:00:00:00:03
            exit
            bgp
                pw-template-binding 1
                exit
            exit
            bgp-ad
                vpls-id 64500:1000
                no shutdown
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
        vpls 1001 name "I-VPLS 1001" customer 1 i-vpls create
            pbb
                backbone-vpls 1000
                exit
            exit
            stp
                shutdown
            exit
            sap 1/2/1:1001 create
            exit
            no shutdown
        exit
```

```
*A:PE-3# show service id 1000 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1000                Vpn Id            : 0
Service Type      : b-VPLS
---snip---

Oper Backbone Src : 00:00:00:00:00:03
Use SAP B-MAC     : Disabled
i-Vpls Count      : 1
```

```
Epipe Count        : 1
Use ESI B-MAC      : Disabled

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                              Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sdp:32765:4294967293 SB(192.0.2.5)      BgpAd        0       8978    Up   Up
sdp:32766:4294967294 SB(192.0.2.4)      BgpAd        0       8978    Up   Up
sdp:32767:4294967295 SB(192.0.2.2)      BgpAd        0       8978    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

Multiple MAC Registration Protocol (MMRP) is not used in the B-VPLS instance. If it were enabled, MMRP would have to be disabled in the network before this migration. If there are ISIDs using B-VPLS SDP-bindings to reach some remote locations and B-VPLS EVPN destinations to reach others, the default multicast list must be used in the current release (the MFIB cannot be used if there is a mix of both types). Therefore, during the migration process, the ISIDs must be added to the default multicast list.

1. Add service-level SHG (if not already there).

   From the first node being migrated to PBB-EVPN to all nodes migrated, PBB-VPLS and PBB-EVPN have to coexist within the same meshed network. That is, EVPN-MPLS destinations and SDP-bindings need to be defined in the same split-horizon group. Therefore, if there is no split-horizon group defined in the B-VPLS, the first step is to add it. In this example, the split-horizon group is defined at the **config>service>pw-template>level**; therefore, it has to be added at the B-VPLS level.

   Note:
   When the **service>split-horizon-group** is removed, an eval-pw-template must be performed.

   Note:
   After adding the **split-horizon-group** at the service level, an eval-pw-template must be performed again so that the SDP-bindings take the new SHG configuration.

   Note:
   During the time between the **split-horizon-group** being removed and added back again, the SDP-bindings can forward BUM traffic to each other, so this operation must be done carefully to avoid loops.

   Assuming that the first node to be migrated is PE-3, the following output shows the procedure for adding the **split-horizon-group** at the service level.

```
# on PE-3:
configure
    service
        pw-template 1
            no split-horizon-group
```

```
*A:PE-3# tools perform service id 1000 eval-pw-template 1 allow-service-impact
eval-pw-template succeeded for Svc 1000 32765:4294967293 Policy 1
eval-pw-template succeeded for Svc 1000 32766:4294967294 Policy 1
eval-pw-template succeeded for Svc 1000 32767:4294967295 Policy 1
```

```
# on PE-3:
configure
    service
```

```
        vpls "B-VPLS 1000"
            split-horizon-group "CORE" create
            exit
            bgp
                pw-template-binding 1 split-horizon-group "CORE"
                exit
            exit
```

```
*A:PE-3# tools perform service id 1000 eval-pw-template 1 allow-service-impact
eval-pw-template succeeded for Svc 1000 32765:4294967293 Policy 1
eval-pw-template succeeded for Svc 1000 32766:4294967294 Policy 1
eval-pw-template succeeded for Svc 1000 32767:4294967295 Policy 1
```

```
*A:PE-3>config>service>vpls# info
----------------------------------------------
            service-mtu 2000
            pbb
                source-bmac 00:00:00:00:00:03
            exit
            split-horizon-group "CORE" create
            exit
            bgp
                pw-template-binding 1 split-horizon-group "CORE"
                exit
            exit
            bgp-ad
                vpls-id 64500:1000
                no shutdown
            exit
            stp
                shutdown
            exit
            no shutdown
```

2. Add BGP-EVPN and ISID-policy configuration to the B-VPLS.

   After the B-VPLS is configured with the split horizon group, the BGP-EVPN configuration (still in **shutdown**) and an ISID-policy can be added, as follows.

```
# on PE-2, PE-3, PE-4, PE-5:
configure
    service
        vpls "B-VPLS 1000"
            bgp-evpn
                evi 1000
                mpls bgp 1
                    shutdown
                    split-horizon-group "CORE"
                    auto-bind-tunnel
                        resolution any
                    exit
                exit
            exit
            isid-policy
                entry 10 create
                    use-def-mcast
                    range 1001 to 3000
                exit
            exit
```

3. Enable BGP-EVPN MPLS on the PE.

When the configuration is ready, the **bgp-evpn mpls bgp 1** context can be enabled, as follows:

```
# on PE-3:
configure
    service
        vpls "B-VPLS 1000"
            bgp-evpn
                mpls bgp 1
                    no shutdown
```

Enabling the **bgp-evpn mpls bgp 1** context triggers a route-refresh message for the EVPN family from PE-3, but no changes happen because PE-3 does not create any EVPN destinations until it imports EVPN routes from the other PEs. The three spoke-SDPs to the remote PEs are still up.

4. Repeat steps 1 to 3 for the second PE (PE-5).

   The same steps 1 to 3 are repeated for PE-5. When the **bgp-evpn mpls bgp 1** context is enabled, PE-5 sends a route-refresh and gets the BGP-EVPN routes from PE-3. As a result of that, PE-3 brings down the spoke-SDP to PE-5 and creates an EVPN destination to PE-5. The same process happens in PE-5. The following CLI output shows the received routes in PE-3 and spoke-SDP going down.

```

25 2021/03/03 11:00:17.445 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 110
    Flag: 0x90 Type: 14 Len: 47 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.5
        Type: EVPN-INCL-MCAST Len: 17 RD: 64500:1000, tag: 1001, orig_addr len: 32,
                              orig_addr: 192.0.2.5
        Type: EVPN-INCL-MCAST Len: 17 RD: 64500:1000, tag: 0, orig_addr len: 32,
                              orig_addr: 192.0.2.5
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.5
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        1.1.1.1
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:1000
        bgp-tunnel-encap:MPLS
    Flag: 0xc0 Type: 22 Len: 9 PMSI:
        Tunnel-type Ingress Replication (6)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf: not required]
        MPLS Label 8388480
        Tunnel-Endpoint 192.0.2.5
"
```

   Log 99 shows that spoke SDP 32765:4294967293 is operationally down:

```
171 2021/03/03 11:00:17.766 UTC MINOR: SVCMGR #2313 Base
"Status of SDP Bind 32765:4294967293 in service 1000 (customer 1) peer PW status bits
 changed to pwNotForwarding "

170 2021/03/03 11:00:16.687 UTC MINOR: SVCMGR #2306 Base
"Status of SDP Bind 32765:4294967293 in service 1000 (customer 1) changed to admin=up oper=
down flags=evpnRouteConflict "

169 2021/03/03 11:00:16.687 UTC MINOR: SVCMGR #2326 Base
```

```
"Status of SDP Bind 32765:4294967293 in service 1000 (customer 1) local PW status bits
 changed to pwNotForwarding "
```

Spoke SDP 32765:4294967293 is the spoke SDP toward PE-5 and it is kept down:

```
*A:PE-3# show service id 1000 base
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sdp:32765:4294967293 SB(192.0.2.5)       BgpAd        0       8978    Up   Down
sdp:32766:4294967294 SB(192.0.2.4)       BgpAd        0       8978    Up   Up
sdp:32767:4294967295 SB(192.0.2.2)       BgpAd        0       8978    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

The reason why the spoke SDP toward PE-5 is down is an EVPN route conflict:

```
*A:PE-3# show service id 1000 sdp 32765:4294967293 detail | match Flag context all
Flags              : PWPeerFaultStatusBits
                     EvpnRouteConflict
```

An EVPN destination to PE-5 is created:

```
*A:PE-3# show service id 1000 evpn-mpls

===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address     Egr Label     Num. MACs   Mcast           Last Change
                Transport:Tnl                             Sup BCast Domain
-------------------------------------------------------------------------------
192.0.2.5       524280        1           bum             03/03/2021 11:00:17
                ldp:65539                                 No
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

5. Repeat Steps 1 to 3 for the rest of the PEs (PE-2, PE-4).

   The same process is repeated in all the PEs, node by node. The service impact for the I-VPLS 1001 is minimal.

6. (Optional) Remove the ISID policy.

   When all the PEs in the B-VPLS 1000 are migrated, the ISID policy can optionally be removed, node by node. This forces the B-VPLS instance to start using the MFIB to send I-VPLS BUM traffic to the remote nodes. This has no effect on Epipes (traffic is always unicast for Epipes).

   Before removing the ISID policy and starting to use the MFIB, it is recommended to check that the Inclusive Multicast routes for an ISID to the remote PEs are all active. Otherwise, connectivity for BUM traffic could be interrupted if any of the expected routes are not active. This is illustrated for PE-3.

```
*A:PE-3# show service id 1000 evpn-mpls

===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address     Egr Label     Num. MACs   Mcast           Last Change
                Transport:Tnl                             Sup BCast Domain
-------------------------------------------------------------------------------
192.0.2.2       524281        1           bum             03/03/2021 11:02:31
                ldp:65537                                 No
192.0.2.4       524280        1           bum             03/03/2021 11:02:32
                ldp:65538                                 No
192.0.2.5       524280        1           bum             03/03/2021 11:00:17
                ldp:65539                                 No
-------------------------------------------------------------------------------
Number of entries : 3
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

The routes for ISID 1001 are valid and used by BGP (flags **u*>i**):

```
*A:PE-3# show router bgp routes evpn incl-mcast tag 1001
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Flag  Route Dist.         OrigAddr
      Tag                 NextHop
-------------------------------------------------------------------------------
u*>i  64500:1000          192.0.2.2
      1001                192.0.2.2

u*>i  64500:1000          192.0.2.4
      1001                192.0.2.4

u*>i  64500:1000          192.0.2.5
      1001                192.0.2.5

-------------------------------------------------------------------------------
Routes : 3
===============================================================================
```

There are no entries in the MFIB:

```
*A:PE-3# show service id 1000 mfib

===============================================================================
Multicast FIB, Service 1000
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Number of entries: 0
===============================================================================
```

The ISID policy is removed as follows:

```
# on PE-2, PE-3, PE-4, PE-5:
configure
    service
        vpls "B-VPLS 1000"
            isid-policy
                no entry 10
```

After removing the ISID-policy, the MFIB is populated with entries for the ISID 1001 group BMAC to the three remote PEs where ISID 1001 is defined:

```
*A:PE-3# show service id 1000 mfib

===============================================================================
Multicast FIB, Service 1000
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
*               01:1e:83:00:03:e9     b-mpls:192.0.2.2:524281      Local    Fwd
                                      b-mpls:192.0.2.4:524280      Local    Fwd
                                      b-mpls:192.0.2.5:524280      Local    Fwd
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

7. (Optional) Remove the BGP-AD configuration.

   The BGP-AD configuration can stay in the B-VPLS services. However, when the entire network is migrated to PBB-EVPN, all the spoke-SDPs will be operationally down and, even if they are not forwarding traffic, they consume resources in the system. Consider removing the BGP-AD configuration and, therefore, the spoke-SDPs.

   The following example shows the removal of BGP-AD in PE-4. Be aware that when BGP-AD is removed from the configuration, if the RD/RT was derived from the VPLS ID (as in this example), a new RD/RT must be auto-derived for the service. Therefore, new updates will be sent for all the EVPN NLRIs, as shown in the following output.

```

*A:PE-4# show service id 1000 bgp

===============================================================================
BGP Information
===============================================================================
Bgp Instance         : 1
Vsi-Import           : None
Vsi-Export           : None
Route Dist           : None
Oper Route Dist      : 64500:1000
Oper RD Type         : derivedVpls
Rte-Target Import    : None                 Rte-Target Export: None
Oper RT Imp Origin   : derivedVpls          Oper RT Import   : 64500:1000
Oper RT Exp Origin   : derivedVpls          Oper RT Export   : 64500:1000

PW-Template Id       : 1                    PW-Template SHG  : CORE
Oper Group           : None
Mon Oper Group       : None
BFD Template         : None
BFD-Enabled          : no                   BFD-Encap        : ipv4
```

```
Import Rte-Tgt       : None
-------------------------------------------------------------------------------
===============================================================================
```

BGP-AD is disabled as follows:

```
# on PE-4:
configure
    service
        vpls "B-VPLS 1000"
            bgp-ad
                shutdown
```

After BGP-AD is disabled, the spoke SDP bindings are deleted.

```
# log "99" on PE-4:
163 2021/03/03 11:05:42.890 UTC MAJOR: SVCMGR #2319 Base
"Dynamic bgp-l2vpn SDP 32765 (192.0.2.5) was deleted."

162 2021/03/03 11:05:42.890 UTC MINOR: SVCMGR #2303 Base
"Status of SDP 32765 changed to admin=down oper=down"

161 2021/03/03 11:05:42.890 UTC MAJOR: SVCMGR #2320 Base
"Service Id 1000, Dynamic bgp-l2vpn SDP Bind Id 32765:4294967293 was deleted."

160 2021/03/03 11:05:42.880 UTC MINOR: SVCMGR #2306 Base
"Status of SDP Bind 32765:4294967293 in service 1000 (customer 1) changed to admin=down
 oper=down flags="
```

The PW template binding is removed as follows:

```
# on PE-4:
configure
    service
        vpls "B-VPLS 1000"
            bgp
                no pw-template-binding 1
```

The BGP-AD configuration is removed as follows:

```
# on PE-4:
configure
    service
        vpls "B-VPLS 1000"
            no bgp-ad
```

Initially, the RD/RT was derived from the VPLS ID (64500:1000). After the BGP-AD configuration is removed, a new RD and RT must be auto-derived from the EVI:

```
*A:PE-4# show service id 1000 bgp

===============================================================================
BGP Information
===============================================================================
Bgp Instance         : 1
Vsi-Import           : None
Vsi-Export           : None
Route Dist           : None
Oper Route Dist      : 192.0.2.4:1000
Oper RD Type         : derivedEvi
```

```
Rte-Target Import    : None                Rte-Target Export: None
Oper RT Imp Origin   : derivedEvi          Oper RT Import   : 64500:1000
Oper RT Exp Origin   : derivedEvi          Oper RT Export   : 64500:1000

PW-Template Id       : None
-------------------------------------------------------------------------------
===============================================================================
```

In this case, the system picks up the RD in the following order:

a. Manual RD or auto-RD always take precedence when configured.

b. If no manual/auto-RD, the RD is derived from the **bgp-ad vpls-id**.

c. If no manual/auto-rd/vpls-id configuration, the RD is derived from the **bgp evpn evi**.

d. If no manual/auto-rd/vpls-id/evi configuration, there will be no RD and the service will fail.

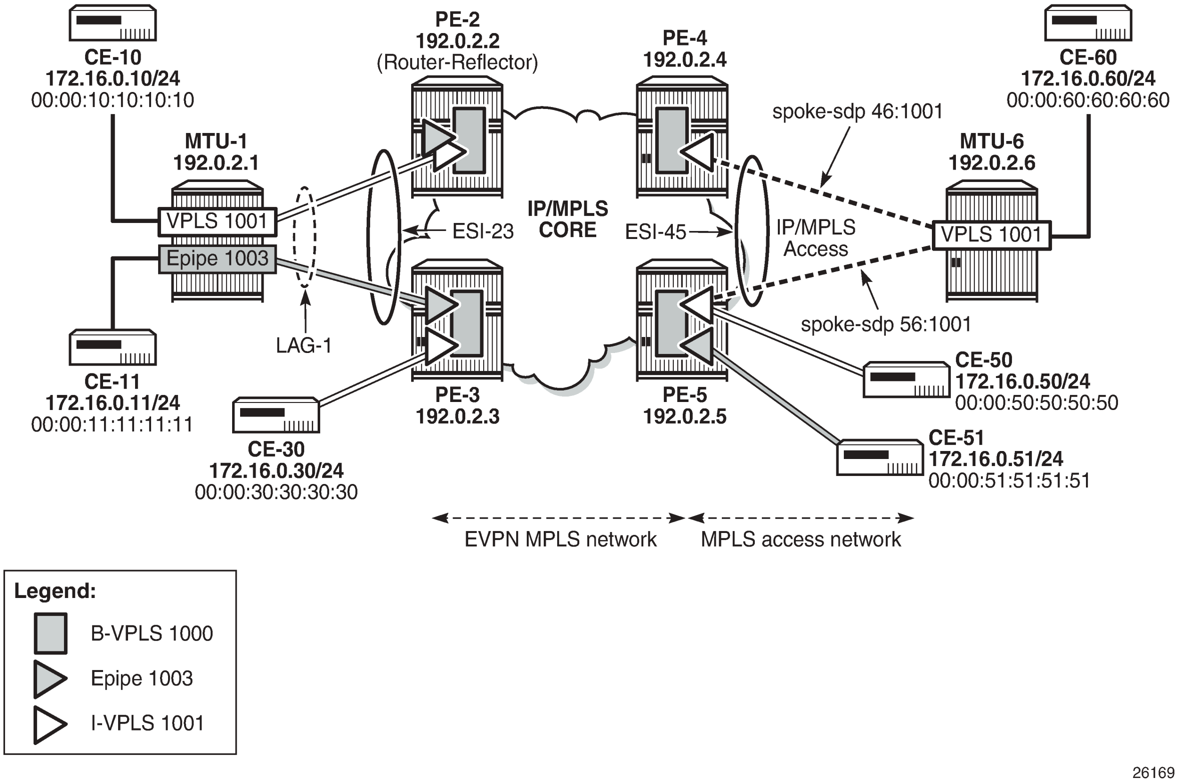
If in the migration from BGP-AD to BGP-EVPN, the advertisement of new updates is not needed, the initial configuration must include manual/auto-RDs. If manual/auto-RDs were not included, disabling BGP-AD would not cause the change of RD and the consequent BGP updates.

PBB-EVPN multi-homing

This section provides configuration guidelines for PBB-EVPN multi-homing. In the same way that EVPN-MPLS supports single-active and all-active multi-homing, PBB-EVPN can also be configured to support both modes. The same Ethernet segment that is used for regular EVPN-MPLS service SAPs and spoke-SDPs can be shared with I-VPLS/Epipe SAPs and spoke-SDPs.

Figure 76: PBB-EVPN multi-homing shows the example topology used in this section.

Figure 76: PBB-EVPN multi-homing



MTU-1 and MTU-6 have been added to the network (compared to Figure 74: PBB-EVPN network without multi-homing). I-VPLS 1001 has two new sites that are multi-homed to the PBB-EVPN network. MTU-1 uses all-active multi-homing, whereas MTU-6 is connected to a single-active ES. As with EVPN-MPLS, all-active multi-homing is only supported when a LAG is used at the access. Single-active multi-homing can be supported with regular Ethernet ports (that can form an independent LAG per PE) or SDPs.

RFC 7623 describes two types of system BMAC assignments that a PE can implement in a B-VPLS when ESs are present:

- Shared BMAC addresses that can be used for all the single-homed CEs and a number of multi-homed CEs connected to Ethernet-segments.
- Dedicated BMAC addresses per Ethernet-segment.

In this chapter and in SR OS terminology:

- A shared BMAC address (in IETF) is a source BMAC address as configured in **service>(b)vpls>pbb>source-bmac**. All the I-VPLS/Epipe traffic coming from single-homed CEs is sent encapsulated in a PBB packet with that source BMAC address.
- A dedicated-BMAC per ES (in IETF) is an ES BMAC address as activated in **service>(b)vpls>pbb>use-es-bmac** and generated from the combination of **vpls>pbb>source-bmac** plus **ethernet-segment>source-bmac-lsb**. If configured, any I-VPLS/Epipe traffic coming from an ES is encapsulated in a PBB packet with the ES-BMAC address as the source BMAC address.

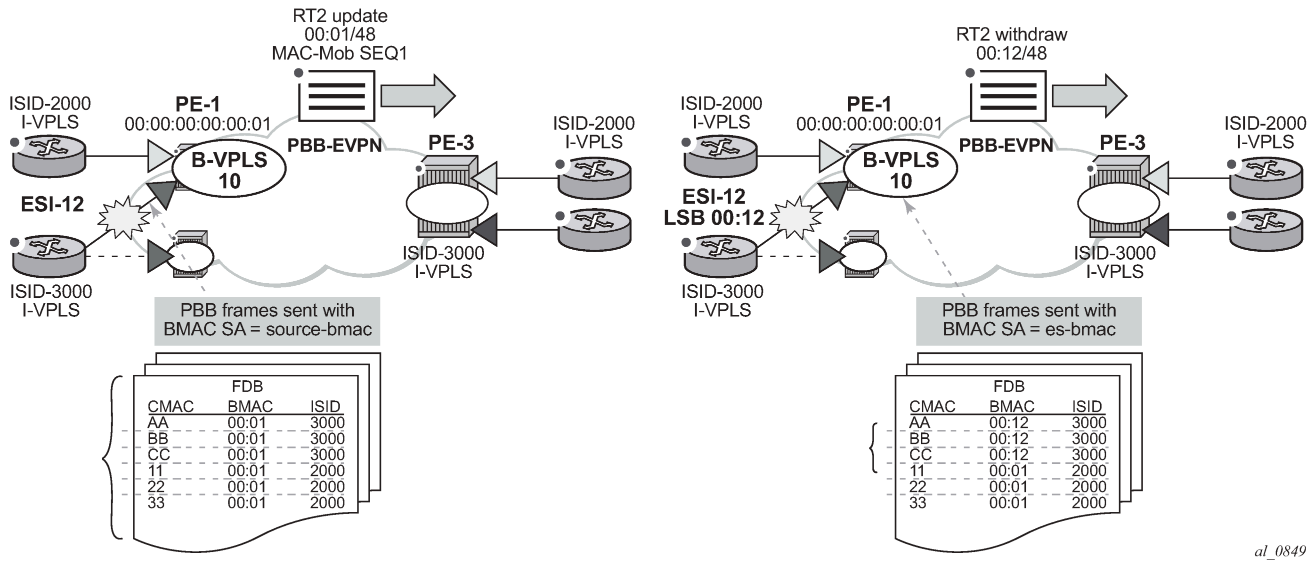
The system allows the following user choices per B-VPLS and ES:

- A dedicated ES BMAC address per ES can be used. In that case, the **pbb use-es-bmac** command is configured in the B-VPLS. In all-active multi-homing, all the PEs that are part of the ES source the PBB packets with the same source ES BMAC address; single-active multi-homing requires the use of a different ES BMAC address per PE.
- A non-dedicated source BMAC address can be used (this is only possible in single-active multi-homing). In this case, the user does not configure **pbb>use-es-bmac** and the regular source BMAC address is used for the traffic. A different source BMAC address has to be advertised per PE.

As discussed, single-active multi-homing can use source BMAC addresses or ES BMAC addresses. Using one type or another has a different impact on CMAC flushing, as illustrated in Figure 77: The use of ES BMAC to minimize CMAC flush.

- If ES BMAC addresses are used, as shown on the right-hand side of Figure 77: The use of ES BMAC to minimize CMAC flush, a less-impacting CMAC flush is achieved, therefore minimizing the flooding after ES failures. In the case of ES failure, PE-1 withdraws the ES BMAC address 00:12 and the remote PE-3 only flushes the CMACs associated with that ES BMAC address (only the CMAC addresses behind the CE are flushed).
- If source BMAC addresses are used, as shown on the left-hand side of Figure 77: The use of ES BMAC to minimize CMAC flush, in the case of ES failure, a BGP update with higher sequence number is issued by PE-1 and the remote PE-3 flushes all the CMAC addresses associated with the source BMAC address. Therefore, all the CMAC addresses behind the B-VPLS of the PEs will be flushed, as opposed to only the CMAC addresses behind the CE of the Ethernet Service Instances (ESIs).

Figure 77: The use of ES BMAC to minimize CMAC flush



FDB

CMAC	BMAC	ISID
AA	00:01	3000
BB	00:01	3000
CC	00:01	3000
11	00:01	2000
22	00:01	2000
33	00:01	2000

FDB

CMAC	BMAC	ISID
AA	00:12	3000
BB	00:12	3000
CC	00:12	3000
11	00:01	2000
22	00:01	2000
33	00:01	2000

Table 5: PBB-EVPN multi-homing supported combinations in SR OS shows the PBB-EVPN multi-homing combinations supported in the current release in the topology of Figure 76: PBB-EVPN multi-homing.

Table 5: PBB-EVPN multi-homing supported combinations in SR OS

CE Connectivity	PE Connectivity	PE Redundancy	BMAC Assignment	I-VPLS Support	Epipe Support
LAG (LACP optional)	LAG SAP	EVPN MH all-active	use-es-bmac (shared BMAC)	Yes	Yes
Ethernet ports (no LAG)	LAG SAP or port SAP	EVPN MH single-active	use-es-bmac (dedicated per PE)	Yes	No
Ethernet ports (no LAG)	LAG SAP or port SAP	EVPN MH single-active	source-bmac (dedicated per PE)	Yes	No
MPLS	spoke-SDP	EVPN MH single-active	source-bmac (dedicated per PE)	Yes	No
MPLS	spoke-SDP	EVPN MH single-active	use-es-bmac (dedicated per PE)	Yes	No

As an example, the configurations of the first, and last two, rows (LAG SAP all-active, MPLS source-BMAC, and MPLS ES-BMAC, respectively) will be discussed in the following three sections.

PBB-EVPN all-active multi-homing for I-VPLS and Epipes

Figure 76: PBB-EVPN multi-homing shows a PBB-EVPN network where ESI-23 is configured as an all-active multi-homing ES on PE-2 and PE-3. Two services are using ESI-23: I-VPLS 1001 and Epipe 1003. The following output shows the relevant configuration in PE-2:

```
# on PE-2:
configure
    service
        pbb
            mac-name "PE-5" 00:00:00:00:00:05
        exit
        system
            bgp-evpn
                ethernet-segment "ESI-23" create
                    esi 01:00:00:00:00:23:00:00:00:01
                    source-bmac-lsb 23-23 es-bmac-table-size 8
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 1
                    no shutdown
                exit
            exit
        exit
        vpls 1000 name "B-VPLS 1000" customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:00:00:00:00:02
                use-es-bmac
            exit
            split-horizon-group "CORE" create
            exit
            bgp
            exit
            bgp-evpn
                evi 1000
                mpls bgp 1
                    split-horizon-group "CORE"
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
        vpls 1001 name "I-VPLS 1001" customer 1 i-vpls create
            pbb
                backbone-vpls 1000
                exit
            exit
            stp
                shutdown
            exit
            sap lag-1:1001 create
                no shutdown
            exit
            no shutdown
        exit
        epipe 1003 name "Epipe 1003" customer 1 create
            pbb
                tunnel 1000 backbone-dest-mac "PE-5" isid 1003
            exit
            sap lag-1:1003 create
                no shutdown
            exit
            no shutdown
        exit
```

The following output shows the relevant configuration in PE-3:

```
# on PE-3:
configure
    service
        pbb
            mac-name "PE-5" 00:00:00:00:00:05
        exit
        system
            bgp-evpn
                ethernet-segment "ESI-23" create
                    esi 01:00:00:00:00:23:00:00:00:01
                    source-bmac-lsb 23-23 es-bmac-table-size 8
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 1
                    no shutdown
                exit
            exit
        exit
        vpls 1000 name "B-VPLS 1000" customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:00:00:00:00:03
                use-es-bmac
            exit
            split-horizon-group "CORE" create
            exit
            bgp
            exit
            bgp-evpn
                evi 1000
                mpls bgp 1
                    split-horizon-group "CORE"
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
        vpls 1001 name "I-VPLS 1001" customer 1 i-vpls create
            pbb
                backbone-vpls 1000
                exit
            exit
            stp
                shutdown
            exit
            sap 1/2/1:1001 create
                no shutdown
            exit
            sap lag-1:1001 create
                no shutdown
            exit
            no shutdown
        exit
        epipe 1003 name "Epipe 1003" customer 1 create
            pbb
                tunnel 1000 backbone-dest-mac "PE-5" isid 1003
            exit
            sap lag-1:1003 create
                no shutdown
            exit
            no shutdown
        exit
```

The preceding configuration shows that Epipe 1003 has a PBB tunnel pointing at the PE-5 source-BMAC. Epipe 1003 has the following configuration in PE-5 (the PBB tunnel points at the ESI-23 ES-BMAC):

```
# on PE-5:
configure
    service
        pbb
```

```
            mac-name "ES-MAC-23" 00:00:00:00:23:23
        exit
        epipe 1003 name "Epipe 1003" customer 1 create
            pbb
                tunnel 1000 backbone-dest-mac "ES-MAC-23" isid 1003
            exit
            sap 1/2/1:1003 create
                no shutdown
            exit
            no shutdown
        exit
```

Source-BMAC addresses and ES-BMAC addresses are distributed in BGP-EVPN. PE-2 and PE-3 will each advertise their own source-BMAC in a MAC route with ESI-0 and the shared ES-BMAC address with ESI-MAX (as per the RFC 7623). The ES-BMAC address that each PE uses in a B-VPLS is derived from the configured **service>(b)vpls>pbb>source-bmac** (four high-order bytes) and the ESI-23 configured **source-bmac-lsb**. In this example, PE-2 and PE-3 will both derive ES-BMAC 00:00:00:00:23:23. For both PEs to derive the required same ES-BMAC address, the four high-order bytes of the source-BMAC address must match on both PEs.

The **es-bmac-table-size** parameter modifies the default value (8) for the maximum number of ES-BMAC addresses that can be associated with the Ethernet segment across different B-VPLS services. When **source-bmac-lsb** is configured, the associated **es-bmac-table-size** is reserved out of the total FDB space.

The following outputs show the source-BMAC addresses and ES-BMAC address and how they are advertised and installed in the B-VPLS FDB.

```
*A:PE-2# show service system bgp-evpn ethernet-segment name "ESI-23" | match BMAC
Source BMAC LSB         : 23-23
```

The following output shows that ES-BMAC is used and that the operational source-BMAC is 00:00:00:00:00:02.

```
*A:PE-2# show service id 1000 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1000                Vpn Id            : 0
Service Type      : b-VPLS
---snip---
Oper Backbone Src : 00:00:00:00:00:02
Use SAP B-MAC     : Disabled
i-Vpls Count      : 1
Epipe Count       : 1
Use ESI B-MAC     : Enabled
---snip---
```

The source BMAC LSB is configured with the same value on PE-2 and PE-3. The two low-order bytes of the ES-BMAC will be 23:23.

```
*A:PE-3# show service system bgp-evpn ethernet-segment name "ESI-23" | match BMAC
Source BMAC LSB         : 23-23
```

On PE-3, ES-BMAC is used and the operational source BMAC is 00:00:00:00:00:03, as follows:

```
*A:PE-3# show service id 1000 base
```

```
===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1000                Vpn Id            : 0
Service Type      : b-VPLS
---snip---
Oper Backbone Src : 00:00:00:00:00:03
Use SAP B-MAC     : Disabled
i-Vpls Count      : 1
Epipe Count       : 2
Use ESI B-MAC     : Enabled
---snip---
```

On PE-2, the FDB for B-VPLS 1000 has an entry for each of the other PEs. PEs do not show their own system BMAC addresses in the FDB:

```
*A:PE-2# show service id 1000 fdb detail

===============================================================================
Forwarding Database, Service 1000
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1000       00:00:00:00:00:03 mpls:                   EvpnS:P  03/03/21 11:06:41
                             192.0.2.3:524277
           ldp:65537
1000       00:00:00:00:00:04 mpls:                   EvpnS:P  03/03/21 11:06:37
                             192.0.2.4:524280
           ldp:65538
1000       00:00:00:00:00:05 mpls:                   EvpnS:P  03/03/21 11:06:39
                             192.0.2.5:524280
           ldp:65539
-------------------------------------------------------------------------------
No. of MAC Entries: 3
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

On PE-4, the FDB for B-VPLS 1000 has an entry for each of the other PEs and an entry for the ES-BMAC address of ES "ESI-23":

```
*A:PE-4# show service id 1000 fdb detail

===============================================================================
Forwarding Database, Service 1000
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1000       00:00:00:00:00:02 mpls:                   EvpnS:P  03/03/21 11:06:42
                             192.0.2.2:524281
           ldp:65537
1000       00:00:00:00:00:03 mpls:                   EvpnS:P  03/03/21 11:06:41
                             192.0.2.3:524277
           ldp:65538
1000       00:00:00:00:00:05 mpls:                   EvpnS:P  03/03/21 11:06:39
                             192.0.2.5:524280
           ldp:65539
1000       00:00:00:00:23:23 eES:                    EvpnS:P  03/03/21 11:07:33
                             MAX-ESI
```

```
-------------------------------------------------------------------------------
No. of MAC Entries: 4
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

On PE-4, there are two BGP routes for ES-BMAC address 00:00:00:00:23:23: one with next hop PE-2 and the other with next hop PE-3, as follows:

```
*A:PE-4# show router bgp routes evpn mac mac-address 00:00:00:00:23:23
===============================================================================
 BGP Router ID:192.0.2.4        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.2:1000      00:00:00:00:23:23 ESI-MAX
      0                   Static            LABEL 524281
                          n/a
                          192.0.2.2

u*>i  192.0.2.3:1000      00:00:00:00:23:23 ESI-MAX
      0                   Static            LABEL 524277
                          n/a
                          192.0.2.3

-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

PBB-EVPN all-active multi-homing is based on the same concepts as EVPN-MPLS all-active multi-homing: DF election, split-horizon, and aliasing.

Designated forwarder (DF) election

Only the DF PE for an ISID will send multicast traffic to the ES. The following command shows that PE-3 is the DF PE in ES “ESI-23” for ISID 1003:

```
*A:PE-3# show service system bgp-evpn ethernet-segment name "ESI-23" isid 1003

===============================================================================
ISID DF and Candidate List
===============================================================================
Isid          SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
1003          1003          0                   yes 03/03/2021 11:07:45
===============================================================================
---snip---
```

The following command shows the DF and DF candidate list in the ES for all EVIs and all ISIDs:

```
*A:PE-3# show service system bgp-evpn ethernet-segment name "ESI-23" all

===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : ESI-23
Eth Seg Type            : None
Admin State             : Enabled            Oper State         : Up
ESI                     : 01:00:00:00:00:23:00:00:00:01
Multi-homing            : allActive          Oper Multi-homing  : allActive
---snip---

===============================================================================
ISID Information
===============================================================================
ISID                SvcId               Actv Timer Rem      DF
-------------------------------------------------------------------------------
1001                1001                0                   yes
1003                1003                0                   yes
-------------------------------------------------------------------------------
Number of entries: 2
===============================================================================

-------------------------------------------------------------------------------
DF Candidate list
-------------------------------------------------------------------------------
ISID                                    DF Address
-------------------------------------------------------------------------------
1001                                    192.0.2.2
1001                                    192.0.2.3
1003                                    192.0.2.2
1003                                    192.0.2.3
-------------------------------------------------------------------------------
Number of entries: 4
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
---snip---
```

The DF PE identifies multicast traffic by looking at either the destination BMAC or the EVPN label (which can be unicast or multicast).

In the case of Epipes, there are also DF and non-DF PEs. However, traffic is usually unicast (sent to the PBB tunnel backbone-destination BMAC). The non-DF PE will usually not discard Epipe traffic to the ES, unless the packet comes with an EVPN multicast label. To avoid packet duplication at the CE for Epipes, it is recommended to either:

- configure **discard-unknown** on all the B-VPLS instances where there are PBB-Epipes. This will prevent the ingress PE from flooding Epipe traffic if the PBB tunnel BMAC is unknown in the FDB.
- configure **ingress-replication-bum-label** so that, when the PBB tunnel BMAC is unknown in the FDB, the ingress PE sends traffic with a multicast label. The non-DF will discard traffic identified as multicast at Epipes.

Ethernet segment split-horizon

In PBB-EVPN all-active multi-homing, the split-horizon function is not based in the ESI label but in a source BMAC check. When BUM traffic is received on an I-VPLS, the PE will encapsulate it in PBB using the ES-

BMAC as source BMAC and the group BMAC for the ISID. When the DF PE for the ISID receives that packet, it will not send it back to the ES if the packet is identified as being originated from the ES itself (based on the ES-BMAC shared between the PEs).

Aliasing

Aliasing is based on the advertisement of the same ES-BMAC with MAX-ESI from the PEs part of the same ES. PE-2 and PE-3 advertise the ES-BMAC 00:00:00:00:23:23 with MAX-ESI (ESI = all FFs, as per the RFC 7623) and as Static (protected). When the remote PEs, PE-4, and PE-5, receive the two routes for the same BMAC and MAX-ESI, they will create a single EVPN-MPLS destination that will give more than one next-hop (in this case 2), as long as ECMP > 1:

```
*A:PE-4# show service id 1000 evpn-mpls

===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address     Egr Label     Num. MACs   Mcast           Last Change
                Transport:Tnl                             Sup BCast Domain
-------------------------------------------------------------------------------
192.0.2.2       524281        1           bum             03/03/2021 11:06:42
                ldp:65537                                 No
192.0.2.3       524277        1           bum             03/03/2021 11:06:41
                ldp:65538                                 No
192.0.2.5       524280        1           bum             03/03/2021 11:06:39
                ldp:65539                                 No
-------------------------------------------------------------------------------
Number of entries : 3
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                       Num. Macs               Last Change
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================

===============================================================================
BGP EVPN-MPLS ES BMAC Dest
===============================================================================
ES BMAC Addr                            Last Change
-------------------------------------------------------------------------------
00:00:00:00:23:23                       03/03/2021 11:07:49
-------------------------------------------------------------------------------
Number of entries: 1
-------------------------------------------------------------------------------
===============================================================================
```

The EVPN-MPLS ES BMAC destination has two next hops: PE-2 and PE-3.

```
*A:PE-4# show service id 1000 evpn-mpls es-bmac 00:00:00:00:23:23

===============================================================================
BGP EVPN-MPLS ES BMAC Dest
===============================================================================
ES BMAC Addr                            Last Change
-------------------------------------------------------------------------------
```

```
00:00:00:00:23:23                       03/03/2021 11:07:49
===============================================================================

===============================================================================
BGP EVPN-MPLS ES BMAC Dest TEP Info
===============================================================================
TEP Address               Egr Label               Last Change          Tunnel-
                          Transport                                    Id
-------------------------------------------------------------------------------
192.0.2.2                 524281                  03/03/2021 11:07:33  65537
                          ldp
192.0.2.3                 524277                  03/03/2021 11:07:49  65538
                          ldp
-------------------------------------------------------------------------------
Number of entries : 2
-------------------------------------------------------------------------------
===============================================================================
```

A similar output will be obtained in PE-5. Unicast traffic entering I-VPLS 1001 in either PE-4 or PE-5 will be hashed and load-balanced to PE-2 and PE-3 if the destination CMAC lookup yields an **es-bmac-dest**:

```
*A:PE-5# show service id 1001 fdb detail pbb

==============================================================================
Forwarding Database, i-Vpls Service 1001
==============================================================================
MAC               Source-Identifier     B-Svc     b-Vpls MAC        Type/Age
 Transport:Tnl-Id
------------------------------------------------------------------------------
00:00:10:10:10:10 eES-BMAC:             1000      00:00:00:00:23:23 L/0
                  00:00:00:00:23:23
00:00:30:30:30:30 b-mpls:               1000      00:00:00:00:00:03 L/0
                  192.0.2.3:524277
00:00:50:50:50:50 sap:1/2/1:1001        1000      N/A               L/0
00:00:60:60:60:60 sdp:56:1001           1000      N/A               L/0
==============================================================================
```

Verify the FDB of I-VPLS 1001 for ES BMAC destination 00:00:00:00:23:23 as follows:

```
*A:PE-5# show service id 1001 fdb evpn-mpls es-bmac-dest 00:00:00:00:23:23

===============================================================================
Forwarding Database, Service 1001
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1001       00:00:10:10:10:10 eES-BMAC:               L/30     03/03/21 11:13:15
                             00:00:00:00:23:23
-------------------------------------------------------------------------------
No. of Entries: 1
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

If a failure occurs in the ES, the PE will withdraw the ES-BMAC and the remote PEs will remove one next-hop of the ES-BMAC EVPN-MPLS destination.

For PBB-Epipes, aliasing will also work, as long as shared-queuing or policing are enabled on the ingress PE Epipe. In Figure 76: PBB-EVPN multi-homing, Epipe 1003 on PE-5 requires shared-queuing or policing at the ingress SAP. Otherwise, the traffic will be sent to only one PE of the ES (usually to the lower-IP PE).

For more information about the configuration of the Ethernet segment and its parameters, see the EVPN for MPLS Tunnels chapter.

PBB-EVPN single-active multi-homing for I-VPLS with source BMAC addresses

ESI-45 is a single-active Ethernet-segment (see Figure 76: PBB-EVPN multi-homing) with SDPs linked to it. As indicated in Table 5: PBB-EVPN multi-homing supported combinations in SR OS , only I-VPLS services can be used in this configuration. As described in section PBB-EVPN multi-homing, single-active ES and B-VPLS services can be configured to either use source-BMAC addresses or ES-BMAC addresses. The following configuration shows the former option on PE-4:

```
# on PE-4:
configure
    service
        sdp 46 mpls create
            far-end 192.0.2.6
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        system
            bgp-evpn
                ethernet-segment "ESI-45" create
                    esi 01:00:00:00:00:45:00:00:00:01
                    source-bmac-lsb 45-04 es-bmac-table-size 8
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 46
                    no shutdown
                exit
            exit
        exit
        vpls 1000 name "B-VPLS 1000" customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:00:00:00:00:04
            exit
            split-horizon-group "CORE" create
            exit
            bgp
            exit
            bgp-evpn
                evi 1000
                mpls bgp 1
                    split-horizon-group "CORE"
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
        vpls 1001 name "I-VPLS 1001" customer 1 i-vpls create
            pbb
                backbone-vpls 1000
                exit
            exit
            stp
                shutdown
            exit
            spoke-sdp 46:1001 create
            exit
            no shutdown
        exit
```

The configuration on PE-5 is similar:

```
# on PE-5:
configure
    service
        sdp 56 mpls create
            far-end 192.0.2.6
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        system
            bgp-evpn
                ethernet-segment "ESI-45" create
                    esi 01:00:00:00:00:45:00:00:00:01
                    source-bmac-lsb 45-05 es-bmac-table-size 8
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 56
                    no shutdown
                exit
            exit
        exit
        vpls 1000 name "B-VPLS 1000" customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:00:00:00:00:05
            exit
            split-horizon-group "CORE" create
            exit
            bgp
            exit
            bgp-evpn
                evi 1000
                mpls bgp 1
                    split-horizon-group "CORE"
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                no shutdown
            exit
            no shutdown
        exit
        vpls 1001 name "I-VPLS 1001" customer 1 i-vpls create
            pbb
                backbone-vpls 1000
                exit
            exit
            stp
                no shutdown
            exit
            sap 1/2/1:1001 create
                no shutdown
            exit
            spoke-sdp 56:1001 create
                no shutdown
            exit
            no shutdown
        exit
```

With the preceding configuration, PE-4 and PE-5 will not advertise ES-BMAC addresses with MAX-ESI. Therefore, all the remote BMACs on PE-2 and PE-3 are associated with regular backbone EVPN-MPLS destinations. The CMAC addresses will be learned in the data plane associated with local SAP/SDP-bindings or remote BMAC addresses. An example for the I-VPLS and B-VPLS FDB in PE-2 follows:

```
*A:PE-2# show service id 1001 fdb detail pbb

==============================================================================
Forwarding Database, i-Vpls Service 1001
==============================================================================
MAC               Source-Identifier     B-Svc     b-Vpls MAC        Type/Age
 Transport:Tnl-Id
------------------------------------------------------------------------------
00:00:10:10:10:10 sap:lag-1:1001        1000      N/A               L/90
00:00:60:60:60:60 b-mpls:               1000      00:00:00:00:00:05 L/90
                  192.0.2.5:524280
==============================================================================
```

The B-VPLS FDB on PE-2 looks as follows:

```
*A:PE-2# show service id 1000 fdb detail

===============================================================================
Forwarding Database, Service 1000
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1000       00:00:00:00:00:03 mpls:                   EvpnS:P  03/03/21 11:06:41
                             192.0.2.3:524277
           ldp:65537
1000       00:00:00:00:00:04 mpls:                   EvpnS:P  03/03/21 11:06:37
                             192.0.2.4:524280
           ldp:65538
1000       00:00:00:00:00:05 mpls:                   EvpnS:P  03/03/21 11:06:39
                             192.0.2.5:524280
           ldp:65539
-------------------------------------------------------------------------------
No. of MAC Entries: 3
-------------------------------------------------------------------------------
```

```
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
================================================================================
```

In the preceding example, the DF for ISID 1001 is PE-5. With a failure event on the SDP to MTU-6, PE-5 will not withdraw the advertised source BMAC (because it is still being used as source BMAC for other services and even CEs within the same service). PE-5 will send an update of the same source BMAC instead, increasing the sequence number in the MAC mobility extended community. That will be a **flush-all-from-me** indication for the remote PEs (they will flush all the CMACs associated with the updated source BMAC, regardless of the service).

When the former DF (PE-5) comes back up, PE-4 will become non-DF and will send a CMAC flush indication using the same mechanism as described above.

The following example shows a failure of SDP 56 in PE-5 and the corresponding DF switchover and CMAC flush.

```
# on PE-5:
186 2021/03/03 11:18:28.912 UTC MINOR: SVCMGR #2095 Base
"Ethernet Segment:ESI-45, ISID:1001, Designated Forwarding state changed to:false"

185 2021/03/03 11:18:28.912 UTC MINOR: SVCMGR #2303 Base
"Status of SDP 56 changed to admin=up oper=down"
```

PE-5 sends a BGP update with the same source BMAC, increasing the sequence number in the MAC mobility extended community—CMAC flush:

```
# on PE-5:
73 2021/03/03 11:18:28.912 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 89
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.5
        Type: EVPN-MAC Len: 33 RD: 192.0.2.5:1000 ESI: ESI-0, tag: 0, mac len: 48
                       mac: 00:00:00:00:00:05, IP len: 0, IP: NULL, label1: 8388480
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:1000
        bgp-tunnel-encap:MPLS
        mac-mobility:Seq:2/Static
"
```

Individual SAP or spoke-SDP failures do not trigger any MAC flush or DF re-election. This is as per RFC 7623. In EVPN-MPLS, individual SAP/spoke-SDP failures are captured by the AD per-EVI withdrawal, which triggers a DF switchover.

PBB-EVPN single-active multi-homing for I-VPLS with ES-BMACs

As discussed throughout this chapter, the use of ES-BMACs for single-active multi-homing can minimize the number of CMACs flushed in a network. A simple change is necessary: activate the **use-es-bmac** command and ensure that the generated ES-BMACs in PE-4 and PE-5 are different (the **source-bmac-lsb** in the previous configuration had different values for PE-4 and PE-5 already):

```
# on PE-4, PE-5:
configure
    service
        vpls "B-VPLS 1000"
            pbb
                use-es-bmac
```

On PE-4, the source BMAC LSB in ESI-45 is configured with a value of 45-04:

```
*A:PE-4# show service system bgp-evpn ethernet-segment name "ESI-45" | match BMAC
Source BMAC LSB         : 45-04
```

On PE-5, the source BMAC LSB in ESI-45 is configured with a value of 45-05:

```
*A:PE-5# show service system bgp-evpn ethernet-segment name "ESI-45" | match BMAC
Source BMAC LSB         : 45-05
```

The remote PEs (such as PE-2 in the following output) will receive additional BGP EVPN-MAC routes for the ES-BMACs. The following FDB on PE-2 shows the source BMAC addresses of PE-4 and PE-5 and the ES BMAC address of DF PE-5.

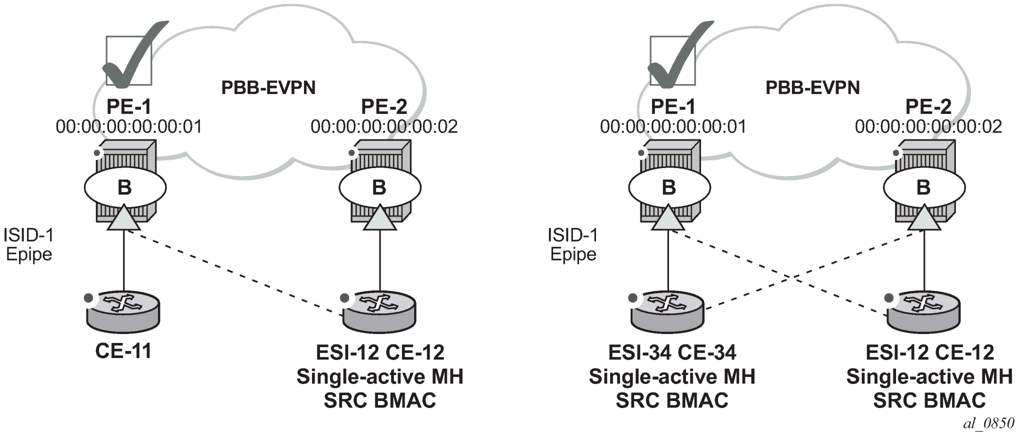
```
*A:PE-2# show service id 1000 fdb detail

===============================================================================
Forwarding Database, Service 1000
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1000       00:00:00:00:00:03 mpls:                   EvpnS:P  03/03/21 11:06:41
                             192.0.2.3:524277
           ldp:65537
1000       00:00:00:00:00:04 mpls:                   EvpnS:P  03/03/21 11:06:37
                             192.0.2.4:524280
           ldp:65538
1000       00:00:00:00:00:05 mpls:                   EvpnS:P  03/03/21 11:06:39
                             192.0.2.5:524280
           ldp:65539
1000       00:00:00:00:45:05 mpls:                   EvpnS:P  03/03/21 11:19:55
                             192.0.2.5:524280
           ldp:65539
-------------------------------------------------------------------------------
No. of MAC Entries: 4
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The benefit is that in case of a failure in ESI-45 (as before) the ES-BMAC is withdrawn and the remote PEs will only flush the CMACs associated with the remote ESI-45, as opposed to all the CMACs associated with PE-5.

PBB-EVPN single-active multi-homing for Epipes

In the network in Figure 76: PBB-EVPN multi-homing, Epipes can only support single-homing or all-active multi-homing but not single-active. For non-local-switching PBB-Epipes (there is a single SAP per Epipe), only all-active multi-homing is supported. Single-active multi-homing for local-switching enabled PBB-Epipes (two SAPs are defined within the PBB-Epipe instance) is only supported in the following scenarios.

Figure 78: PBB-EVPN single-active support for Epipes



Single-active multi-homing is supported for redundancy in a two-node, three or four SAP, scenario, as displayed in Figure 78: PBB-EVPN single-active support for Epipes. In these two cases, the Epipe PBB tunnel will be configured with the source BMAC of the remote PE node. When two SAPs are active in the same Epipe, local-switching is used to exchange frames between the CEs.

All-active multi-homing is not supported for redundancy in this scenario because the PE-1 PBB tunnel cannot point at a locally defined ES-BMAC.

PBB-EVPN multi-homing operation

See the EVPN for MPLS Tunnels chapter for the commands to operate Ethernet-segments. Consider that there are no AD routes in PBB-EVPN. Also, the DF election algorithm will be based on the ISID values as opposed to EVIs.

Troubleshooting and debug commands

When troubleshooting PBB-EVPN networks, most of the troubleshooting commands discussed in EVPN for MPLS Tunnels can be used in the B-VPLS service and the base **service>system>bgp-evpn** instance. Some examples of useful commands are:

- show redundancy bgp-evpn-multi-homing
- show router bgp routes evpn (and filters)
- show service evpn-mpls [<TEP ip-address>]
- show service id bgp-evpn
- show service id evpn-mpls (and modifiers)

- show service id fdb pbb (and modifiers)
- show service system bgp-evpn
- show service system bgp-evpn ethernet-segment (and modifiers)
- debug router bgp update
- log-id 99

In addition, the following **tools dump** commands also discussed in EVPN for MPLS Tunnels can help too:

- tools dump service evpn usage
- tools dump service system bgp-evpn ethernet-segment <name> isid <isid> df (Note: **isid** is used instead of **evi**.)

There are two aspects that are specific to PBB-EVPN and not EVPN:

1. Consumption of virtual BMAC addresses in the system— source BMACs, SAP BMACs, SDP BMACs, and ES BMACs are system BMACs that use FDB space but are not shown in the FDB together with the rest of the learned MAC addresses. The following command provides information about the virtual system MAC addresses consumed in the system.

```
*A:PE-3# tools dump redundancy src-bmac-lsb
Src-bmac-lsb:      3 (00-03) User: B-Vpls - 1 service(s)
Src-bmac-lsb:   8995 (23-23) User: Evpn Mpls

Total Src-bmac-lsbs = 2
```

2. Consumption of MFIBs — when ISIDs are not using the default-multicast list in the B-VPLS context for sending BUM traffic, an MFIB is consumed per ISID. The following command provides information about the consumption of MFIBs per system and per B-VPLS.

```
*A:PE-3# tools dump service vpls-pbb-mfib-stats detail

Service Manager VPLS PBB MFIB statistics at 03/03/2021 11:21:15:

Usage per Service
   ServiceId    MFIB User      Count
  ------------+--------------+-------
   1000         Evpn               1
  ------------+--------------+-------
                       Total       1

MMRP
  Current Usage     :     0
  System Limit      :  8191 Full, 40959 ESOnly
  Per Service Limit :  2048 Full,  8192 ESOnly

SPB
  Current Usage     :     0
  System Limit      :  8191
  Per Service Limit :  8191

Evpn
  Current Usage     :     1
  System Limit      : 40959
  Per Service Limit :  8191
```

Conclusion

In addition to a full RFC 7432 EVPN-MPLS implementation, SR OS supports PBB-EVPN as per RFC 7623 for large Layer 2 deployments, including single-active and all-active multi-homing. This example has shown how to configure and operate a PBB-EVPN network focusing on the specific aspects of PBB-EVPN compared to EVPN-MPLS.

EVPN for VXLAN Tunnels (Layer 2)

This chapter provides information about Ethernet Virtual Private Network (EVPN) for Virtual eXtensible Local Area Network (VXLAN) tunnels in VPLS services.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter is applicable to SR OS and was initially written for SR OS Release 12.0.R4. The CLI in the current edition is based on SR OS Release 21.2.R1. Ethernet Virtual Private Network (EVPN) is a control plane technology and does not have line card hardware dependencies.

Overview

SR OS supports the EVPN control plane with Virtual eXtensible Local Area Network (VXLAN) data plane in VPLS services.

EVPN (RFC 7432) is an IETF technology that uses a dedicated BGP address family which allows VPLS services to be operated in a similar way to IP-VPNs, where the MAC addresses, IP addresses, and the information to set up the flooding tree are distributed by BGP. EVPN can be used as the control plane for different data plane encapsulations, such as VXLAN and MPLS.

VXLAN (RFC 7348) is an overlay IP tunneling technology used to carry Ethernet traffic over any IP network, and it is becoming the de facto standard for overlay data centers and networks. Compared to other IP overlay tunneling technologies, such as GRE, VXLAN supports multi-tenancy and multi-pathing:

- A tenant identifier, the VXLAN Network Identifier (VNI), is encoded in the VXLAN header and allows each tenant to have an isolated Layer 2 domain.
- VXLAN supports multi-pathing scalability through ECMP. VXLAN uses the outer source UDP port as an entropy field that can be used by the core IP routers to balance the load across different paths.

In SR OS, EVPN and VXLAN can be enabled in VPLS or R-VPLS services. In this chapter, EVPN-VXLAN services will refer to VPLS or R-VPLS services with EVPN and VXLAN enabled. These services can terminate/originate VXLAN tunnels and may have SAPs and/or SDP bindings at the same time. Some other SR OS implementation-specific considerations are the following:

- VXLAN is only supported on network or hybrid ports on Ethernet/LAG/POS/APS interfaces.
- VXLAN packets are originated/terminated with the system IPv4 address, in other words, a system originating VXLAN packets will use the system IP address as source outer IPv4 address and systems

will only process VXLAN packets if their destination outer IPv4 address matches its own system IP address.

- Data plane MAC learning is not supported over VXLAN bindings. Only the control plane (EVPN) will be used for populating the FDB with MAC addresses associated to VXLAN bindings.
- EVPN provides support for the following features that are described in this chapter:
  - The BGP advertisement of the MAC addresses learned on SAPs, SDP-bindings and conditional static MACs to the remote BGP peers. The advertisement of MAC addresses in BGP can optionally be disabled.
  - The optional advertisement of an unknown MAC route, that allows the remote EVPN PEs or Network Virtualization Edge devices (NVEs) to suppress the unknown unicast flooding and send any unknown unicast frame to the owner of the unknown MAC route.
  - Ingress replication of Broadcast, Unknown unicast, and Multicast (BUM) packets over VXLAN.
  - A proxy-ARP table per service populated by the MAC-IP pairs received in BGP MAC advertisements. When an ARP request is received on a SAP or SDP-binding, the system will perform a lookup on this table and will reply to the ARP request if the lookup yields a valid result.
  - MAC mobility and static-MAC protection as described in RFC 7432, as well as MAC duplication detection.
- Multi-homing redundancy for SAPs and SDP-bindings in EVPN-VXLAN services is supported through BGP Multi-homing (L2VPN BGP address family). Only one BGP-MH site is supported in an EVPN-VXLAN service.

One of the main applications for EVPN-VXLAN services in SR OS is the Data Center Gateway (DC GW) function. In such an application, EVPN and VXLAN are expected to be used within the data center and VPLS SDP-bindings or SAPs are expected to be used for the connectivity to the WAN. When the system is used as a DC GW, a VPLS service is configured per Layer 2 domain that has to be extended to the WAN. In those VPLS services, BGP EVPN automatically sets up the VXLAN auto-bindings that connect the DC GW to the data center Network Virtual Edge devices (NVEs). The WAN connectivity is based on regular VPLS constructs where SAPs (null, dot1q, and QinQ), spoke-SDPs (FEC type 128 and 129, BGP-VPLS), and mesh-SDPs are supported. B-VPLS or I-VPLS services are not supported.

Although the DC GW application is one of the most common uses for this feature, this chapter focuses on the configuration and operation of EVPN-VXLAN for Layer 2 services in general, and its integration with regular VPLS services in MPLS networks.

Configuration

This section describes the configuration of EVPN-VXLAN on SR OS as well as the available troubleshooting and show commands. This example focuses on the following configuration aspects:

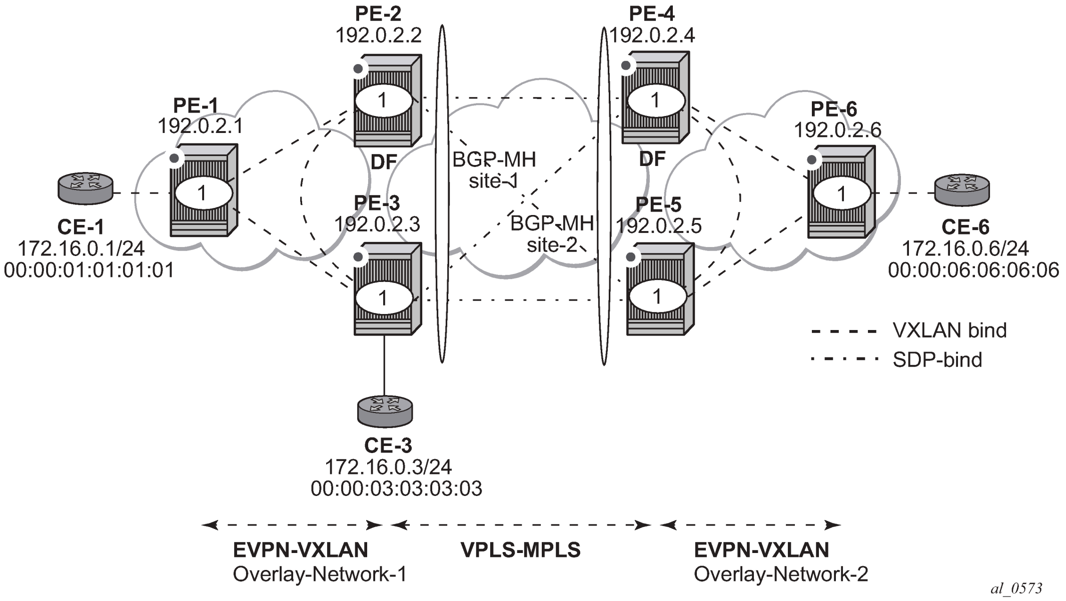
- Enabling EVPN and VXLAN in a VPLS service, including the use of BGP-EVPN, BGP Auto-discovery (BGP-AD), and BGP-Multi-homing (BGP-MH) in the same VPLS instance.
- Scaling BGP-MH resiliency with the use of operational groups (oper-groups).
- Use of proxy-ARP in EVPN-VXLAN services
- MAC mobility, MAC duplication, and MAC protection in EVPN-VXLAN services.

The configuration will be shown for PE-1, PE-2, and PE-3 only; the PEs in Overlay-Network-2 (Figure 79: EVPN-VXLAN example topology) have an equivalent configuration.

Enabling EVPN-VXLAN in a VPLS service

Figure 79: EVPN-VXLAN example topology shows the topology used in this example.

Figure 79: EVPN-VXLAN example topology



The example topology shows two overlay (VXLAN) networks interconnected by an MPLS network:

- PE-1, PE-2, and PE-3 are part of Overlay-Network-1
- PE-4, PE-5, and PE-6 are part of Overlay-Network-2

CE-1, CE-3, and CE-6 belong to the same IP subnet, therefore, Layer 2 connectivity must be provided to them.

The example topology can illustrate a Data Center Interconnect (DCI) example, where Overlay-Network-1 and Overlay-Network-2 are two data centers interconnected through an MPLS WAN. In this application, CE-1, CE-3, and CE-6 simulate virtual machines or appliances, PE-2/3/4/5 act as DC GWs and PE-1/6 as NVEs (or virtual PEs running on compute infrastructure).

The following protocols and objects are configured beforehand:

- The ports interconnecting the six PEs in Figure 79: EVPN-VXLAN example topology are configured as network ports (or hybrid) and have router network interfaces defined on them. Only the ports connected to the CEs are configured as access ports.
- The six PEs shown in the Figure 79: EVPN-VXLAN example topology are running IS-IS for the global routing table with the four core PEs interconnected using IS-IS Level-2 point-to-point interfaces and each overlay network is using IS-IS Level-1 point-to-point interfaces.
- LDP is used as the MPLS protocol to signal transport tunnel labels among PE-2, PE-3, PE-4, and PE-5. There is no LDP running in the two overlay networks.

- The network port MTU (in all the ports sending/receiving VXLAN packets) must be at least 50 bytes (54 if dot1q encapsulation is used) greater than the service MTU in order to accommodate the size of the VXLAN header.

Once the IGP infrastructure and LDP are enabled in the core, BGP has to be configured. In this example, two BGP families have to be enabled: EVPN within each overlay network for the exchange of MAC/IP addresses and setting up the flooding domains, and L2-VPN for the use of BGP-MH and BGP-AD in the VPLS-MPLS network.

As an example, the following CLI output shows the relevant BGP configuration of PE-1, which only needs the EVPN family. PE-6 would have a similar BGP configuration. The use of route reflectors (RRs) in this type of scenarios is common. Although this example does not use RRs, an EVPN RR could have been used in Overlay-Network-1 and Overlay-Network-2 and an L2-VPN RR could have been used in the core VPLS-MPLS network.

```
# on PE-1:
configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            rapid-update evpn
            group "DC"
                family evpn
                peer-as 64500
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
        exit
```

The BGP configuration on PE-2 is as follows:

```
# on PE-2:
configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            rapid-update l2-vpn evpn
            group "DC"
                family l2-vpn evpn
                peer-as 64500
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.3
                exit
            exit
            group "WAN"
                family l2-vpn
                peer-as 64500
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
        exit
```

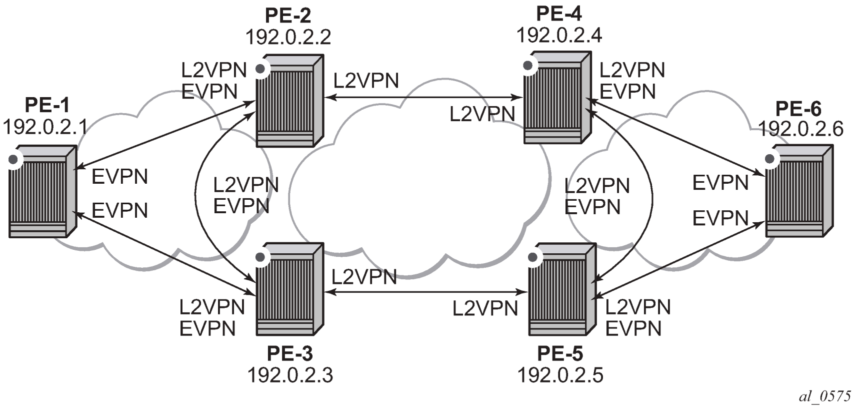
The BGP configuration on PE-3 is as follows:

```
# on PE-3:
configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            rapid-update l2-vpn evpn
            group "DC"
                family l2-vpn evpn
                peer-as 64500
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.2
                exit
            exit
            group "WAN"
                family l2-vpn
                peer-as 64500
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
        exit
```

The BGP configuration on PE-4 and PE-5 is equivalent.

Figure 80: BGP adjacencies and enabled families shows the BGP peering sessions among the PEs and the enabled BGP families. PE-1 will only establish an EVPN peering session with its peers (only the EVPN family is enabled on PE-1), even though PE-2 and PE-3 have EVPN and L2-VPN families configured.

Figure 80: BGP adjacencies and enabled families



Once the network infrastructure is running properly, the actual service configuration can be carried out. The following CLI outputs show the configuration of VPLS 1 in PE-1, PE-2, and PE-3 as per the topology illustrated in Figure 79: EVPN-VXLAN example topology.

VPLS 1 in those three PEs are interconnected using VXLAN bindings, whereas PE-2 and PE-3 are connected to the remote PEs by means of BGP-AD SDP-bindings. Although BGP-AD SDP-bindings are used in this example for the connectivity of the EVPN-VXLAN PEs to a regular VPLS network, SAPs, BGP-VPLS spoke-SDPs, manual spoke-SDPs, or mesh-SDPs could have been used instead.

VPLS 1 is configured on PE-1, as follows:

```
# on PE-1:
configure
    service
        vpls 1 name "VPLS1" customer 1 create
            vxlan instance 1 vni 1 create
            exit
            bgp
                route-distinguisher 64500:1
                route-target export target:64500:12 import target:64500:12
            exit
            bgp-evpn
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/2/1:1 create
                no shutdown
            exit
            no shutdown
```

EVPN-VXLAN is enabled by the configuration of a valid VXLAN Network Identifier (VNI) and the **bgp-evpn vxlan no shutdown** command. These two commands, along with the required BGP Route Distinguisher (RD) and Route Target (RT) information, are the minimum mandatory attributes:

- The VNI is a 24-bit identifier with valid values in the [1..16777215] range. This defines the VNI that SR OS will use in the EVPN routes generated for the VPLS service, and therefore the VNI that the system expects to see in the VXLAN packets destined to that particular VPLS service. The configured VNI determines the VNI that has to be received in the packets for the VPLS service, but not the VNI that will be sent in VXLAN packets to remote PEs for the service. In other words, in this example, VPLS 1 is configured with VNI=1 in all the PEs; however, each PE could have used a different VNI. The VNI is a system-wide significant value and two VPLS services cannot be configured with the same VNI.
- The **bgp-evpn vxlan no shutdown** command enables the use of EVPN for VXLAN. It requires the previous configuration of the VNI, RD, and RT. As soon as this command is executed, EVPN will advertise an inclusive multicast route to all of the BGP EVPN peers (regardless of the existing SAP/SDP-binding operational status). The exchange of inclusive multicast routes allows the establishment of the VXLAN bindings among the PEs.

Upon the reception of the EVPN inclusive multicast routes from PE-2 and PE-3, PE-1 will automatically set up its VXLAN bindings for VPLS 1. A VXLAN binding is represented by an (egress VTEP, egress VNI) pair, where VTEP is a VXLAN Termination End Point. This can be shown with the following show commands on PE-1:

```
*A:PE-1# show service vxlan

===============================================================================
VXLAN Tunnel Endpoints (VTEPs)
===============================================================================
VTEP Address                                 VXLAN Dest        ES Dest
```

```
-------------------------------------------------------------------------------
192.0.2.2                                    1                 0
192.0.2.3                                    1                 0
-------------------------------------------------------------------------------
Number of VTEPs: 2
-------------------------------------------------------------------------------
===============================================================================
```

```
*A:PE-1# show service id 1 vxlan instance 1 destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
Instance    VTEP Address                                Egress VNI  EvpnStatic Num
 Mcast       Oper State                                  L2 PBR     SupBcasDom MACs
-------------------------------------------------------------------------------
1           192.0.2.2                                   1           evpn       1
 BUM         Up                                          No          No
1           192.0.2.3                                   1           evpn       1
 BUM         Up                                          No          No
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 2
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

To actually see this output, the VPLS service needs to be configured on all PEs, with import and export policy “vsi-policy-1” defined on the core PEs; see further. As can be seen in the CLI output, PE-1 has two VXLAN bindings: one to PE-2 and one to PE-3. Both use egress VNI=1 (the actual VNI used in its egress VXLAN packets) and both are part of the flooding multicast list (BUM) for VPLS 1 and are up. There is no layer 2 Policy-Based Routing (L2 PBR).

- The **Mcast**=*BUM* entry is set when the proper inclusive multicast route is received from the remote VTEP. The VXLAN binding will be used to flood BUM packets.
- The **Oper State** is based on the existence of the VTEP in the global routing table.

The VPLS 1 configuration of PE-2 and PE-3 is as follows:

```
# on PE-2:
configure
    service
        pw-template 1 name "PW1" create
        exit
        vpls 1 name "VPLS1" customer 1 create
            vxlan instance 1 vni 1 create
            exit
            bgp
                route-distinguisher 192.0.2.2:1
                vsi-export "vsi-policy-1"
                vsi-import "vsi-policy-1"
                pw-template-binding 1 split-horizon-group "CORE"
                exit
            exit
            bgp-ad
```

```
                vpls-id 64500:1
                no shutdown
            exit
            bgp-evpn
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            site "site-1" create
                site-id 1
                split-horizon-group CORE
                no shutdown
            exit
            no shutdown
```

```
# on PE-3:
configure
    service
        pw-template 1 name "PW1" create
        exit
        vpls 1 name "VPLS1" customer 1 create
            vxlan instance 1 vni 1 create
            exit
            bgp
                route-distinguisher 192.0.2.3:1
                vsi-export "vsi-policy-1"
                vsi-import "vsi-policy-1"
                pw-template-binding 1 split-horizon-group "CORE"
                exit
            exit
            bgp-ad
                vpls-id 64500:1
                no shutdown
            exit
            bgp-evpn
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            site "site-1" create
                site-id 1
                split-horizon-group CORE
                no shutdown
            exit
            sap 1/2/1:1 create
            exit
            no shutdown
```

In addition to the VNI and **bgp-evpn vxlan no shutdown** commands for enabling EVPN-VXLAN in VPLS 1, PE-2 and PE-3 require the configuration of BGP-AD for the discovery and establishment of FEC129 spoke-SDPs to the remote PEs in the core, as well as BGP-MH for redundancy. As outlined in Figure 79: EVPN-VXLAN example topology, there are two BGP-MH sites defined in the network: site-1 is used on PE-2/PE-3 and site-2 is used on PE-4/PE-5. Only one of the two gateway PEs in each overlay network will be the designated forwarder (DF) for VPLS 1, and only the DF will send/receive traffic for VPLS 1 in the overlay network. The following considerations must be taken into account when configuring the connectivity of EVPN-VXLAN services to regular VPLS objects:

- As discussed, in this example, BGP-AD spoke-SDPs are used, but SAPs, BGP-VPLS spoke-SDPs, manual spoke-SDPs, or mesh-SDPs are also supported.
- In this example, BGP-AD spoke-SDPs are auto-instantiated using **pw-template-binding 1 split-horizon-group "CORE".**
  - This requires the creation of the pw-template 1 (**config service pw-template 1 name <..> create**).
- The split-horizon group CORE is added to the BGP-MH site "site-1". This statement will ensure that all the spoke-SDPs automatically established to the remote PEs are part of the BGP-MH site.
- Although the route targets for the overlay network and the VPLS-MPLS network can have the same value for the same VPLS service, they are usually different. This example assumes the use of RT-DC-1 in Overlay-Network-1 and RT-WAN-1 in the VPLS-MPLS core for VPLS 1. The "vsi-policy-1" allows the system to export and import the right RTs for VPLS 1 on the core PEs:

```
# on PE-2 and PE-3:
configure
    router Base
        policy-options
            begin
            community "RT-DC-1"
                members "target:64500:12"
            exit
            community "RT-WAN-1"
                members "target:64500:11"
            exit
            policy-statement "vsi-policy-1"
                entry 10     # to import all the EVPN routes with RT-DC-1
                    from
                        community "RT-DC-1"
                        family evpn
                    exit
                    action accept
                    exit
                exit
                entry 20     # to import all the BGP-AD/MH routes from the WAN
                    from
                        community "RT-WAN-1"
                        family l2-vpn
                    exit
                    action accept
                    exit
                exit
                entry 30      # to export all the EVPN routes with "RT-DC-1"
                    from
                        family evpn
                    exit
                    action accept
                        community add "RT-DC-1"
                    exit
                exit
                entry 40      # to export all the BGP-AD/MH routes with "RT-WAN-1"
                    from
                        family l2-vpn
                    exit
                    action accept
                        community add "RT-WAN-1"
                    exit
                exit
                default-action drop
                exit
            exit
            commit
```

Once PE-2 and PE-3 are configured as shown, they will set up the spoke-SDPs and will run the DF election algorithm to determine the operational status of those spoke-SDPs. See chapters LDP VPLS Using BGP Auto-Discovery and BGP Multi-Homing for VPLS Networks for more information about the use of BGP-AD and BGP-MH.

In the configuration for VPLS 1, both gateway PEs, PE-2 and PE-3, will attempt to establish two parallel Layer 2 paths between each other (a BGP-AD spoke-SDP and an EVPN VXLAN binding). Because that would create a Layer 2 loop, the SR OS implementation gives priority to the EVPN path and only the VXLAN binding will be active. In other words, when a VXLAN (egress VTEP, VNI) and a spoke-SDP are attempted to be set up to the same far-end IP address at the same time, the VXLAN path will prevail and the spoke-SDP will be kept down. The spoke-SDP will only be brought up if the VXLAN (egress VTEP, VNI) goes down.

This behavior can be easily observed in this setup by using the following **show** commands. In PE-2, the spoke-SDP to far-end PE-3 will be down with a *EvpnRouteConflict* Flag. The (egress VTEP, VNI) = (192.0.2.3, 1) VXLAN bind will be up.

```
*A:PE-2# show service id 1 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : VPLS
---snip---

Admin State       : Up                  Oper State        : Up
MTU               : 1514
SAP Count         : 0                   SDP Bind Count    : 3
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sdp:32765:4294967293 SB(192.0.2.5)       BgpAd        0       8978    Up   Up
sdp:32766:4294967294 SB(192.0.2.4)       BgpAd        0       8978    Up   Up
sdp:32767:4294967295 SB(192.0.2.3)       BgpAd        0       8978    Up   Down
===============================================================================
```

```
*A:PE-2# show service id 1 all | match Flag context all
Flags              : None
Flags              : None
Flags              : PWPeerFaultStatusBits
                     EvpnRouteConflict
```

```
*A:PE-2# show service id 1 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
Instance    VTEP Address                            Egress VNI  EvpnStatic Num
 Mcast       Oper State                              L2 PBR     SupBcasDom MACs
```

```
-------------------------------------------------------------------------------
1           192.0.2.1                              1           evpn       1
 BUM         Up                                     No          No
1           192.0.2.3                              1           evpn       1
 BUM         Up                                     No          No
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 2
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

At the non-DF, PE-3, all the spoke-SDPs will be down due to BGP-MH:

```
*A:PE-3# show service id 1 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : VPLS
---snip---

Admin State       : Up                  Oper State        : Up
MTU               : 1514
SAP Count         : 1                   SDP Bind Count    : 3
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/2/1:1                              q-tag        1518    1518    Up   Up
sdp:32765:4294967293 SB(192.0.2.2)       BgpAd        0       8978    Up   Down
sdp:32766:4294967294 SB(192.0.2.5)       BgpAd        0       8978    Up   Down
sdp:32767:4294967295 SB(192.0.2.4)       BgpAd        0       8978    Up   Down
===============================================================================
```

```
*A:PE-3# show service id 1 all | match Flag context all
Flags              : StandbyForMHProtocol
                     PWPeerFaultStatusBits
                     EvpnRouteConflict
Flags              : StandbyForMHProtocol
Flags              : StandbyForMHProtocol
Flags              : None
```

MAC learning and unknown-mac-route

Once the VPLS service (VPLS 1) is configured, the network allows the CEs to exchange unicast and BUM traffic over the overlay and VPLS-MPLS service infrastructure. BUM traffic sent by CE-1 will be ingress-replicated by PE-1 to PE-2 and PE-3, and propagated by PE-2 (the DF) to the remote network. From this point on, MAC addresses will be learned on active SAPs and spoke-SDPs and advertised in EVPN MAC routes. No data plane MAC learning is carried out on VXLAN bindings. MACs associated with (egress VTEP, VNI) bindings will always be learned through EVPN.

The following CLI output shows the reception of an EVPN MAC route on PE-1 and how the (CE-3) MAC address appears in the FDB for VPLS 1.

```
# on PE-1:
11 2021/02/10 15:48:52.094 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 88
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-MAC Len: 33 RD: 192.0.2.3:1 ESI: ESI-0, tag: 0, mac len: 48
                       mac: 00:00:03:03:03:03, IP len: 0, IP: NULL, label1: 1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:12
        bgp-tunnel-encap:VXLAN
"
```

```
*A:PE-1# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:01:01:01:01 sap:1/2/1:1             L/120    02/10/21 15:47:35
1          00:00:03:03:03:03 vxlan-1:                Evpn     02/10/21 15:48:52
                             192.0.2.3:1
1          00:00:06:06:06:06 vxlan-1:                Evpn     02/10/21 15:52:31
                             192.0.2.2:1
-------------------------------------------------------------------------------
No. of MAC Entries: 3
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

When a frame destined to 00:00:03:03:03:03 enters SAP 1/2/1:1, it is encapsulated into a VXLAN packet with outer destination IP 192.0.2.3 and VNI 1, and sent on the wire.

In virtualized data center networks where all the MACs are known beforehand (all the virtual machine and appliance MACs are distributed by EVPN before any traffic flows), unknown MAC addresses are always outside the data center. If that is the case, the DC GWs can make use of the **unknown-mac-route** so that the DC NVEs supporting the concept of this route send the unknown unicast traffic only to the DC GW. This minimizes the flooding within the data center, as explained in draft-ietf-bess-dci-evpn-overlay.

In this example, the unknown MAC route is configured in the gateway PEs (in Overlay-Network-1: PE-2 and PE-3) in the following way:

```
# on PE-2, PE-3:
configure
    service
        vpls "VPLS1"
```

```
            bgp-evpn
                unknown-mac-route
            exit
```

```
# on PE-2:
47 2021/02/10 16:00:36.068 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 88
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-MAC Len: 33 RD: 192.0.2.2:1 ESI: ESI-0, tag: 0, mac len: 48
                       mac: 00:00:00:00:00:00, IP len: 0, IP: NULL, label1: 1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:12
        bgp-tunnel-encap:VXLAN
"
```

Note that:

- Although SR OS can generate the unknown MAC route, it will never honor it and normal flooding applies when an unknown unicast packet arrives at an ingress SAP/SDP-binding.
- When **unknown-mac-route** is configured, it will only be generated when: a) no BGP-MH site is configured within the same VPLS service or b) a site is configured and the site is DF (Designated Forwarder) in the PE. If the site becomes a non-DF site, the unknown MAC route will be withdrawn.
- If the **unknown-mac-route** is used in the DC GW and all the NVEs in the DC understand it, the advertisement of MAC addresses can be disabled with the **[no] mac-advertisement** command. If so, SR OS will only advertise the unknown MAC route.

```
# on DC GWs PE-2 and PE-3:
configure
    service
        vpls "VPLS1"
            bgp-evpn
                unknown-mac-route
                no mac-advertisement
            exit
```

Scaling BGP-MH resiliency with the use of operational groups

In Figure 79: EVPN-VXLAN example topology, VPLS 1 in PE-2 and PE-3 is configured with a BGP-MH site that controls which of the two PEs forwards the traffic to the remote PEs (in this case, PE-2 is the DF and the GW responsible for forwarding packets to the remote PEs).

When new VPLS services are required in PE-2 and PE-3, the same BGP-MH configuration can be used. However, if the number of VPLS services grows significantly, the use of individual BGP-MH sites per service will not scale. Because all the services in these two PEs share the same physical topology, the use of operational groups can provide a simple and scalable way of providing resiliency to as many services as the user needs (up to the maximum number of VPLS services per system).

The way operational groups can be used to scale this type of deployments is the following (using the network topology in Figure 79: EVPN-VXLAN example topology and focusing on Overlay-Network-1):

- A control-VPLS service is defined in PE-2 and PE-3. For instance, VPLS 1.
  - This service is configured with a BGP-MH site in both PEs.
  - An oper-group "control-vpls-1" is created and associated to the pw-template-binding 1 in VPLS 1.
- Data VPLS services are defined in both PEs. For instance: VPLS 2, VPLS 3,... VPLS 999.
  - In all these services, the pw-template-binding is configured with **monitor-oper-group "control-vpls-1"**.
  - The status of the spoke-SDPs in the data VPLS services depends on the status of the operational group. If there is a DF switchover in VPLS 1 and VPLS 1 spoke-SDPs go down on PE-2, all the spoke-SDPs in all the data VPLS services controlled by "control-vpls-1" in PE-2 will go down too. In the same way, the spoke-SDPs in PE-3 will come up.
- To allow per-service load balancing, a second control-VPLS service with a different BGP-MH site should be configured.
  - For instance, VPLS 1 might have PE-2 as the DF and VPLS 1000 might be a second control-VPLS service with PE-3 as the DF.
  - Each control-VPLS would control a group of data VPLS services based on the definition and association of a second operational group.

The following example shows the modification of VPLS 1 as the control-VPLS and the configuration of VPLS 2 as a data-VPLS on PE-2. VPLS 1 controls the VPLS 2 spoke-SDP status.

```
# on PE-2:
configure
    service
        oper-group "control-vpls-1" create
        exit
        vpls 1 name "VPLS1" customer 1 create
            description "control-VPLS"
            bgp
                pw-template-binding 1 split-horizon-group "CORE"
                    oper-group "control-vpls-1"
                exit
            exit
        exit
        vpls 2 name "VPLS2" customer 1 create
            description "data-VPLS"
            vxlan instance 1 vni 2 create
            exit
            bgp
                route-distinguisher 192.0.2.2:2
                vsi-export "vsi-policy-2"
                vsi-import "vsi-policy-2"
                pw-template-binding 1
                    monitor-oper-group "control-vpls-1"
                exit
            exit
            bgp-ad
                vpls-id 64500:2
                no shutdown
            exit
            bgp-evpn
                unknown-mac-route
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

Use of proxy-ARP in EVPN-VXLAN services

EVPN-VXLAN services support proxy-ARP functionality that is enabled by the **proxy-arp [no] shutdown** command. By default, proxy-ARP is disabled. When proxy-ARP is enabled, the following applies:

- MAC and IP addresses contained in the received valid EVPN MAC routes are populated in the proxy-ARP table.
- ARP-request messages received on SAPs and SDP-bindings are intercepted and the target IP address is looked up. If the IP address is found, an ARP reply will be issued based on the information found in the proxy-ARP table, otherwise the ARP request would be flooded in the VPLS service (except for the source SAP/SDP binding).
- ARP-reply messages received on SAPs and SDP-bindings are also intercepted and sent to the CPM. These ARP-reply messages are re-injected in the data plane and forwarded based on the FDB information to the destination MAC address. If the destination MAC address is not in the FDB, the ARP-reply message will be flooded in the VPLS service (except for the source SAP/SDP binding).

The following CLI output shows the proxy-ARP configuration in PE-3 and a received valid MAC route that includes the MAC address 00:00:01:01:01:01 and IP address 172.16.0.1 of CE-1. This MAC-IP pair is installed in the proxy-ARP table for VPLS 1.

```
# on PE-3:
configure
    service
        vpls "VPLS1"
            proxy-arp
                no shutdown
            exit
```

```
# on PE-3:
120 2021/02/10 16:12:53.542 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 127
    Flag: 0x90 Type: 14 Len: 83 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.1
        Type: EVPN-MAC Len: 37 RD: 192.0.2.1:1 ESI: ESI-0, tag: 0, mac len: 48
                  mac: 00:00:01:01:01:01, IP len: 4, IP: 172.16.0.1, label1: 1
        Type: EVPN-MAC Len: 33 RD: 192.0.2.1:1 ESI: ESI-0, tag: 0, mac len: 48
                  mac: 00:00:01:01:01:01, IP len: 0, IP: NULL, label1: 1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:12
        bgp-tunnel-encap:VXLAN
"
```

This MAC-IP pair is installed in the proxy-ARP table for VPLS 1 on PE-3, as follows:

```
*A:PE-3# show service id 1 proxy-arp detail
-------------------------------------------------------------------------------
Proxy Arp
-------------------------------------------------------------------------------
Admin State       : enabled
Dyn Populate      : disabled
Age Time          : disabled            Send Refresh      : disabled
Table Size        : 250                 Total             : 1
Static Count      : 0                   EVPN Count        : 1
Dynamic Count     : 0                   Duplicate Count   : 0

Dup Detect
-------------------------------------------------------------------------------
Detect Window     : 3 mins              Num Moves         : 5
Hold down         : 9 mins
Anti Spoof MAC    : None

EVPN
-------------------------------------------------------------------------------
Garp Flood        : enabled             Req Flood         : enabled
Static Black Hole : disabled
EVPN Route Tag    : 0
-------------------------------------------------------------------------------

===============================================================================
VPLS Proxy Arp Entries
===============================================================================
IP Address          Mac Address         Type      Status    Last Update
-------------------------------------------------------------------------------
172.16.0.1          00:00:01:01:01:01   evpn      active    02/10/2021 16:12:54
-------------------------------------------------------------------------------
Number of entries : 1
===============================================================================
```

SR OS does not include a host IP address in any EVPN MAC advertisement for a MAC learned on a SAP or SDP-binding. Host IP addresses are only included in the EVPN MAC advertisements corresponding to R-VPLS IP interfaces. When deployed as DC GW in a Nuage architecture, the Nuage Networks Virtual Services Controller (VSC) or Virtual Services Gateway (VSG) will send virtual machine and host MAC/IP pairs in EVPN MAC routes. See the Nokia Nuage documentation for more information about the Nuage DC architecture. The SR OS DC GW will populate the proxy-ARP tables with those MAC/IP pairs.

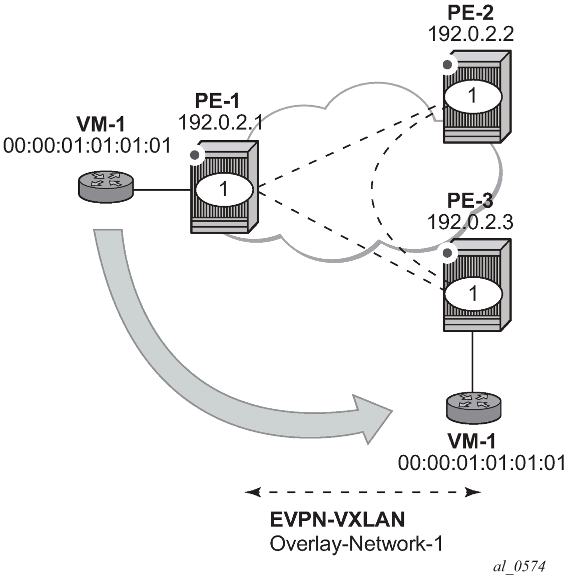
In the preceding CLI excerpt, assume that PE-1 is replaced by a Nuage VSC that sends the pair <172.16.0.1, 00:00:01:01:01:01> in an EVPN MAC route. PE-3 receives the advertisement and adds the entry to its proxy-ARP table for VPLS 1.

The proxy-ARP feature was significantly improved in SR OS Release 13.0; see chapter EVPN for MPLS Tunnels.

MAC mobility, MAC duplication, and MAC protection in EVPN

MAC mobility, duplication and protection are fully supported as specified in RFC 7432. Figure 81: EVPN MAC mobility illustrates the concept of mobility (Virtual Machine VM-1 moves from PE-1 to PE-3).

Figure 81: EVPN MAC mobility



MAC mobility is handled in EVPN by the use of sequence numbers in the MAC routes. When 00:00:01:01:01:01 moves from PE-1 to PE-3, SR OS will gracefully handle it in this way:

- 00:00:01:01:01:01 moves to PE-3 SAP 1/2/1:1
- PE-3 advertises 00:00:01:01:01:01 using a higher sequence number (the first time a MAC is advertised, EVPN uses sequence number 0).
- PE-2 at this point has two valid MAC routes for 00:00:01:01:01:01. It picks up the one coming from PE-3 because the sequence number is higher.
- PE-1 receives the MAC route, and because the sequence number is higher than the one for its own route, it updates the FDB and withdraws its own MAC route.

However, if MAC 00:00:01:01:01:01 is constantly learned on the PE-1 and PE-3 SAPs, the preceding process causes an endless exchange of MAC route advertisements and withdraws that has a negative impact on all the PEs in the EVPN network. This issue is known as “MAC duplication” and is originated by a loop at the access or a duplicated MAC address in two hosts of the same service. SR OS solves this issue through the use of the MAC duplication detection feature. MAC duplication is always enabled with the following default settings:

```
*A:PE-1>config>service>vpls>bgp-evpn# info detail | match mac-duplication context all
----------------------------------------------
                mac-duplication
                    detect num-moves 5 window 3
                    retry 9
                    no black-hole-dup-mac
```

Where:

- **num-moves** — Identifies the number of MAC moves in a VPLS service. The counter is incremented when a MAC is locally relearned in the FDB or flushed from the FDB due to the reception of a better remote EVPN route for that MAC. When the threshold is reached for a MAC address, this MAC address is put in hold-down state (this hold-down state is described below). Range: <3..10>. Default value: 5.

- **window** — Identifies the timer within which a MAC is considered duplicate if it reaches the configured num-moves. Range: <1..15> minutes. Default value: 3 minutes.
- **Retry** — The timer after which the MAC in hold-down state is automatically flushed and the mac-duplication process starts again. This value is expected to be equal to two times or more than the window. If no retry is configured, this implies that, once MAC duplication is detected, MAC updates for that MAC will be held down until the user intervenes or a network event (that flushes the MAC) occurs. Range: <2..60> minutes. Default value: 9 minutes.
- **black-hole-dup-mac** — If enabled and a duplicate MAC address is detected, the router adds the MAC address to the duplicate MAC list and it programs the MAC in the FDB as a protected MAC associated with a black-hole (with type EvpnD:P and source ID "black-hole")

When a MAC address is considered a duplicate or in the hold-down state, no further BGP advertisements are issued for this MAC and an alarm is triggered (by the first MAC address in hold-down state). The following CLI output shows how PE-3 detects that MAC 00:00:01:01:01:01 is a duplicate (after reaching the **num-moves** in **window**) and the corresponding alarm.

```
# on PE-3:
144 2021/02/10 16:16:44.974 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 96
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-MAC Len: 33 RD: 192.0.2.3:1 ESI: ESI-0, tag: 0, mac len: 48
                       mac: 00:00:01:01:01:01, IP len: 0, IP: NULL, label1: 1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:12
        bgp-tunnel-encap:VXLAN
        mac-mobility:Seq:5
"
```

Log 99 on PE-3 shows the following message when EVPN has detected a duplicate MAC address in VPLS 1:

```
# on PE-3:
154 2021/02/10 16:18:58.902 UTC MINOR: SVCMGR #2331 Base
"VPLS Service 1 has MAC(s) detected as duplicates by EVPN mac-duplication detection."
```

The **show service id bgp-evpn** command shows the MAC duplication settings and the list of duplicate MAC addresses on hold-down.

```
*A:PE-3# show service id 1 bgp-evpn

===============================================================================
BGP EVPN Table
===============================================================================
MAC Advertisement  : Enabled           Unknown MAC Route  : Enabled
CFM MAC Advertise  : Disabled
Creation Origin    : manual
```

```
MAC Dup Detn Moves : 3                  MAC Dup Detn Window: 1
MAC Dup Detn Retry : 2                  Number of Dup MACs : 1
MAC Dup Detn BH    : Disabled
IP Route Advert    : Disabled
Sel Mcast Advert   : Disabled

EVI                : n/a
Ing Rep Inc McastAd: Enabled
Accept IVPLS Flush : Disabled

-------------------------------------------------------------------------------
Detected Duplicate MAC Addresses            Time Detected
-------------------------------------------------------------------------------
00:00:01:01:01:01                           02/10/2021 16:18:58
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

SR OS stops sending and processing any BGP MAC advertisement routes for that MAC address until:

- The MAC is flushed due to a local event (SAP/SDP-binding associated to the MAC fails) or the reception of a remote withdraw for the MAC (due to a MAC flush at the remote SR OS) or
- The **retry <in_minutes>** timer expires, which flushes the MAC and restart the process.

When the last duplicate MAC address is removed from the duplicate list, log 99 on PE-3 will show the following message:

```
155 2021/02/10 16:21:58.885 UTC MINOR: SVCMGR #2332 Base
"VPLS Service 1 no longer has MAC(s) detected as duplicates by EVPN mac-duplication
detection."
```

EVPN also provides a mechanism to protect certain MAC addresses that do not move for which connectivity must be guaranteed. These addresses must be protected in case there is an attempt to dynamically learn them in a different place in the EVPN-VXLAN VPLS service (on the same or different PE).

The protected MAC addresses are configured in SR OS as conditional static MAC addresses. A conditional static MAC address defined in an EVPN-VXLAN VPLS service is advertised by BGP-EVPN as a static address. An example of the configuration of a conditional static MAC addressis as follows:

```
# on PE-1:
configure
    service
        vpls "VPLS1"
            static-mac
                mac 00:00:05:05:05:05 create sap 1/2/1:1 monitor fwd-status
            exit
        exit
```

The protected MAC addresses advertised in EVPN are shown in the receiving BGP RIB as Static (MAC mobility extended community with Sequence 0 and sticky bit set) and *EvpnS:P* (Evpn Static: Protected) in the FDB. The advertising PE shows the protected MAC as *CStatic:P* (Conditional Static: Protected) in the FDB:

On the advertising PE:

```
*A:PE-1# show service id 1 fdb mac 00:00:05:05:05:05

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:05:05:05:05 sap:1/2/1:1             CStatic: 02/10/21 16:31:03
                                                     P
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

On the receiving PE:

```
*A:PE-3# show service id 1 fdb mac 00:00:05:05:05:05

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:05:05:05:05 vxlan-1:                EvpnS:P  02/10/21 16:31:03
                             192.0.2.1:1
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

```
*A:PE-3# show router bgp routes evpn mac mac-address 00:00:05:05:05:05 hunt
===============================================================================
 BGP Router ID:192.0.2.3       AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : n/a
Nexthop        : 192.0.2.1
From           : 192.0.2.1
Res. Nexthop   : 192.168.13.1
Local Pref.    : 100                    Interface Name : int-PE-3-PE-1
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64500:12 bgp-tunnel-encap:VXLAN
                 mac-mobility:Seq:0/Static
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.1
Flags          : Used Valid Best IGP
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : MAC
ESI            : ESI-0
Tag            : 0
IP Address     : n/a
Route Dist.    : 192.0.2.1:1
```

```
Mac Address    : 00:00:05:05:05:05
MPLS Label1    : VNI 1                  MPLS Label2    : n/a
Route Tag      : 0
Neighbor-AS    : n/a
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h02m32s

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

The following procedures are supported in order to protect the configured static MAC addresses:

- All the SAP/SDP-bindings are internally configured as MAC protect restrict-protected-src as soon as BGP-EVPN is enabled in the VPLS service.
- Local static MAC addresses or remote EVPN static MAC addresses are considered as protected.
- If a frame with a source MAC address matching one of the protected MAC addresses is received on a different SAP/SDP-binding than the owner of the protected MAC address, the frame is discarded and an alarm triggered. This MAC protection is not performed for frames received on VXLAN bindings.
- The same throttled alarm mechanism used in MAC protect for restrict-protected-src with discard-frame is used here: the offending frames are captured to a list to be polled by the CPM every ~10min.

In this example, PE-3 has 00:00:05:05:05:05 in its FDB as EvpnS. If SAP 1/2/1:1 on PE-3 receives a frame with source MAC address 00:00:05:05:05:05, the frame is discarded and an alarm triggered. The following is logged in log 99 on PE-3:

```
164 2021/02/10 16:44:03.736 UTC MINOR: SVCMGR #2208 Base Slot 1
"Protected MAC 00:00:05:05:05:05 received on SAP 1/2/1:1 in service 1. "
```

Debug and show commands

In addition to the previously mentioned **show service id vxlan destinations**, **show service id bgp-evpn** and **show service id fdb detail** commands, the following commands provide valuable information when troubleshooting an EVPN-VXLAN VPLS service.

The **show router bgp routes evpn** command supports filtering by route type as well as many other route fields.

```
# on any PE:
show router bgp routes evpn ?
  - evpn <evpn-type>

      auto-disc       - Display BGP EVPN Auto-Disc Routes
      eth-seg         - Display BGP EVPN Eth-Seg Routes
      incl-mcast      - Display BGP EVPN Inclusive-Mcast Routes
      ip-prefix       - Display BGP EVPN IPv4-Prefix Routes
      ipv6-prefix     - Display BGP EVPN IPv6-Prefix Routes
      mac             - Display BGP EVPN Mac Routes
      mcast-join-syn* - Display BGP EVPN Mcast Join Sync Routes
      mcast-leave-sy* - Display BGP EVPN Mcast Leave Sync Routes
      smet            - Display BGP EVPN Smet Routes
      spmsi-ad        - Display BGP EVPN Spmsi AD Routes
```

```
# on any PE:
show router bgp routes evpn mac ?
  - mac [hunt|detail] [rd <rd>] [next-hop <next-hop>] [mac-address <mac-address>]
        [community <comm-id>] [tag <tag>]
    [aspath-regex <reg-exp>]

 <hunt|detail>        : keywords
 <rd>                 : {<ip-addr:comm-val>|
                        <2byte-asnumber:ext-comm-val>|
                        <4byte-asnumber:comm-val>}
 <next-hop>           : ipv4-address   - a.b.c.d
                        ipv6-address   - x:x:x:x:x:x:x:x   (eight 16-bit pieces)
                                         x:x:x:x:x:x:d.d.d.d
                                         x - [0..FFFF]H
                                         d - [0..255]D
 <mac-address>        : xx:xx:xx:xx:xx:xx or xx-xx-xx-xx-xx-xx
 <comm-id>            : <as-number1:comm-val1>|<ext-comm>|
                        <well-known-comm>
                        ext-comm        - <type>:{<ip-address:comm-val1>|
                                                  <as-number1:comm-val2>|
                                                  <as-number2:comm-val1>}
                        as-number1      - [0..65535]
                        comm-val1       - [0..65535]
                        type            - target|origin
                        ip-address      - a.b.c.d
                        comm-val2       - [0..4294967295]
                        as-number2      - [0..4294967295]
                        well-known-comm - null|no-export|no-export-subconfed|
                                          no-advertise
 <tag>                : [0..4294967295] | MAX-ET
 <reg-exp>            : [80 chars max]
```

```
*A:PE-3# show router bgp routes evpn mac tag 0
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.1:1         00:00:05:05:05:05 ESI-0
      0                   Static            VNI 1
                          n/a
                          192.0.2.1

u*>i  192.0.2.1:1         02:0f:ff:00:03:3a ESI-0
      0                   Static            VNI 1
                          n/a
                          192.0.2.1

u*>i  192.0.2.2:1         00:00:00:00:00:00 ESI-0
```

```
      0                   Seq:0            VNI 1
                          n/a
                          192.0.2.2

-------------------------------------------------------------------------------
Routes : 3
===============================================================================
```

The **tools dump service id vxlan** commad displays the number of times a service could not add a VXLAN binding or <VTEP, Egress VNI> due to the following limits:

- The per system VTEP limit has been reached
- The per system (egress VTEP, egress VNI) limit has been reached
- The per service (egress VTEP, egress VNI) limit has been reached
- The per system Bind limit: Total bind limit or VXLAN bind limit has been reached.

```
*A:PE-1# tools dump service id 1 vxlan

VTEP, Egress VNI Failure statistics at 02/10/2021 17:03:07:

statistics last cleared at 02/10/2021 10:43:55:

Failures: None
```

```
*A:PE-1# tools dump service id 1 evpn usage

Evpn Tunnel Interface IP Next Hop: N/A
```

The **tools dump service evpn usage** command displays the consumed resources in the system:

```
*A:PE-1# tools dump service evpn usage

vxlan-evpn-mpls usage statistics at 02/10/2021 17:03:07:

MPLS-TEP                                        :             0
VXLAN-TEP                                       :             2
Total-TEP                                       :      2/ 16383

Mpls Dests (TEP, Egress Label + ES + ES-BMAC)   :             0
Mpls Etree Leaf Dests                           :             0
Vxlan Dests (TEP, Egress VNI + ES)              :             2
Total-Dest                                      :      2/196607

Sdp Bind +  Evpn Dests                          :      2/245759
ES L2/L3 PBR                                    :      0/ 32767
Evpn Etree Remote BUM Leaf Labels               :             0
```

Conclusion

SR OS supports the EVPN control plane for VXLAN tunnels terminated in VPLS services. VXLAN is an overlay IP tunneling mechanism that is being used in data centers, data center interconnect, and other applications. EVPN is a scalable and flexible control plane that provides control over the MAC addresses being learned and advertised, as well as other mechanisms to optimize Layer 2 services such as proxy-ARP, MAC mobility, MAC duplication detection, and MAC protection. SR OS provides a resilient and scalable EVPN-VXLAN solution for Layer 2 services, including interoperability to existing VPLS networks. This chapter showed all of those functions and how they are configured and operated.

EVPN for VXLAN Tunnels (Layer 3)

This chapter provides information about EVPN for VXLAN tunnels (Layer 3).

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter is applicable to SR OS and was initially written for Release 12.0.R4. The CLI in the current edition is based on SR OS Release 21.10.R3. Ethernet Virtual Private Network (EVPN) is a control plane technology and does not have line card hardware dependencies.

Chapter EVPN for VXLAN Tunnels (Layer 2) is prerequisite reading.

Overview

As discussed in the EVPN for VXLAN Tunnels (Layer 2) chapter, EVPN and VXLAN can be enabled on VPLS or R-VPLS services in SR OS. Where that chapter focuses on the use of EVPN-VXLAN layer 2 services, in other words, how EVPN-VXLAN is configured in VPLS services, this chapter describes how EVPN-VXLAN can be used to provide inter-subnet forwarding in R-VPLS and VPRN services. Inter-subnet forwarding can be provided by regular R-VPLS and VPRN services. However, EVPN provides an efficient and unified way to populate Forwarding Databases (FDBs), Address Resolution Protocol (ARP) tables and routing tables using a single BGP address family. Inter-subnet forwarding in overlay networks would otherwise require data plane learning and the use of routing protocols on a per VPRN basis.

The SR OS solution for inter-subnet forwarding using EVPN is based on building blocks described in *draft-ietf-bess-evpn-inter-subnet-forwarding* and the use of the EVPN IP-prefix routes (route type 5) as explained in RFC 9136. This example describes three supported common scenarios and provides the CLI configuration and required tools to troubleshoot EVPN-VXLAN in each case. The scenarios configured and explained are:

- EVPN-VXLAN in R-VPLS services
- EVPN-VXLAN in Integrated Routing Bridging (IRB) backhaul R-VPLS services
- EVPN-VXLAN in EVPN tunnel R-VPLS services

In all these scenarios, redundant PEs are usually deployed. If that is the case, the interaction of EVPN, IP-VPN, and the Routing Table Manager (RTM) may lead to some routing loop situations that must be avoided by using routing policies (this also may happen in traditional IP-VPN deployments when eBGP and MP-BGP interact to populate VPRN routing tables in multi-homed networks). This chapter explains when those routing loops can happen and how to avoid them.

The term IRB interface refers to an R-VPLS service bound to a VPRN IP interface. The terms IRB interface and R-VPLS interface are used interchangeably throughout this chapter.

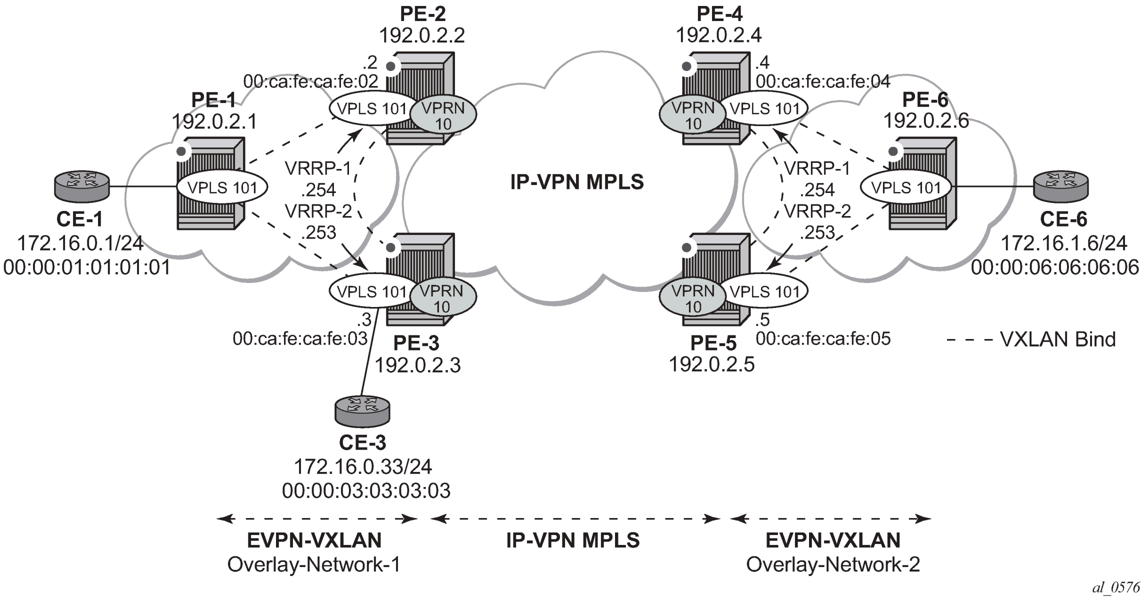
Configuration

This section describes the configuration of EVPN-VXLAN for Layer 3 services on SR OS, as well as the available troubleshooting and show commands. The three scenarios described in the overview are analyzed independently.

EVPN-VXLAN in an R-VPLS service

Figure 82: EVPN-VXLAN for R-VPLS services shows the topology used in the first scenario.

Figure 82: EVPN-VXLAN for R-VPLS services



The network topology shows two overlay (VXLAN) networks interconnected by an MPLS network:

- PE-1, PE-2, and PE-3 are part of Overlay-Network-1
- PE-4, PE-5, and PE-6 are part of Overlay-Network-2

A Layer 2/Layer 3 service is provided to a customer to connect CE-1, CE-3, and CE-6. In this scenario, Layer 2 connectivity is provided within each overlay network and inter-subnet connectivity (Layer 3) is provided between the overlay networks. VPLS 101 is defined within each overlay network and VPRN 10 connects both Layer 2 services through an IP-VPN MPLS network.

This topology can illustrate a Data Center Interconnect (DCI) example, where Overlay-Network-1 and Overlay-Network-2 are two data centers interconnected through an MPLS WAN. In this application, CE-1, CE-3, and CE-6 simulate virtual machines or appliances, PE-2/3/4/5 act as Data Center Gateways (DC GWs) and PE-1/6 as Network Virtualization Edge devices (or virtual PEs running on a compute infrastructure).

The following protocols and objects are configured beforehand:

- The ports interconnecting the six PEs in Figure 82: EVPN-VXLAN for R-VPLS services are configured as network or hybrid ports and have router network interfaces defined in them. Only the ports connected to the CEs are configured as access ports.
- The six PEs are running IS-IS for the global routing table with the four core PEs interconnected using IS-IS Level-2 point-to-point interfaces and each overlay network using IS-IS Level-1 point-to-point interfaces.
- LDP is used as the MPLS protocol to signal transport tunnel labels among PE-2, PE-3, PE-4, and PE-5. There is no LDP running within each overlay network.
- The network port MTU (in all the ports sending/receiving VXLAN packets) must be at least 50 bytes (54 if dot1q encapsulation is used) greater than the service MTU to accommodate the size of the VXLAN header.

Once the IGP infrastructure and LDP in the core are enabled, BGP is configured. In this scenario, two BGP families must be enabled: EVPN within each overlay network for the exchange of MAC/IP addresses and setting up the flooding domains, and VPN-IPv4 among the four core PEs so that IP prefixes can be exchanged and resolved to MPLS tunnels in the core.

The following CLI output shows the BGP configuration of PE-1, which only needs the EVPN family. PE-6 has a similar BGP configuration, that is, only EVPN family is configured for its peers. The use of Route Reflectors (RRs) in these scenarios is common. Although this scenario does not use RRs, an EVPN RR could have been used in Overlay-Network-1 and Overlay-Network-2 and a separate VPN-IPv4 RR could have been used in the core IP-VPN MPLS network.

```
# on PE-1:
configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            rapid-update evpn
            group "DC"
                family evpn
                peer-as 64500
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
            no shutdown
        exit
```

The BGP configuration on the DC GWs is as follows:

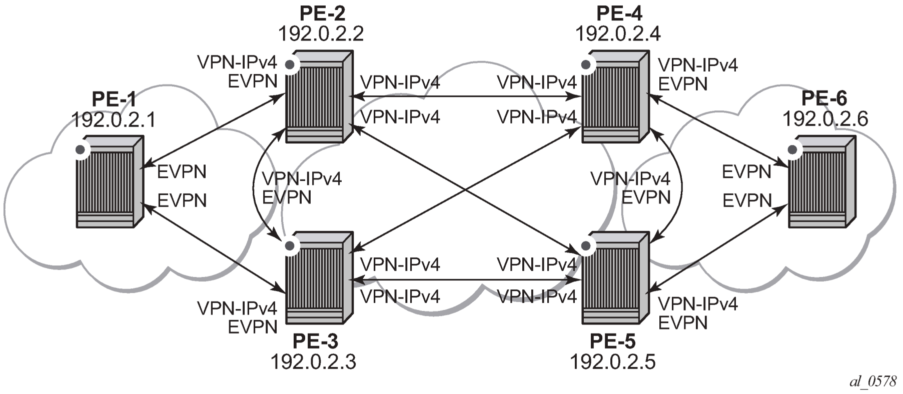
```
# on PE-2:
configure
    router
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            rapid-update evpn
            group "DC"
                family vpn-ipv4 evpn
                peer-as 64500
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.3
                exit
            exit
            group "WAN"
                family vpn-ipv4
                peer-as 64500
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
            no shutdown
        exit
```

```
# on PE-3:
configure
    router
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            rapid-update evpn
            group "DC"
                family vpn-ipv4 evpn
                peer-as 64500
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.2
                exit
            exit
            group "WAN"
                family vpn-ipv4
                peer-as 64500
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
            no shutdown
        exit
```

The DC GWs PE-4 and PE-5 have an equivalent BGP configuration.

Figure 83: BGP adjacencies and enabled families shows the BGP peering sessions among the PEs and the enabled BGP families. PE-1 and PE-6 only establish an EVPN peering session with their peers (only the EVPN family is enabled on PE-1 and PE-6, even if the peer PEs are VPN-IPv4 capable as well).

Figure 83: BGP adjacencies and enabled families



Once the network infrastructure is running properly, the actual service configuration, as illustrated in Figure 82: EVPN-VXLAN for R-VPLS services, can be carried out. The following CLI shows the configuration for VPLS 101 and VPRN 10 in PE-1, PE-2, and PE-3. The other overlay network has a similar configuration.

```
# on PE-1:
configure
    service
        vpls 101 name "evi-101" customer 1 create
            vxlan instance 1 vni 101 create
            exit
            bgp
                route-distinguisher 192.0.2.1:101
                route-target export target:64500:101 import target:64500:101
            exit
            bgp-evpn
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            sap 1/2/1:101 create
                no shutdown
            exit
            proxy-arp
                no shutdown
            exit
            no shutdown
```

Proxy-ARP is disabled (default) on PE-2, as well as on the other core PEs:

```
# on PE-2:
configure
    service
        vpls 101 name "evi-101" customer 1 create
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 101 create
            exit
            bgp
                route-distinguisher 192.0.2.2:101
                route-target export target:64500:101 import target:64500:101
            exit
            bgp-evpn
                vxlan bgp 1 vxlan-instance 1
```

```
                    no shutdown
                exit
            exit
            no shutdown
        exit
        vprn 10 name "VPRN10" customer 1 create
            ecmp 2
            interface "int-1" create
                address 172.16.0.2/24
                mac 00:ca:fe:ca:fe:02
                vrrp 1
                    backup 172.16.0.254
                    priority 254
                    ping-reply
                    traceroute-reply
                    mac 00:ca:fe:ca:fe:54
                exit
                vrrp 2
                    backup 172.16.0.253
                    ping-reply
                    traceroute-reply
                    mac 00:ca:fe:ca:fe:53
                exit
                vpls "evi-101"
                exit
            exit
            bgp-ipvpn
                mpls
                    auto-bind-tunnel
                        resolution-filter
                            ldp
                        exit
                        resolution filter
                    exit
                    route-distinguisher 192.0.2.2:10
                    vrf-target target:64500:10
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

```
# on PE-3:
configure
    service
        vpls 101 name "evi-101" customer 1 create
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 101 create
            exit
            bgp
                route-distinguisher 192.0.2.3:101
                route-target export target:64500:101 import target:64500:101
            exit
            bgp-evpn
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            sap 1/2/1:101 create
                no shutdown
            exit
            no shutdown
        exit
        vprn 10 name "VPRN10" customer 1 create
            ecmp 2
            interface "int-1" create
                address 172.16.0.3/24
                mac 00:ca:fe:ca:fe:03
                vrrp 1
                    backup 172.16.0.254
                    ping-reply
                    traceroute-reply
                    mac 00:ca:fe:ca:fe:54
                exit
                vrrp 2
                    backup 172.16.0.253
                    priority 254
                    ping-reply
                    traceroute-reply
                    mac 00:ca:fe:ca:fe:53
                exit
                vpls "evi-101"
                exit
            exit
            bgp-ipvpn
                mpls
                    auto-bind-tunnel
                        resolution-filter
                            ldp
                        exit
                        resolution filter
                    exit
                    route-distinguisher 192.0.2.3:10
                    vrf-target target:64500:10
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

For details about the EVPN and VXLAN configuration on PE-1 VPLS 101, see chapter EVPN for VXLAN Tunnels (Layer 2). The configuration of VPLS 101 on PE-2 and PE-3 has the following important aspects:

- The **allow-ip-int-bind** command is required so that the R-VPLS can be bound to VPRN 10.
- The service name "evi-101" is configured when the service is created and cannot be modified afterward. The service name must match the name configured in the VPRN 10 VPLS interface.
- Even though EVPN and VXLAN are properly configured, proxy-ARP cannot be enabled in VPLS 101. In an R-VPLS with EVPN-VXLAN, proxy-ARP is not supported and the VPRN ARP table is used instead. When an EVPN MAC route that includes an IP address is received in an R-VPLS, the MAC-IP pair encoded in the route is added to the ARP table of the VPRN, as opposed to the proxy-ARP table.

```
*A:PE-2>config>service>vpls>proxy-arp$ no shutdown
MINOR: SVCMGR #8007 Cannot modify proxy arp - service is routed
```

When configuring VPRN 10 on PE-2 and PE-3, the following considerations must be taken into account:

- When trying to enable existing VPRN features on interfaces linked to EVPN-VXLAN R-VPLS interfaces, the authentication-policy command is not supported:

```
*A:PE-2>config>service>vprn>if# authentication-policy "authPol1"
INFO: PIP #1875 Cannot configure auth-policy on routed-vpls interface
```

- Dynamic routing protocols such as IS-IS, RIP, or OSPF are not supported.
- In general, no SR OS control plane generated packets are sent to the egress VXLAN bindings except for ARP, VRRP, ICMP, BFD, and Eth-CFM.
- As shown in Figure 82: EVPN-VXLAN for R-VPLS services and in the CLI excerpts, VRRP can be configured on the VPRN 10 VPLS interfaces to provide default gateway redundancy to the hosts connected to VPLS 101. Two VRRP instances are configured so that VPLS 101 upstream traffic can be load-balanced to PE-2 and PE-3. With VRRP on EVPN-VXLAN R-VPLS interfaces:
  - **Ping-reply** and **traceroute-reply** can be configured and are supported. BFD is also supported to speed up the fault detection.
  - **standby-forwarding**, even if it were configured for VRRP, would not have any effect in this configuration: the standby PE will never see any flooded traffic sent to it, so this command is not applicable to this scenario.
- When a VPRN 10 VPLS interface is bound to VPLS 101, EVPN advertises all the IP addresses configured for that VPLS interface as MAC routes with a static MAC indication. For the remote EVPN peers, that means that those MAC addresses linked to remote IP interfaces are protected. VRRP virtual IP/MACs are also advertised by EVPN as "static" and so protected. In the example of Figure 82: EVPN-VXLAN for R-VPLS services, the VPLS 101 FDB in PE-1 shows the IP interface MAC addresses and VRRP MAC addresses as *EvpnS:P* (Static and protected MAC) as shown in the following output:

```
*A:PE-1# show service id 101 fdb detail

===============================================================================
Forwarding Database, Service 101
===============================================================================
ServId     MAC                Source-Identifier       Type     Last Change
            Transport:Tnl-Id                          Age
-------------------------------------------------------------------------------
101        00:00:01:01:01:01 sap:1/2/1:101           L/0      03/02/22 11:34:55
101        00:00:03:03:03:03 vxlan-1:                Evpn     03/02/22 11:35:37
                             192.0.2.3:101
101        00:ca:fe:ca:fe:02 vxlan-1:                EvpnS:P  03/02/22 11:35:05
                             192.0.2.2:101
101        00:ca:fe:ca:fe:03 vxlan-1:                EvpnS:P  03/02/22 11:35:37
                             192.0.2.3:101
101        00:ca:fe:ca:fe:53 vxlan-1:                EvpnS:P  03/02/22 11:35:40
                             192.0.2.3:101
101        00:ca:fe:ca:fe:54 vxlan-1:                EvpnS:P  03/02/22 11:35:08
                             192.0.2.2:101
-------------------------------------------------------------------------------
No. of MAC Entries: 6
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The VPRN 10 VRRP instances on PE-2 are the following:

```
*A:PE-2# show router 10 vrrp instance

===============================================================================
VRRP Instances
===============================================================================
Interface Name                   VR Id Own Adm  State       Base Pri   Msg Int
                                 IP        Opr  Pol Id      InUse Pri  Inh Int
-------------------------------------------------------------------------------
int-1                            1     No  Up   Master       254       1
```

```
                                 IPv4      Up   n/a         254        No
  Backup Addr: 172.16.0.254
int-1                            2     No  Up   Backup       100       1
                                 IPv4      Up   n/a         100        No
  Backup Addr: 172.16.0.253
-------------------------------------------------------------------------------
Instances : 2
===============================================================================
```

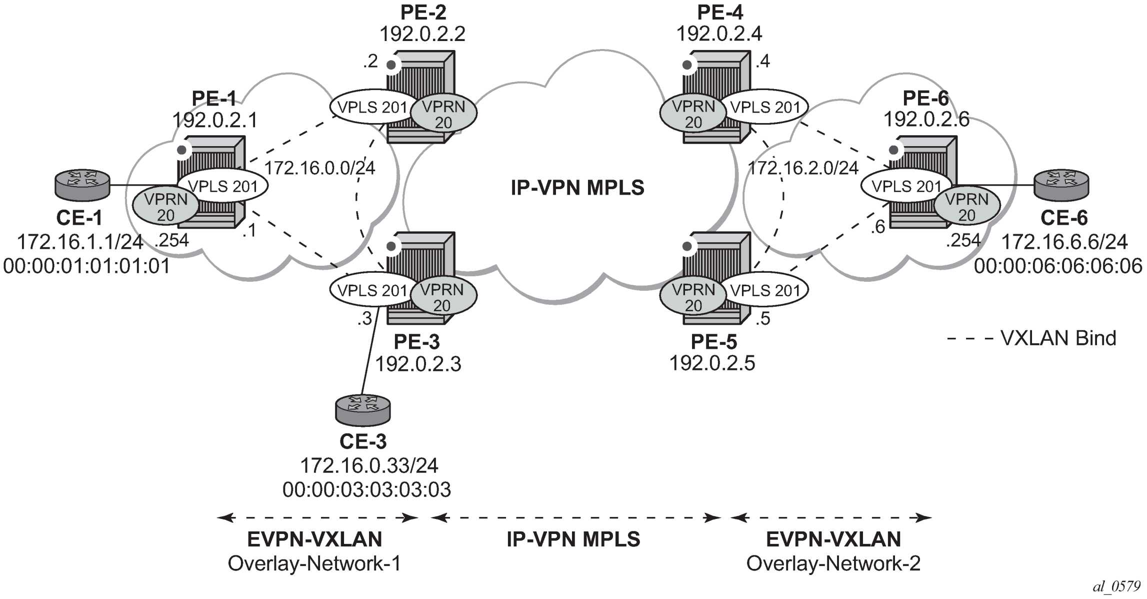
The ARP entries for PE-2 are the following:

```
*A:PE-2# show router 10 arp

===============================================================================
ARP Table (Service: 10)
===============================================================================
IP Address      MAC Address       Expiry    Type   Interface
-------------------------------------------------------------------------------
172.16.0.2      00:ca:fe:ca:fe:02 00h00m00s Oth[I] int-1
172.16.0.3      00:ca:fe:ca:fe:03 00h00m00s Evp[I] int-1
172.16.0.253    00:ca:fe:ca:fe:53 00h00m00s Oth    int-1
172.16.0.254    00:ca:fe:ca:fe:54 00h00m00s Oth[I] int-1
-------------------------------------------------------------------------------
No. of ARP Entries: 4
===============================================================================
```

EVPN-VXLAN in IRB backhaul R-VPLS services

Figure 84: EVPN-VXLAN for IRB backhaul R-VPLS services illustrates the second inter-subnet forwarding scenario, where Layer 3 connectivity must be provided not only between the overlay networks but also within each overlay network. In the example shown in Figure 84: EVPN-VXLAN for IRB backhaul R-VPLS services, a customer (tenant) has different subnets and connectivity must be provided across all of them (CE-1, CE-3, and CE-6 must be able to communicate), bearing in mind that EVPN-VXLAN is enabled in each overlay network and IP-VPN MPLS is used to interconnect both overlay networks. VPLS 201 is an IRB Backhaul R-VPLS service because it provides connectivity to the VPRN instances.

Figure 84: EVPN-VXLAN for IRB backhaul R-VPLS services



From a BGP peering perspective, there is no change in this scenario compared to the previous one: PE-1 and PE-6 only support the EVPN address family. However, in this scenario, CE-1 is now connected to an R-VPLS directly linked to the VPRN instances in PE-2/PE-3. As a result of that, IP prefixes must be exchanged between PE-1 and PE-2/PE-3. EVPN can advertise not only MAC routes and Inclusive Multicast routes, but also IP prefix routes that contain IP prefixes that can be installed in the attached VPRN routing table.

As an example, the VPRN 20 and VPLS "evi-201" configurations on PE-1, PE-2, and PE-3 are shown. Similar configurations are needed in PE-4, PE-5, and PE-6.

On PE-1, VPRN 20 and VPLS "evi-201" are configured as follows:

```
# on PE-1:
configure
    service
        vprn 20 name "VPRN20" customer 1 create
            interface "int-evi-201" create
                address 172.16.0.1/24
                vpls "evi-201"
                exit
            exit
            interface "int-PE-1-CE-1" create
                address 172.16.1.254/24
                sap 1/2/1:20 create
                exit
            exit
            no shutdown
        exit
        vpls 201 name "evi-201" customer 1 create
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 201 create
            exit
            bgp
```

```
                route-distinguisher 192.0.2.1:201
                route-target export target:64500:201 import target:64500:201
            exit
            bgp-evpn
                ip-route-advertisement
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

On PE-2, VPRN 20 and VPLS "evi-201" are configured as follows:

```
# on PE-2:
configure
    service
        vprn 20 name "VPRN20" customer 1 create
            interface "int-evi-201" create
                address 172.16.0.2/24
                vpls "evi-201"
                exit
            exit
            bgp-ipvpn
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    route-distinguisher 192.0.2.2:20
                    vrf-target target:64500:20
                    no shutdown
                exit
            exit
            no shutdown
        exit
        vpls 201 name "evi-201" customer 1 create
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 201 create
            exit
            bgp
                route-distinguisher 192.0.2.2:201
                route-target export target:64500:201 import target:64500:201
            exit
            bgp-evpn
                ip-route-advertisement
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
```

On PE-3, VPRN 20 and VPLS "evi-201" are configured as follows:

```
# on PE-3:
configure
    service
        vprn 20 name "VPRN20" customer 1 create
            interface "int-evi-201" create
```

```
                address 172.16.0.3/24
                vpls "evi-201"
                exit
            exit
            bgp-ipvpn
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    route-distinguisher 192.0.2.3:20
                    vrf-target target:64500:20
                    no shutdown
                exit
            exit
            no shutdown
        exit
        vpls 201 name "evi-201" customer 1 create
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 201 create
            exit
            bgp
                route-distinguisher 192.0.2.3:201
                route-target export target:64500:201 import target:64500:201
            exit
            bgp-evpn
                ip-route-advertisement
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            sap 1/2/1:20 create
                no shutdown
            exit
            no shutdown
        exit
```

As shown in the CLI excerpt, the configuration in the three nodes (PE-1, PE-2, and PE-3) for VPLS "evi-201" and VPRN 20 is very similar. The main difference is the **auto-bind-tunnel** command in VRPN 20 on PE-2 and PE-3. This command allows the VPRN 20 on PE-2 and PE-3 to receive IP-VPN routes from the core and resolve them to MPLS tunnels. VPRN 20 on PE-1 does not require such command because all its IP prefixes are resolved to local interfaces or to EVPN peers.

The **ip-route-advertisement** command enables:

- The advertisement of IP prefixes in EVPN, in routes type 5. All the existing IP prefixes in the attached VPRN 20 routing table are advertised in EVPN within the VPLS 201 context (except for the ones associated to VPLS 201 itself).
- The installation of IP prefixes in the attached VPRN 20 routing table with a preference of 169 (BGP-VPN routes for IP-VPN have a preference of 170) and a next-hop of the gateway IP (GW IP) address included in the EVPN IP prefix route.

For instance, the following output shows that PE-1 advertises the IP prefix 172.16.1.0/24 as an EVPN route to PE-3 (a similar route is sent to PE-2), captured by a **debug router bgp update** session.

```
44 2022/03/02 11:38:45.956 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 82
    Flag: 0x90 Type: 14 Len: 45 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.1
        Type: EVPN-IP-PREFIX Len: 34 RD: 192.0.2.1:201, tag: 0,
              ip_prefix: 172.16.1.0/24 gw_ip 172.16.0.1 Label: 201 (Raw Label: 0xc9)
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:201
        bgp-tunnel-encap:VXLAN
"
```

The VPRN 20 routing table in PE-1 includes two EVPN Interface-ful (EVPN-IFF) routes with preference 169, as follows:

```
*A:PE-1# show router 20 route-table

===============================================================================
Route Table (Service: 20)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
172.16.0.0/24                                 Local   Local     00h22m22s  0
       int-evi-201                                                  0
172.16.1.0/24                                 Local   Local     00h22m22s  0
       int-PE-1-CE-1                                                0
172.16.2.0/24                                 Remote  EVPN-IFF  00h01m41s  169
       172.16.0.2                                                   0
172.16.6.0/24                                 Remote  EVPN-IFF  00h01m41s  169
       172.16.0.2                                                   0
-------------------------------------------------------------------------------
No. of Routes: 4
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The subnet 172.16.0.0/24 is used on the interfaces "int-evi-201" in overlay network 1 and subnet 172.16.2.0/24 is used on similar interfaces in overlay network 2. CE-1 has an IP address in subnet 172.16.1.0/24 and CE-6 has an IP address in subnet 172.16.6.0/24. The next hop to reach 172.16.2.0/24 (overlay network 2) or CE-6, is 172.16.0.2 (PE-2), but it could have been PE-3.

There is redundancy in the example setup and therefore, loops can occur. To avoid loops, routing policies need to be configured on the core PEs (PE-2, PE-3, PE-4, and PE-5). These policies are described in the Use of routing policies to avoid routing loops in redundant PEs section for routing loop use case 1.

The routing table on PE-2 shows a EVPN-IFF route toward CE-1 (subnet 172.16.1.0/24) via PE-1. The route toward CE-6 uses a tunnel toward PE-4 in overlay network 2.

```
*A:PE-2# show router 20 route-table

===============================================================================
Route Table (Service: 20)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
172.16.0.0/24                                 Local   Local     00h20m36s  0
       int-evi-201                                                  0
```

```
172.16.1.0/24                                 Remote  EVPN-IFF  00h02m17s  169
       172.16.0.1                                                   0
172.16.2.0/24                                 Remote  BGP VPN   00h01m43s  170
       192.0.2.4 (tunneled)                                         10
172.16.6.0/24                                 Remote  BGP VPN   00h01m43s  170
       192.0.2.4 (tunneled)                                         10
-------------------------------------------------------------------------------
No. of Routes: 4
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The routing table on PE-3 is as follows:

```
*A:PE-3# show router 20 route-table

===============================================================================
Route Table (Service: 20)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
172.16.0.0/24                                 Local   Local     00h09m20s  0
       int-evi-201                                                  0
172.16.1.0/24                                 Remote  EVPN-IFF  00h02m46s  169
       172.16.0.1                                                   0
172.16.2.0/24                                 Remote  BGP VPN   00h02m20s  170
       192.0.2.4 (tunneled)                                         10
172.16.6.0/24                                 Remote  BGP VPN   00h01m53s  170
       192.0.2.4 (tunneled)                                         10
-------------------------------------------------------------------------------
No. of Routes: 4
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

When checking the operation of EVPN in this scenario, it is important to observe that the right next hops and prefixes are successfully installed in the VPRN 20 routing table:

- EVPN IP prefixes are sent using a GW IP matching the primary IP interface address of the R-VPLS for which the routes are sent. For instance, as shown above, IP prefix 172.16.1.0/24 is advertised from PE-1 with GW IP 172.16.0.1, which is the IP address configured for the VPRN 20 VPLS interface in PE-1. In the VPRN 20 routing tables on PE-2 and PE-3, IP prefix 172.16.1.0/24 is installed with next hop 172.16.0.1. Traffic arriving at PE-2 or PE-3 on VPRN 20 with IP Destination Address (DA) in the 172.16.1.0/24 subnet matches the mentioned routing table entry. As usual, the next-hop is resolved by the ARP table to a MAC address and the MAC address resolved by the FDB table to an egress VTEP, VNI.
- IP prefixes in the VPRN 20 routing table are advertised in IP-VPN to the remote IP-VPN MPLS peers. Received IP-VPN prefixes are installed in the VPRN 20 routing table using the remote PE system IP address as the next hop, as usual. For instance, 172.16.6.0/24 is installed in the routing table of VPRN 20 on PE-2 with next-hop (tunneled) 192.0.2.4 and preference 170.

The following considerations of how the routing table manager (RTM) handles EVPN and IP-VPN prefixes must be taken into account:

- Only VPRN interface primary addresses are advertised as GW IP in EVPN IP prefix routes. Secondary addresses are never sent as GW IP addresses.
- EVPN IP prefixes are advertised by default as soon as the **ip-route-advertisement** command is enabled and there are active IP prefixes in the attached VPRN routing table.
- If the same IP prefix is received on a PE via EVPN and IP-VPN at the same time for the same VPRN, by default, the EVPN prefix is selected because its preference (169) is better than the IP-VPN preference (170).
- Because EVPN has a better preference compared to IP-VPN, when the VPRNs on redundant PEs are attached to the same R-VPLS service, routing loops may occur. The use case described here is an example where routing loops can occur. Check Use of routing policies to avoid routing loops in redundant PEs to avoid routing loops in redundant PEs for more information.
- When the command **ip-route-advertisement** is enabled, the subnet IP prefixes are advertised in EVPN but not the host IP prefixes (/32 prefixes associated with the local interfaces). If the user wants to advertise the host IP prefixes as well, the **incl-host** keyword must be added to the **ip-route-advertisement** command. The following example illustrates this.

```
*A:PE-1# show router 20 route-table

===============================================================================
Route Table (Service: 20)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
172.16.0.0/24                                 Local   Local     00h10m12s  0
       int-evi-201                                                 0
172.16.1.0/24                                 Local   Local     00h10m12s  0
       int-PE-1-CE-1                                               0
172.16.2.0/24                                 Remote  EVPN-IFF  00h02m51s  169
       172.16.0.2                                                  0
172.16.6.0/24                                 Remote  EVPN-IFF  00h02m51s  169
       172.16.0.2                                                  0
-------------------------------------------------------------------------------
No. of Routes: 4
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The host routes can be shown with the **show router route-table all** command:

```
*A:PE-1# show router 20 route-table all

===============================================================================
Route Table (Service: 20)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                         Active     Metric
-------------------------------------------------------------------------------
172.16.0.0/24                                 Local   Local     00h10m49s  0
       int-evi-201                                      Y          0
172.16.0.1/32                                 Local   Host      00h10m49s  0
       int-evi-201                                      Y          0
172.16.1.0/24                                 Local   Local     00h10m49s  0
       int-PE-1-CE-1                                    Y          0
```

```
172.16.1.254/32                               Local   Host      00h10m49s  0
       int-PE-1-CE-1                                   Y            0
172.16.2.0/24                                 Remote  EVPN-IFF  00h03m28s  169
       172.16.0.2                                      Y            0
172.16.6.0/24                                 Remote  EVPN-IFF  00h03m28s  169
       172.16.0.2                                      Y            0
-------------------------------------------------------------------------------
No. of Routes: 6
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
       E = Inactive best-external BGP route
===============================================================================
```

When the **incl-host** keyword is added to VPLS "evi-201" on PE-1, PE-1 advertises the host routes as well and these are installed in the routing tables on the remote PEs.

```
# on PE-1:
configure
    service
        vpls "evi-201"
            bgp-evpn
                ip-route-advertisement incl-host
            exit
```

```
*A:PE-2# show router 20 route-table

===============================================================================
Route Table (Service: 20)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
172.16.0.0/24                                 Local   Local     00h11m40s  0
       int-evi-201                                                  0
172.16.1.0/24                                 Remote  EVPN-IFF  00h04m59s  169
       172.16.0.1                                                   0
172.16.1.254/32                               Remote  EVPN-IFF  00h00m11s  169
       172.16.0.1                                                   0
172.16.2.0/24                                 Remote  BGP VPN   00h04m27s  170
       192.0.2.4 (tunneled)                                         10
172.16.6.0/24                                 Remote  BGP VPN   00h04m27s  170
       192.0.2.4 (tunneled)                                         10
-------------------------------------------------------------------------------
No. of Routes: 5
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

- ECMP is fully supported for the VPRN for EVPN IP prefix routes coming from different GW IP next-hops. However, ECMP is not supported for IP prefixes routes belonging to different owners (EVPN and IP-VPN). ECMP is enabled in VPRN 20 on PE-1, as follows:

```
# on PE-1:
configure
    service
        vprn "VPRN20"
```

```
            ecmp 2
```

When policies are applied that prevent routing loops, as described in section Use of routing policies to avoid routing loops in redundant PEs, both PE-2 and PE-3 have IP-VPN tunnels for IP prefixes 172.16.2.0/24 and 172.16.6.0/24. In that case, an additional route with a different GW IP as next hop is installed in the routing table for these IP prefixes:

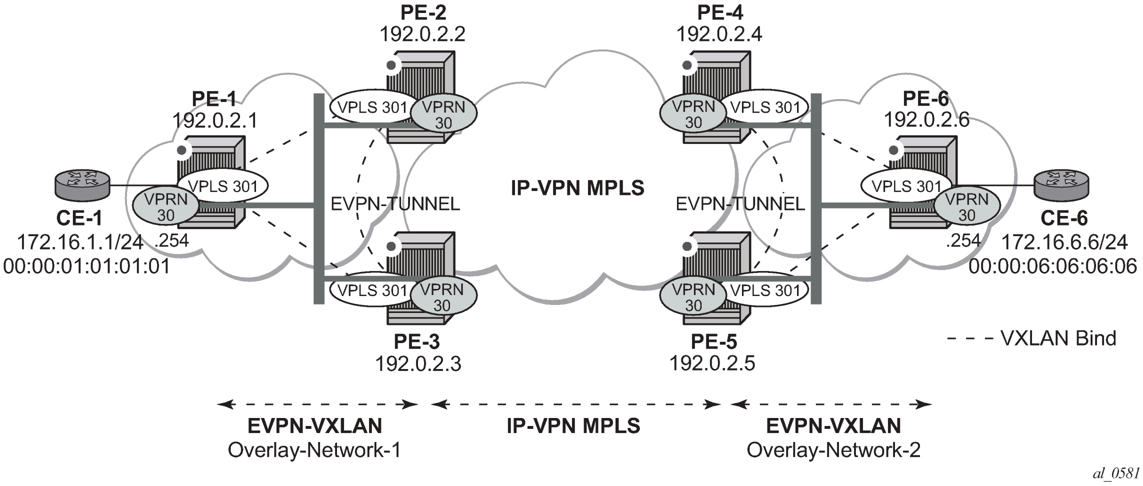
```

*A:PE-1# show router 20 route-table

===============================================================================
Route Table (Service: 20)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
172.16.0.0/24                                 Local   Local     00h12m39s  0
       int-evi-201                                                  0
172.16.1.0/24                                 Local   Local     00h12m39s  0
       int-PE-1-CE-1                                                0
172.16.2.0/24                                 Remote  EVPN-IFF  00h00m08s  169
       172.16.0.2                                                   0
172.16.2.0/24                                 Remote  EVPN-IFF  00h00m08s  169
       172.16.0.3                                                   0
172.16.6.0/24                                 Remote  EVPN-IFF  00h00m08s  169
       172.16.0.2                                                   0
172.16.6.0/24                                 Remote  EVPN-IFF  00h00m08s  169
       172.16.0.3                                                   0
-------------------------------------------------------------------------------
No. of Routes: 6
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

EVPN-VXLAN in EVPN tunnel R-VPLS services

The previous scenario shows how to use EVPN-VXLAN to provide inter-subnet forwarding for a tenant, where R-VPLS services can contain hosts and also offer transit services between VPRN instances. For example, in the use case shown in Figure 84: EVPN-VXLAN for IRB backhaul R-VPLS services, VPLS 201 in Overlay-Network-1 is an R-VPLS that can provide intra-subnet connectivity to all the hosts in subnet 172.16.0.0/24 (for example, CE-3 belongs to this subnet) but it can also provide transit or backhaul connectivity to hosts in subnet 172.16.1.0/24 (for example, CE-1) sending packets to subnets 172.16.2.0/24 or 172.16.6.0/24.

In some cases, the R-VPLS where EVPN-VXLAN is enabled does not need to provide intra-subnet connectivity and it is purely a transit or backhaul service where VPRN IRB interfaces are connected. Figure 85: EVPN-VXLAN in EVPN-tunnel R-VPLS services illustrates this use case.

Figure 85: EVPN-VXLAN in EVPN-tunnel R-VPLS services



Compared to the preceding use case in Figure 84: EVPN-VXLAN for IRB backhaul R-VPLS services, in this case the R-VPLS connecting the IRB interfaces in Overlay-Network-1 (VPLS 301) does not have any connected host. If that is the case, VPLS 301 can be configured as an EVPN tunnel.

EVPN tunnels are enabled using the **evpn-tunnel** command under the R-VPLS interface configured on the VPRN. EVPN tunnels bring the following benefits to EVPN-VXLAN IRB backhaul R-VPLS services:

- Easier and simpler provisioning of the tenant service: if an EVPN tunnel is configured in an IRB backhaul R-VPLS, there is no need to provision the IRB IP addresses in the VPRN. This makes the provisioning easier to automate and saves IP addresses from the tenant IP space.
- Higher scalability of the IRB backhaul R-VPLS: if EVPN tunnels are enabled, BUM traffic is suppressed in the EVPN-VXLAN IRB backhaul R-VPLS service (it is not required). As a result, the number of VXLAN bindings in IRB backhaul R-VPLS services with EVPN tunnels can be much higher.

As an example, the VPRN 30 and VPLS 301 configurations on PE-1, PE-2, and PE-3 are shown. Similar configurations are needed in PE-4, PE-5, and PE-6.

```
# on PE-1:
configure
    service
        vprn 30 name "VPRN30" customer 1 create
            interface "int-PE-1-CE-1" create
                address 172.16.0.254/24
                sap 1/2/1:30 create
                exit
            exit
            interface "int-evi-301" create
                vpls "evi-301"
                    evpn-tunnel
                exit
            exit
            no shutdown
        exit
        vpls 301 name "evi-301" customer 1 create
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 301 create
            exit
            bgp
                route-distinguisher 192.0.2.1:301
                route-target export target:64500:301 import target:64500:301
            exit
            bgp-evpn
                ip-route-advertisement
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
```

```
# on PE-2:
configure
    service
        vprn 30 name "VPRN30" customer 1 create
            interface "int-evi-301" create
                vpls "evi-301"
                    evpn-tunnel
                exit
            exit
            bgp-ipvpn
                mpls
                    auto-bind-tunnel
                        resolution-filter
                            ldp
                        exit
                        resolution filter
                    exit
                    route-distinguisher 192.0.2.2:30
                    vrf-target target:64500:30
                    no shutdown
                exit
            exit
            no shutdown
        exit
        vpls 301 name "evi-301" customer 1 create
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 301 create
            exit
            bgp
                route-distinguisher 192.0.2.2:301
                route-target export target:64500:301 import target:64500:301
            exit
            bgp-evpn
                ip-route-advertisement
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
```

```
# on PE-3:
```

```
configure
    service
        vprn 30 name "VPRN30" customer 1 create
            interface "int-evi-301" create
                vpls "evi-301"
                    evpn-tunnel
                exit
            exit
            bgp-ipvpn
                mpls
                    auto-bind-tunnel
                        resolution-filter
                            ldp
                        exit
                        resolution filter
                    exit
                    route-distinguisher 192.0.2.3:30
                    vrf-target target:64500:30
                    no shutdown
                exit
            exit
            no shutdown
        exit
        vpls 301 name "evi-301" customer 1 create
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 301 create
            exit
            bgp
                route-distinguisher 192.0.2.3:301
                route-target export target:64500:301 import target:64500:301
            exit
            bgp-evpn
                ip-route-advertisement
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
```

As shown in the preceding output, the configuration in the three nodes (PE-1/2/3) for VPLS 301 and VPRN 30 is similar to the configuration of VPLS 201 and VPRN 20 in the previous scenario, however, when the **evpn-tunnel** command is added to the VPRN interface, there is no need to configure an IP interface address. The option **evpn-tunnel** can be enabled independently of **ip-route-advertisement** (although no route type 5 advertisements are sent in that case).

A VPRN supports regular IRB backhaul R-VPLS services as well as EVPN tunnel R-VPLS services. A maximum of eight R-VPLS services with **ip-route-advertisement** enabled per VPRN is supported (in any combination of regular IRB R-VPLS or EVPN tunnel R-VPLS services). EVPN tunnel R-VPLS services do not support SAPs or SDP-bindings. No frames are flooded in an EVPN tunnel R-VPLS service, and, in fact no inclusive multicast routes are exchanged in R-VPLS services that are configured as EVPN tunnels.

The **show service id vxlan destinations** command for an R-VPLS service configured as an EVPN tunnel shows <egress VTEP, VNI> bindings excluded from Mcast, in other words, the VXLAN bindings are not used to flood BUM traffic:

```
*A:PE-2# show service id 301 vxlan destinations
```

```
===============================================================================
Egress VTEP, VNI
===============================================================================
Instance    VTEP Address                            Egress VNI  EvpnStatic Num
 Mcast       Oper State                              L2 PBR     SupBcasDom MACs
-------------------------------------------------------------------------------
1           192.0.2.1                               301         evpn       1
 -           Up                                      No          No
1           192.0.2.3                               301         evpn       1
 -           Up                                      No          No
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 2
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

The process followed upon receiving a route type 5 on a regular IRB R-VPLS interface (previous scenario) differs from the one for an EVPN tunnel type (this scenario):

- IRB backhaul R-VPLS VPRN interface:
  - When a route type 2 that includes an IP address is received and it becomes active, the MAC/IP information is added to the FDB and ARP tables. This can be checked with the **show router arp** command and the **show service id fdb detail** command.
  - When a route type 5 is received on (for instance) PE-2, and becomes active for the R-VPLS service, the IP prefix is added to the VPRN routing table regardless of the existence of a route type 2 that can resolve the GW IP address. If a packet is received from the WAN side and the IP lookup hits an entry for which the GW IP (IP next-hop) does not have an active ARP entry, the system will ARP to get the MAC. If the ARP is resolved but the MAC is unknown in the FDB table, the system will flood the ARP message into the R-VPLS multicast list. Routes type 5 can be checked in the routing table with the **show router route-table** command and the **show router fib** command.
- EVPN tunnel R-VPLS VPRN interface:
  - When a route type 2 is received and becomes active, the MAC address is added to the FDB (only). This MAC address is normally a GW MAC.
  - When a route type 5 is received on (for instance) PE-1, the system looks for the GW MAC. The IP prefix is added to the VPRN routing table with next hop equal to EVPN-tunnel GW MAC; for example, ET-02:13:ff:00:00:6a is an EVPN tunnel with GW MAC 02:13:ff:00:00:6a. The GW MAC is added from the GW MAC extended community sent along with the route type 5 for prefix 172.16.6.0/24. If a packet is received from CE-1 and the IP lookup hits an entry for which the next hop is an EVPN tunnel: GW MAC, the system looks up the GW MAC in the FDB. Normally a route type 2 with the GW MAC has already been received so that the GW MAC has been added to the FDB. If the GW MAC is not present in the FDB, the packet will be dropped.
  - The IP prefixes with GW MACs as next hops for the setup in Figure 85: EVPN-VXLAN in EVPN-tunnel R-VPLS services are displayed in the **show router route-table** command, as follows:

```
*A:PE-1# show router 30 route-table
```

```
===============================================================================
Route Table (Service: 30)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
172.16.1.0/24                                 Local   Local     00h02m40s  0
       int-PE-1-CE-1                                               0
172.16.6.0/24                                 Remote  EVPN-IFF  00h00m36s  169
       int-evi-301 (ET-02:13:ff:00:00:6a)                          0
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The same routing policies are applied on the core PEs to prevent loops; see Use of routing policies to avoid routing loops in redundant PEs.

The **show service id 301 fdb detail** command can be used to look for the forwarding information for a GW MAC:

```
*A:PE-1# show service id 301 fdb detail

===============================================================================
Forwarding Database, Service 301
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
301        02:0f:ff:00:00:6a cpm                     Intf     03/02/22 11:52:54
301        02:13:ff:00:00:6a vxlan-1:                EvpnS:P  03/02/22 11:53:02
                             192.0.2.2:301
301        02:17:ff:00:00:6a vxlan-1:                EvpnS:P  03/02/22 11:53:09
                             192.0.2.3:301
-------------------------------------------------------------------------------
No. of MAC Entries: 3
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

IP prefix routes sent for EVPN tunnel R-VPLS services do not contain a GW IP (the GW IP will be zero) but convey a GW MAC address that is used in the peer VPRN routing table. The following output shows PE-2’s VPRN 30 interface MAC address sent to PE-1:

```
*A:PE-2# show router 30 interface "int-evi-301" detail | match "MAC "
MAC Address       : 02:13:ff:00:00:6a    Mac Accounting    : Disabled
```

When **ip-route-advertisement** is configured, PE-2 sends route type 5 messages to PE-1, as can be seen in the following BGP update for the route toward subnet 172.16.6.0/24 in overlay network 2, using the MAC as GW MAC:

```
# on PE-2:
configure
    service
        vpls "evi-301"
            bgp-evpn
                ip-route-advertisement
            exit
```

```
221 2022/03/02 11:54:51.734 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 90
    Flag: 0x90 Type: 14 Len: 45 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-IP-PREFIX Len: 34 RD: 192.0.2.2:301, tag: 0,
              ip_prefix: 172.16.6.0/24 gw_ip 0.0.0.0 Label: 301 (Raw Label: 0x12d)
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:301
        mac-nh:02:13:ff:00:00:6a
        bgp-tunnel-encap:VXLAN
"
```

In the VPRN 30 routing table on PE-2, IP prefixes are shown with an EVPN tunnel next-hop (GW MAC) as opposed to an IP next-hop, therefore, the user may think that no ARP entries are consumed by VPRN 30. However, internal ARP entries are still consumed in VPRN 30. Although not shown in the **show router 30 arp** command, the **summary** option shows the consumption of internal ARP entries for EVPN.

```
*A:PE-2# show router 30 route-table

===============================================================================
Route Table (Service: 30)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
172.16.1.0/24                                 Remote  EVPN-IFF  00h04m28s  169
       int-evi-301 (ET-02:0f:ff:00:00:6a)                           0
172.16.6.0/24                                 Remote  BGP VPN   00h02m40s  170
       192.0.2.4 (tunneled)                                         10
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

There are no entries in the ARP table:

```
*A:PE-2# show router 30 arp

===============================================================================
ARP Table (Service: 30)
===============================================================================
IP Address      MAC Address       Expiry    Type   Interface
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================
```

One internal BGP-EVPN ARP entry is consumed, as can be seen as follows:

```
*A:PE-2# show router 30 arp summary

============================================================
ARP Table Summary (Service: 30)
============================================================
Local ARP Entries    : 1
Static ARP Entries   : 0
Dynamic ARP Entries  : 0
Managed ARP Entries  : 0
Internal ARP Entries : 0
BGP-EVPN ARP Entries : 1
------------------------------------------------------------
No. of ARP Entries   : 2
============================================================
```

The number of BGP-EVPN ARP entries in the **show router 30 arp summary** command matches the number of remote valid GW MACs for VPRN 30.

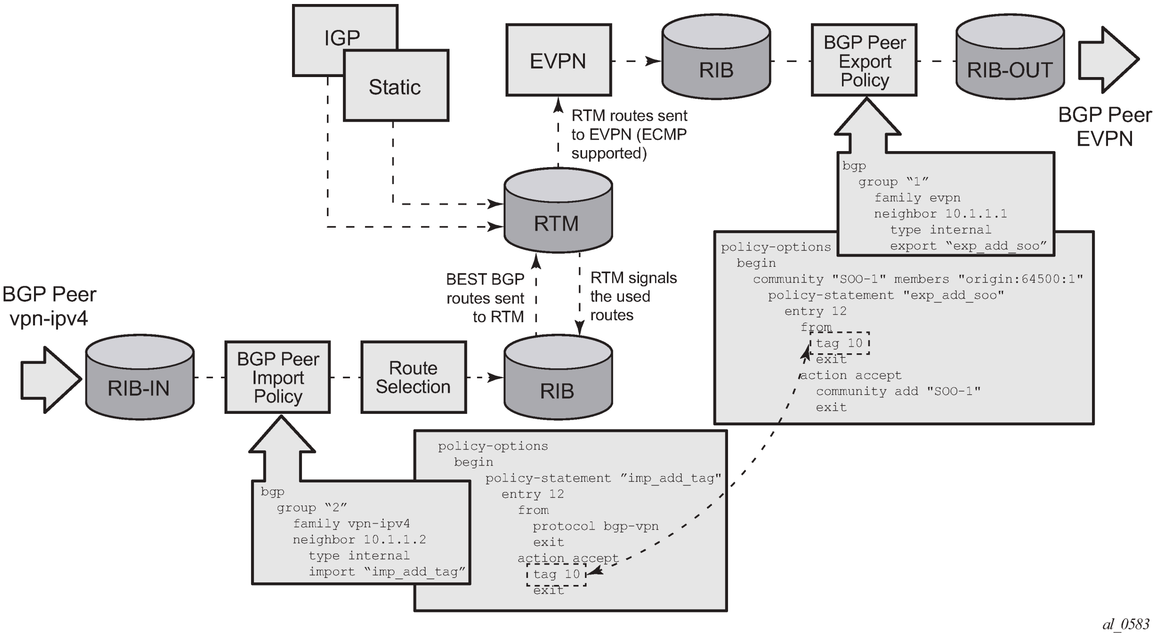
Routing policies for IP prefixes in EVPN

Routing policies are supported for IP prefixes imported or exported through BGP EVPN. The default import and export behavior for IP prefixes in EVPN can be modified by the use of routing policies applied either at peer level (**config router bgp group/group neighbor import/export**) or VPLS level (**config service vpls bgp vsi-import/vsi-export**).

When applying routing policies to control the distribution of prefixes between EVPN and IP-VPN, the user must take into account that both families are completely separated as far as BGP is concerned and that when prefixes from a family are imported in the RTM, the BGP attributes are lost to the other family. The use of tags allows the controlled distribution of prefixes across the two families.

Figure 86: Routing policies for egress EVPN routes illustrates how VPN-IPv4 routes are imported into the RTM and then passed onto EVPN for its own processing. VPN-IPv4 routes can be tagged at ingress and this tag is preserved throughout the RTM and EVPN processing so that the tag can be matched by the egress BGP routing policy. In this example, egress EVPN routes matching tag 10, are modified to add a site-of-origin (SOO) community origin:64500:1.

Figure 86: Routing policies for egress EVPN routes



```
bgp
  group “1”
    family evpn
    neighbor 10.1.1.1
      type internal
      export “exp_add_soo”
```

```
policy-options
  begin
    community "SOO-1" members "origin:64500:1"
      policy-statement "exp_add_soo"
        entry 12
          from
            tag 10
            exit
          action accept
            community add "SOO-1"
            exit
```

```
bgp
  group “2”
    family vpn-ipv4
    neighbor 10.1.1.2
      type internal
      import “imp_add_tag”
```

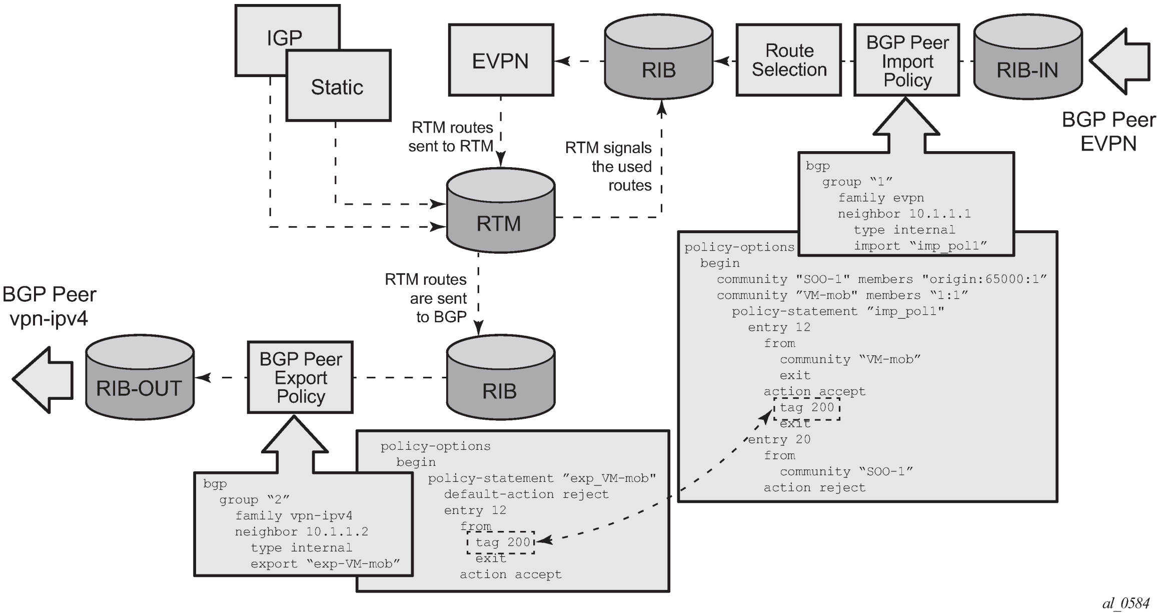
```
policy-options
  begin
      policy-statement ”imp_add_tag"
        entry 12
          from
            protocol bgp-vpn
            exit
          action accept
            tag 10
            exit
```

Policy tags can be used to match EVPN IP-prefixes that were learned not only from BGP VPN-IPv4 but also from other routing protocols. The tag range supported for each protocol is different:

```
*A:PE-2>config>router>policy-options>policy-statement>entry>action$ tag ?
  - no tag
  - tag <tag>

 <tag>                : tag-value      - accepts in decimal or hex
                                         [0x1..0xFFFFFFFF]H (for OSPF and ISIS)
                                         [0x1..0xFFFF]H (for RIP)
                                         [0x1..0xFFFFFFFF]H (for BGP)
                        param-name     - [32 chars max] - Must start and end with an at-sign
                                         (@)
```

Figure 87: Routing policies for ingress EVPN routes illustrates the reverse workflow: routes imported from EVPN and exported from RTM to BGPVPN-IPv4. In this example, EVPN routes received with community VM-mob are tagged with tag 200. At the egress VPN-IPv4 peers, only the routes with tag 200 are advertised.

Figure 87: Routing policies for ingress EVPN routes



```
bgp
  group “1”
    family evpn
    neighbor 10.1.1.1
      type internal
      import “imp_pol1”
```

```
policy-options
  begin
    community "SOO-1" members "origin:65000:1”
    community ”VM-mob" members “1:1”
      policy-statement ”imp_pol1"
        entry 12
          from
            community “VM-mob”
            exit
          action accept
            tag 200
            exit
        entry 20
          from
            community “SOO-1”
          action reject
```

```
bgp
  group “2”
    family vpn-ipv4
    neighbor 10.1.1.2
      type internal
      export “exp-VM-mob”
```

```
policy-options
  begin
      policy-statement ”exp_VM-mob"
        default-action reject
        entry 12
          from
            tag 200
            exit
          action accept
```

The preceding behavior and the use of tags is also valid for **vsi-import** and **vsi-export** policies. The behavior can be summarized in the following statements:

- For EVPN prefix routes received and imported in RTM:
  - Routes can be matched on communities and tags can be added to them. This works at peer level or vsi-import level.
  - Well-known communities **[no-export | no-export-subconfed | no-advertise]** also require that the routing policies add a tag if the user wants to modify the behavior when exporting to BGP.
  - Routes can be matched based on family EVPN.
  - Routes cannot be matched on prefix list.
- For exporting RTM to EVPN prefix routes:
  - Routes can be matched on tags and based on that, communities added, or routes accepted or rejected (dropped), and so on. This works at peer level or vsi-export level.
  - Tags can be added for static routes, RIP, OSPF, IS-IS, and BGP and then be matched in the vsi-export policy for EVPN IP-prefix route advertisement.
  - Tags cannot be added for direct routes.

Use of routing policies to avoid routing loops in redundant PEs

When redundant PE VPRN instances are connected to the same R-VPLS service (IRB backhaul or EVPN tunnel R-VPLS) with the **ip-route-advertisement** command enabled, routing loops can occur in two different use cases:

1. Routing loop caused by EVPN and IP-VPN interaction in the RTM.
2. Routing loop caused by EVPN in parallel R-VPLS services.

Policy configuration examples for both cases are provided in the following sections.

Routing loop use-case 1: EVPN and IP-VPN interaction

This use case refers to scenarios with redundant PEs and VPRNs connected to the same R-VPLS with **ip-route-advertisement**. The scenarios in Figure 84: EVPN-VXLAN for IRB backhaul R-VPLS services (EVPN-VXLAN for IRB Backhaul R-VPLS services) and Figure 85: EVPN-VXLAN in EVPN-tunnel R-VPLS services (EVPN-VXLAN in EVPN tunnel R-VPLS services) are examples of this use case. In both scenarios, the following process causes a routing loop:

1. PE-4 advertises IP prefix 172.16.6.0/24 with preference 170 (IP-VPN) to PE-2 and PE-3.
2. PE-2 and PE-3 import prefix 172.16.6.0/24 in the VPRN routing table. PE-2 re-advertises prefix 172.16.6.0/24 with preference 169 (EVPN) to PE-1 and PE-3; PE-3 re-advertises the IP prefix in EVPN to PE-1 and PE-2.
3. PE-2 and PE-3 already have the 172.16.6.0/24 prefix in the VPRN routing table with preference 170 (IP-VPN) but because the IP prefix from EVPN has a lower preference (169), the RTM installs the EVPN prefix in the VPRN routing table.
4. PE-2 advertises the EVPN-learned IP prefix 172.16.6.0/24 to all MP-BGP VPN-IPv4 peers, including PE-3; PE-3 advertises the prefix 172.16.6.0/24 to all MP-BGP VPN-IPv4 peers, including PE-2.
5. PE-2 receives the IP prefix 172.16.6.0/24 again from PE-3 and advertises it in EVPN again, creating a routing loop. The same thing happens in PE-3.

This routing loop also happens in traditional multi-homed IP-VPN scenarios where the PE-CE eBGP and MP-BGP VPN-IPv4/v6 protocols interact in the same VPRN RTM, with different router preferences. In either case (EVPN or eBGP interaction with MP-BGP) the issue can be solved by using routing policies and site-of-origin communities.

Routing policies are applied to PE-2 and PE-3 (also to PE-4 and PE-5) and allow the redundant PEs to reject their own generated routes to avoid the loops. These routing policies can be applied at vsi-import/export level or BGP group/neighbor level. The following output shows an example of routing policies applied at BGP neighbor level for PE-2 (similar policies are applied on PE-3/4/5). Neighbor or group level policies are the preferred way in this kind of use case: a single set of policies is sufficient, as opposed to a set of policies per service (if the policies are applied at vsi-import/export level).

The following policies are applied in the BGP group or BGP group/neighbor context on PE-2:

```
# on PE-2:
configure
    router Base
        policy-options
            begin
            community "SOO-PE-2"
                members "origin:2:1"
            exit
            community "SOO-PE-3"
                members "origin:3:1"
            exit
            policy-statement "add-SOO_on_export"
                entry 10
                    from
                        tag 2
                    exit
                    action accept
                        community add "SOO-PE-2"
                    exit
                exit
                entry 20
                    from
```

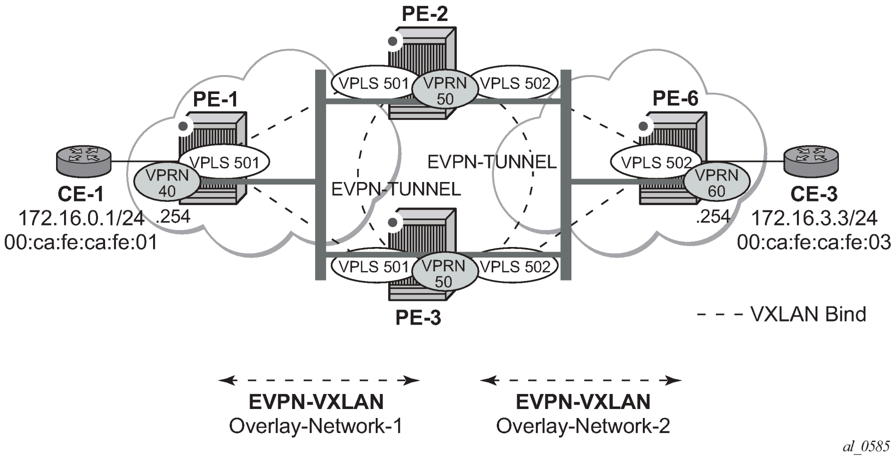
```
                        tag 3
                    exit
                    action accept
                        community add "SOO-PE-3"
                    exit
                exit
            exit
            policy-statement "reject_based_on_SOO"
                entry 10
                    from
                        community "SOO-PE-2"
                    exit
                    action drop
                    exit
                exit
                entry 20
                    from
                        community "SOO-PE-3"
                    exit
                    action drop
                    exit
                exit
            exit
            policy-statement "add-tag_to_bgp-vpn_routes"
                entry 10
                    from
                        protocol bgp-vpn
                    exit
                    action accept
                        tag 2
                    exit
                exit
            exit
            policy-statement "add-tag_to_bgp-evpn_routes"
                entry 10
                    from
                        family evpn
                    exit
                    action accept
                        tag 2
                    exit
                exit
            exit
            commit
        exit
        bgp
            group "DC"
                neighbor 192.0.2.1
                    import "add-tag_to_bgp-evpn_routes"
                exit
                neighbor 192.0.2.3
                    import "reject_based_on_SOO"
                    export "add-SOO_on_export"
                exit
            exit
            group "WAN"
                import "add-tag_to_bgp-vpn_routes"
            exit
```

EVPN and MP-BGP routes are tagged at import; on export, a site-of-origin community is added. Routes exchanged between the two redundant PEs are dropped if they are received by a PE with its own site-of-origin.

Routing loop use-case 2: EVPN in parallel R-VPLS services

If a VPRN is connected to more than one R-VPLS with **ip-route-advertisement** enabled, IP prefixes that belong to one R-VPLS are advertised into the other R-VPLS and vice versa. When redundant PEs are used, a routing loop will occur. Figure 88: EVPN in parallel R-VPLS services illustrates this use case. The example shows R-VPLS with an EVPN tunnel configuration, but the same routing loop occurs for regular IRB backhaul R-VPLS services.

Figure 88: EVPN in parallel R-VPLS services



The configuration of VPRN 50 as well as VPLS 501/502 and the required policies are as follows. For this use case, policies must be applied at vsi-import/export level because more granularity is required when modifying the imported/exported routes.

```
# on PE-2:
configure
    service
        vprn 50 name "VPRN50" customer 1 create
            interface "int-evi-501" create
                vpls "evi-501"
                    evpn-tunnel
                exit
            exit
            interface "int-evi-502" create
                vpls "evi-502"
                    evpn-tunnel
                exit
            exit
            no shutdown
        exit
        vpls 501 name "evi-501" customer 1 create
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 501 create
            exit
            bgp
                route-distinguisher 192.0.2.2:501
                vsi-export "vsi-export-policy-501"
                vsi-import "vsi-import-policy-501"
            exit
            bgp-evpn
```

```
                ip-route-advertisement
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            no shutdown
        exit
        vpls 502 name "evi-502" customer 1 create
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 502 create
            exit
            bgp
                route-distinguisher 192.0.2.2:502
                vsi-export "vsi-export-policy-502"
                vsi-import "vsi-import-policy-502"
            exit
            bgp-evpn
                ip-route-advertisement
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            no shutdown
        exit
    exit
    router Base
        policy-options
            begin
            community "exp_RVPLS501"
                members "origin:2:11" "target:64500:501"
            exit
            community "exp_RVPLS502"
                members "origin:2:11" "target:64500:502"
            exit
            community "SOO-PE-2-RVPLS"
                members "origin:2:11"
            exit
            community "SOO-PE-3-RVPLS"
                members "origin:3:11"
            exit
            community "SOO_PE-3_RVPLS501"
                members "origin:3:11" "target:64500:501"
            exit
            community "SOO_PE-3_RVPLS502"
                members "origin:3:11" "target:64500:502"
            exit
            policy-statement "vsi-export-policy-501"
                entry 10
                    from
                        tag 12
                    exit
                    action accept
                        community add "SOO_PE-3_RVPLS501"
                    exit
                exit
                entry 20
                    action accept
                        community add "exp_RVPLS501"
                    exit
                exit
            exit
            policy-statement "vsi-export-policy-502"
                entry 10
```

```
                    from
                        tag 12
                    exit
                    action accept
                        community add "SOO_PE-3_RVPLS502"
                    exit
                exit
                entry 20
                    action accept
                        community add "exp_RVPLS502"
                    exit
                exit
            exit
            policy-statement "vsi-import-policy-501"
                entry 10
                    from
                        community "SOO-PE-2-RVPLS"
                    exit
                    action drop
                    exit
                exit
                entry 20
                    from
                        community "SOO_PE-3_RVPLS501"
                    exit
                    action accept
                        tag 12
                    exit
                exit
                default-action accept
                exit
            exit
            policy-statement "vsi-import-policy-502"
                entry 10
                    from
                        community "SOO-PE-2-RVPLS"
                    exit
                    action drop
                    exit
                exit
                entry 20
                    from
                        community "SOO_PE-3_RVPLS502"
                    exit
                    action accept
                        tag 12
                    exit
                exit
                default-action accept
                exit
            exit
            commit
```

Troubleshooting and debug commands

For general information on EVPN and VXLAN troubleshooting and debug commands, see chapter EVPN for VXLAN Tunnels (Layer 2). The following information focuses on specific commands for Layer-3 applications.

When troubleshooting and operating an EVPN-VXLAN scenario with inter-subnet forwarding, it is important to check the IP prefixes and next-hops, as well as ARP tables and FDB tables:

- **show router** <..> **route-table**
- **show router** <..> **arp**
- **show service id** <..> **fdb detail**

ICMP commands can also help checking the connectivity. When traceroute is used on EVPN-VXLAN in EVPN tunnel interfaces, EVPN tunnel interface hops in the traceroute commands are showing the VPRN loopback address or the other non EVPN-tunnel interface address. In VPRN services where all the interfaces are EVPN tunnels, ICMP packets fail until an IP address is configured. The following output shows a traceroute from VPRN 30 in PE-1 to CE-6 and from PE-2 to CE-1 (see Figure 85: EVPN-VXLAN in EVPN-tunnel R-VPLS services):

```
*A:PE-1# traceroute router 30 172.16.6.6
traceroute to 172.16.6.6, 30 hops max, 40 byte packets
  1  0.0.0.0  * * *
  2  0.0.0.0  * * *
  3  172.16.6.254 (172.16.6.254)    4.98 ms  4.77 ms  4.97 ms
  4  172.16.6.6 (172.16.6.6)    7.64 ms  4.88 ms  5.14 ms
```

```
*A:PE-2# traceroute router 30 172.16.1.1
traceroute to 172.16.1.1, 30 hops max, 0 byte packets
No route to destination. Address: 172.16.1.1, Service: 30
```

When troubleshooting R-VPLS services, specifically R-VPLS services configured as EVPN tunnels, the limit of peer PEs per EVPN tunnel service is much higher than for a regular R-VPLS service because the egress <VTEP, VNI> bindings do not have to be added to the multicast flooding list. For this reason, the following **tools dump** command has been added to check the consumed/total EVPN tunnel next hops. The number of EVPN tunnel next hops matches the number of remote GW MAC addresses per EVPN tunnel R-VPLS service.

```
*A:PE-1# tools dump service id 501 evpn usage

Evpn Tunnel Interface IP Next Hop: 2/8189
```

Finally, when troubleshooting EVPN routes and routing policies, the **show router bgp routes evpn** command and its filters can help:

- Check that the expected routes are received, properly imported, and communities/tags added/replaced/removed.
- Check that the expected routes are sent, properly exported, and communities added/replaced/removed.

Examples of EVPN IP prefix routes including communities and tags are the following.

```
*A:PE-2# show router bgp routes evpn ?
  - evpn <evpn-type>

      auto-disc       - Display BGP EVPN Auto-Disc Routes
      eth-seg         - Display BGP EVPN Eth-Seg Routes
      incl-mcast      - Display BGP EVPN Inclusive-Mcast Routes
      ip-prefix       - Display BGP EVPN IPv4-Prefix Routes
      ipv6-prefix     - Display BGP EVPN IPv6-Prefix Routes
      mac             - Display BGP EVPN Mac Routes
      mcast-join-syn* - Display BGP EVPN Mcast Join Sync Routes
      mcast-leave-sy* - Display BGP EVPN Mcast Leave Sync Routes
      smet            - Display BGP EVPN Smet Routes
      spmsi-ad        - Display BGP EVPN Spmsi AD Routes
```

```
*A:PE-2# show router bgp routes evpn ip-prefix ?
  - ip-prefix [hunt|detail] [rd <rd>] [prefix <ip-prefix/ip-prefix-length>]
      [community <comm-id>] [tag <tag>] [next-hop <next-hop>] [aspath-regex <reg-exp>]

 <hunt|detail>        : keywords
 <rd>                 : {<ip-addr:comm-val>|
                        <2byte-asnumber:ext-comm-val>|
                        <4byte-asnumber:comm-val>}
 <ip-prefix/ip-pref*> : ip-address     - a.b.c.d (host bits must be 0)
                        mask           - [0..32]
 <comm-id>            : <as-number1:comm-val1>|<ext-comm>|
                        <well-known-comm>
                        ext-comm        - <type>:{<ip-address:comm-val1>|
                                                <as-number1:comm-val2>|
                                                <as-number2:comm-val1>}
                        as-number1      - [0..65535]
                        comm-val1       - [0..65535]
                        type            - target|origin
                        ip-address      - a.b.c.d
                        comm-val2       - [0..4294967295]
                        as-number2      - [0..4294967295]
                        well-known-comm - null|no-export|no-export-subconfed|
                                          no-advertise
 <tag>                : [0..4294967295] | MAX-ET
 <next-hop>           : ipv4-address   - a.b.c.d
                        ipv6-address   - x:x:x:x:x:x:x:x   (eight 16-bit pieces)
                                         x:x:x:x:x:x:d.d.d.d
                                         x - [0..FFFF]H
                                         d - [0..255]D
 <reg-exp>            : [80 chars max]
```

Routing policy "vsi-export-policy-502" adds community "origin:2:11 target:64500:502" to the outgoing routes, as can be verified as follows:

```
*A:PE-2# show router bgp routes evpn ip-prefix hunt prefix 172.16.1.0/24
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
Network        : n/a
Nexthop        : 192.0.2.2
Path Id        : None
To             : 192.0.2.1
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
```

```
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : n/a
Connector      : None
Community      : origin:2:11 target:64500:502
                 mac-nh:02:13:ff:00:01:33 bgp-tunnel-encap:VXLAN
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.1
Origin         : IGP
AS-Path        : No As-Path
EVPN type      : IP-PREFIX
ESI            : n/a
Tag            : 0
Gateway Address: 02:13:ff:00:01:33
Prefix         : 172.16.1.0/24
Route Dist.    : 192.0.2.2:502
MPLS Label     : VNI 502
Route Tag      : 0
Neighbor-AS    : n/a
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
---snip---
```

On PE-2, policy "add-tag_to_bgp-evpn_routes" adds route tag 2 to all BGP EVPN routes, as can be verified in the following output:

```
*A:PE-2# show router bgp routes evpn ip-prefix prefix 172.16.1.0/24 detail
===============================================================================
 BGP Router ID:192.0.2.2       AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
Original Attributes

Network        : n/a
Nexthop        : 192.0.2.1
Path Id        : None
From           : 192.0.2.1
Res. Nexthop   : 192.168.12.1
Local Pref.    : 100                    Interface Name : int-PE-2-PE-1
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64500:201 bgp-tunnel-encap:VXLAN
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.1
Flags          : Used Valid Best IGP
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : IP-PREFIX
ESI            : n/a
Tag            : 0
Gateway Address: 172.16.0.1
Prefix         : 172.16.1.0/24
Route Dist.    : 192.0.2.1:201
MPLS Label     : VNI 201
Route Tag      : 0
```

```
Neighbor-AS    : n/a
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h04m30s

Modified Attributes

Network        : n/a
Nexthop        : 192.0.2.1
Path Id        : None
From           : 192.0.2.1
Res. Nexthop   : 192.168.12.1
Local Pref.    : 100                    Interface Name : int-PE-2-PE-1
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64500:201 bgp-tunnel-encap:VXLAN
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.1
Flags          : Used Valid Best IGP
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : IP-PREFIX
ESI            : n/a
Tag            : 0
Gateway Address: 172.16.0.1
Prefix         : 172.16.1.0/24
Route Dist.    : 192.0.2.1:201
MPLS Label     : VNI 201
Route Tag      : 2
Neighbor-AS    : n/a
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h04m30s

-------------------------------------------------------------------------------
---snip---
```

Conclusion

SR OS supports not only the EVPN control plane for VXLAN tunnels in Layer 2 applications but also the simultaneous use of EVPN and VXLAN for VPN customers (tenants) with intra and inter-subnet connectivity requirements. R-VPLS services can be configured to provide default gateway connectivity to hosts, IRB backhaul connectivity to VPRN services, and EVPN tunnel connectivity to VPRN services.

When configured to do so, EVPN can advertise IP prefixes and interact with the VPRN RTM to propagate IP prefix connectivity between EVPN and other routing protocols in the VPRN, including IP-VPN. This example has shown how to configure R-VPLS services for all these functions, as well as how to configure routing policies for EVPN-based IP prefixes.

EVPN Interconnect Ethernet Segments

This chapter provides information about EVPN Interconnect Ethernet Segments.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written based on SR OS Release 15.0.R4, but the CLI in the current edition corresponds to SR OS Release 21.2.R2.

Chapters EVPN for MPLS Tunnels, EVPN for VXLAN Tunnels (Layer 2) and EVPN-MPLS Interconnect for EVPN-VXLAN VPLS Services are prerequisite reading.

Overview

SR OS supports Interconnect Ethernet Segments (I-ESs) for VXLAN as per the IETF *draft-ietf-bess-dci-evpn-overlay*. An I-ES is a virtual Ethernet Segment (vES) that allows Data Center Gateways (DCGWs) with two BGP instances (one for EVPN-MPLS and one for EVPN-VXLAN) to handle redundancy in VXLAN access networks. I-ESs support the RFC 7432 multi-homing functions, including single-active and all-active, ESI-label based split-horizon filtering, Designated Forwarder (DF) election, aliasing, and backup functions on remote EVPN-MPLS PEs.

The chapter EVPN-MPLS Interconnect for EVPN-VXLAN VPLS Services describes how VPLS services with two BGP instances are configured and describes a redundant mechanism referred to as Multi-homed anycast configuration for dual BGP-instance VPLS services. The use of I-ESs is recommended over this anycast configuration.

In addition to the EVPN multi-homing features, the main advantages of the I-ES solution compared to the redundant solution (described in Anycast Redundant Solution for Dual BGP-instance Services) are as follows:

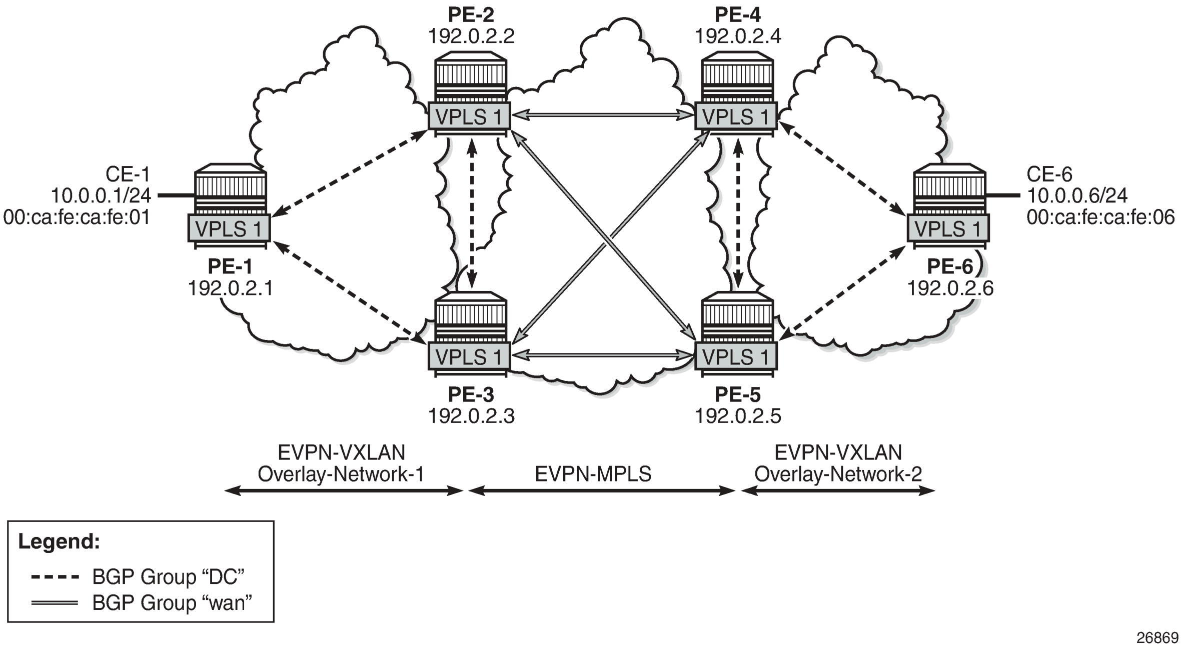
- The use of I-ES for redundancy in dual BGP-instance services allows local SAPs on the DCGWs. This is not supported in the anycast solution.
- P2MP mLDP can be provisioned to transport Broadcast, Unknown unicast, and Multicast (BUM) traffic between DCs that use I-ES, without any risk of packet duplication. As described in The use of provider tunnels on multi-homed anycast solutions, packet duplication may occur in the anycast DCGW solution when mLDP is used in the WAN.

When EVPN-MPLS networks are interconnected to EVPN-VXLAN networks, the I-ES concept and procedures apply only to the access VXLAN network; the EVPN-MPLS network does not modify its existing behavior compared to any other ES.

Configuration

Figure 89: EVPN-MPLS interconnect for EVPN-VXLAN - BGP topology shows the topology and infrastructure configuration, which are the same as in chapter EVPN-MPLS Interconnect for EVPN-VXLAN VPLS Services. Read that chapter to see how the PEs are configured at port, IS-IS, and base BGP level.

Figure 89: EVPN-MPLS interconnect for EVPN-VXLAN - BGP topology



PE-1, PE-2, and PE-3 simulate a data center (DC), shown as Overlay-Network-1, where PE-2 and PE-3 are DCGWs. In the same way, PE-4, PE-5, and PE-6 simulate a remote DC, Overlay-Network-2. Inside each DC, EVPN-VXLAN is used and the two DCGW pairs are connected by EVPN-MPLS. CE-1 and CE-6 are end-to-end connected by EVPN without any VLAN or Pseudowire (PW) hand-off, maintaining all the EVPN advantages across the DC Interconnect (DCI) network.

Interconnect Ethernet Segment (I-ES) configuration

After the base infrastructure is configured (interfaces, IGP, LDP in the core, and BGP EVPN peering sessions, as per Figure 89: EVPN-MPLS interconnect for EVPN-VXLAN - BGP topology), two I-ESs configured on the DCGWs show the use of the Interconnect Ethernet Segments.

The I-ES "I-ES231" is configured on PE-2 and PE-3 as follows:

```
# on PE-2:
configure
```

```
    service
        system
            bgp-evpn
                ethernet-segment "I-ES231" virtual create
                    esi 00:23:23:23:23:23:23:00:00:01
                    service-carving
                        mode manual
                        manual
                            preference non-revertive create
                                value 150
                            exit
                            evi 101 to 200
                        exit
                    exit
                    multi-homing all-active
                    network-interconnect-vxlan 1
                    service-id
                        service-range 1 to 100
                        service-range 101 to 200
                    exit
                    no shutdown
                exit
```

```
# on PE-3:
configure
    service
        system
            bgp-evpn
                ethernet-segment "I-ES231" virtual create
                    esi 00:23:23:23:23:23:23:00:00:01
                    service-carving
                        mode manual
                        manual
                            preference non-revertive create
                                value 50
                            exit
                            evi 101 to 200
                        exit
                    exit
                    multi-homing all-active
                    network-interconnect-vxlan 1
                    service-id
                        service-range 1 to 100
                        service-range 101 to 200
                    exit
                    no shutdown
                exit
```

On PE-1 and PE-2, the preceding configuration associates I-ES "I-ES231" with the VXLAN instance 1 in services contained in the range VPLS 1 to 100 and 101 to 200. The I-ES is modeled as a virtual ES, where:

- Two commands are needed within the ethernet-segment context: **network-interconnect-vxlan** and **service-id service-range** <*svc-id*> **[to** <*svc-id*>**]**.
  - The **[no] network-interconnect-vxlan** command identifies the VXLAN instance associated with the virtual ES. Only value 1 is supported in SR OS Release 21.2.R2.

```
*A:PE-2>config>service>system>bgp-evpn>eth-seg# network-interconnect-vxlan ?
  - no network-interconnect-vxlan
  - network-interconnect-vxlan <instance>
```

```
 <instance>           : [1..1]
```

  The **[no] network-interconnect-vxlan** command is rejected in non-virtual ESs:

```
*A:PE-2>config>service>system>bgp-evpn# ethernet-segment "ES-23" create
*A:PE-2>config>service>system>bgp-evpn>eth-seg# network-interconnect-vxlan 1
MINOR: SVCMGR #8065 Supported only on virtual ethernet segments
```

  - The **[no] service-range** command associates the specific service range with the ES. The ES must be configured as **network-interconnect-vxlan** before any service range can be added.
  - The other ES association options (port, lag, sdp, vc-id-range, dot1q, and qinq) are blocked in the ES when a **network-interconnect-vxlan** instance is configured.
  - The rest of the ES configuration options are supported. The **source-bmac-lsb** is blocked because the I-ES cannot be associated with I-VPLS or PBB-Epipe services.
  - All the services with two BGP instances associate the VXLAN destinations and ingress VXLAN instance with the ES.

- Multiple services (for example, 1 to 200 in the CLI above) can be associated with the same ES.
  - Up to eight service ranges per VXLAN instance can be configured. Ranges may overlap within the same ES (and not between different ESs). In this example, two non-overlapping ranges are configured to show the service range configuration, although a single range containing all the services could have been configured.
  - The service range may be configured before the service is, and it can be changed on the fly without having to disable the ES first.
- When the **network-interconnect-vxlan** I-ES is configured, the ES operational state depends exclusively on the ES admin state.
  - Because the I-ES is not associated with a physical port or SDP, when testing the non-revertive service-carving manual mode, an ethernet-segment shutdown/no shutdown will result in the node sending its own administrative preference and "Do not preempt" (pref/DP) values, and taking over if pref/DP is higher than the current DF. This is because when the ES is no shutdown, the peer ES routes are not present at the EVPN application layer, so the PE will send its own admin pref/DP values. Therefore, for I-ESs, the non-revertive mode will only work for node failures. See the chapter for more information about the preference-based and non-revertive DF election modes.
- There are no restrictions in the service-carving mode supported by I-ESs. In this example, preference-based service-carving is configured, but modes auto and (non-preference-based) manual are also supported.
- As described in the Preference-based and Non-revertive EVPN DF Election chapter, the service-carving context is configured with an EVI range that will pick up the lowest preference value when electing a DF for the service, whereas the non-configured EVI services will pick up the highest value when electing a DF. In this example, this means that, of the services allowed in the I-ES, that is, 1 to 200, services 1 to 100 will elect the highest Preference PE as DF, whereas services 101 to 200 will elect the lowest Preference PE.

PE-4 and PE-5 are configured with I-ES "I-ES451". The configuration of I-ES451 is similar to that of I-ES231; only single-active mode is configured, instead of all-active mode.

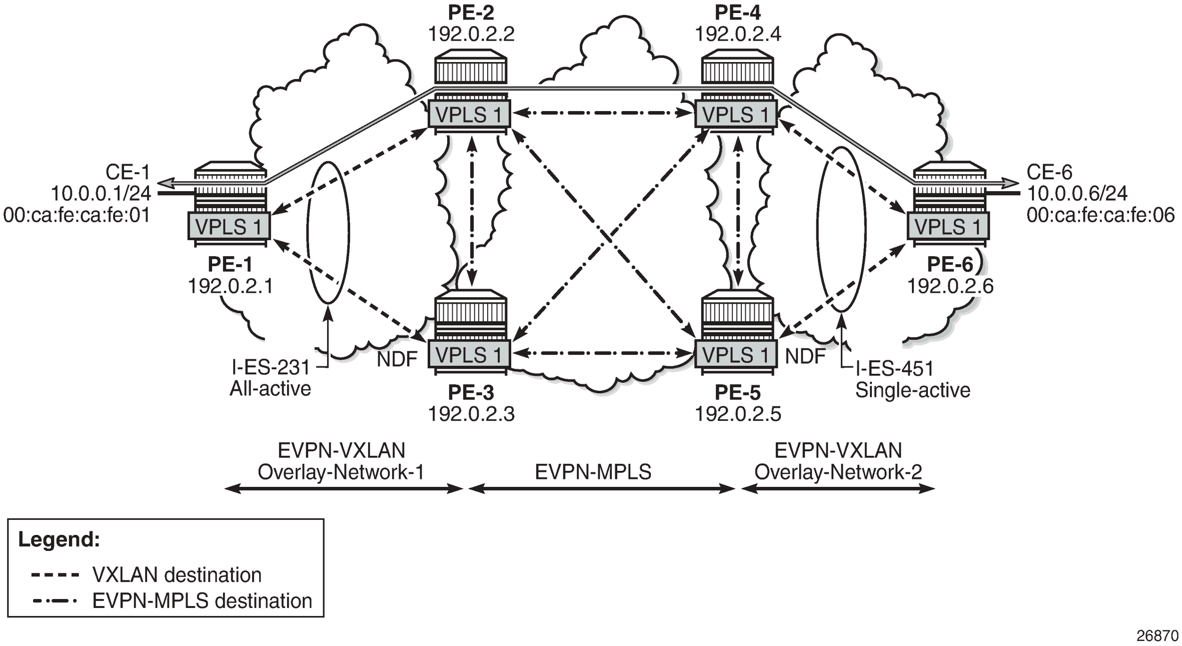
```
# on PE-4:
configure
    service
        system
            bgp-evpn
```

```
                ethernet-segment "I-ES451" virtual create
                    esi 00:45:45:45:45:45:45:00:00:01
                    service-carving
                        mode manual
                        manual
                            preference non-revertive create
                                value 150
                            exit
                            evi 101 to 200
                        exit
                    exit
                    multi-homing single-active
                    network-interconnect-vxlan 1
                    service-id
                        service-range 1 to 100
                        service-range 101 to 200
                    exit
                    no shutdown
                exit
```

```
# on PE-5:
configure
    service
        system
            bgp-evpn
                ethernet-segment "I-ES451" virtual create
                    esi 00:45:45:45:45:45:45:00:00:01
                    service-carving
                        mode manual
                        manual
                            preference non-revertive create
                                value 50
                            exit
                            evi 101 to 200
                        exit
                    exit
                    multi-homing single-active
                    network-interconnect-vxlan 1
                    service-id
                        service-range 1 to 100
                        service-range 101 to 200
                    exit
                    no shutdown
                exit
```

In this example, VPLS 1 will be configured and associated with the preceding I-ESs. Figure 90: VPLS service and association with I-ESs shows an example of VPLS 1 and how it is associated with the I-ESs.

Figure 90: VPLS service and association with I-ESs



The configuration of VPLS 1 for PE-1, PE-2, and PE-3 is as follows. VPLS 101 is also configured in all the PEs in a similar way as VPLS 1, but not shown here. Also, the VPLS 1 configuration on the rest of the PEs is equivalent to the one in PE-1, PE-2, and PE-3, as follows:

```
# on PE-1:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            vxlan instance 1 vni 1 create
            exit
            bgp
            exit
            bgp-evpn
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
                mpls
                    shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/2/1:1 create
                no shutdown
            exit
            no shutdown
```

```
# on PE-2:
```

```
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            vxlan instance 1 vni 1 create
            exit
            bgp
                route-distinguisher 192.0.2.2:1
            exit
            bgp 2
                route-distinguisher 192.0.2.2:2
            exit
            bgp-evpn
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
                mpls bgp 2
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
```

```
# on PE-3:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            vxlan instance 1 vni 1 create
            exit
            bgp
                route-distinguisher 192.0.2.3:1
            exit
            bgp 2
                route-distinguisher 192.0.2.3:2
            exit
            bgp-evpn
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
                mpls bgp 2
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
```

As in the case of any other ESs, the association of instance and service is based on the ES configuration and there is no extra configuration required at the service level to make that association. The existing

show commands that are used to check the status of the ES can be used to check the I-ESs. For example, on I-ES231:

```
*A:PE-2# show service system bgp-evpn ethernet-segment name "I-ES231" all

===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : I-ES231
Eth Seg Type            : Virtual
Admin State             : Enabled            Oper State         : Up
ESI                     : 00:23:23:23:23:23:23:00:00:01
Multi-homing            : allActive          Oper Multi-homing  : allActive
ES SHG Label            : 524278
Source BMAC LSB         : <none>
VXLAN Instance Id       : 1
ES Activation Timer     : 3 secs (default)
Oper Group              : (Not Specified)
Svc Carving             : manual             Oper Svc Carving   : manual
Cfg Range Type          : lowest-pref

-------------------------------------------------------------------------------
DF Pref Election Information
-------------------------------------------------------------------------------
Preference     Preference     Last Admin Change        Oper Pref      Do No
Mode           Value                                   Value          Preempt
-------------------------------------------------------------------------------
non-revertive  150            05/03/2021 13:01:53      150            Enabled
-------------------------------------------------------------------------------

-------------------------------------------------------------------------------
EVI Ranges
-------------------------------------------------------------------------------
From                                    To
-------------------------------------------------------------------------------
101                                     200
-------------------------------------------------------------------------------
ISID Ranges: <none>
===============================================================================

===============================================================================
EVI Information
===============================================================================
EVI                 SvcId               Actv Timer Rem      DF
-------------------------------------------------------------------------------
1                   1                   0                   yes
101                 101                 0                   no
-------------------------------------------------------------------------------
Number of entries: 2
===============================================================================

-------------------------------------------------------------------------------
DF Candidate list
-------------------------------------------------------------------------------
EVI                                     DF Address
-------------------------------------------------------------------------------
1                                       192.0.2.2
1                                       192.0.2.3
101                                     192.0.2.2
101                                     192.0.2.3
-------------------------------------------------------------------------------
Number of entries: 4
-------------------------------------------------------------------------------
```

```
-------------------------------------------------------------------------------
---snip---

===============================================================================
Vxlan Instance Service Ranges
===============================================================================
Svc Range Start          Svc Range End            Last Changed
-------------------------------------------------------------------------------
1                        100                      05/03/2021 13:01:53
101                      200                      05/03/2021 13:01:53
-------------------------------------------------------------------------------
Number of Entries: 2
===============================================================================
```

The **show service id 1 vxlan instance 1 oper-flags** command shows the status of a VXLAN instance in the service. A service VXLAN instance will raise the oper-flag **MhStandby** (multi-homing standby) due to any of the following reasons:

- The PE is (single-active) non-Designated Forwarder (NDF) for that I-ES.
- The VXLAN service is added to the I-ES and either the ES is **shutdown** or **bgp-evpn>mpls** is **shutdown** in all the services included in the ES.

For example, because PE-5 is an NDF in I-ES451, the MhStandby flag will show "true":

```
*A:PE-5# show service id 1 vxlan instance 1 oper-flags

===============================================================================
VPLS VXLAN oper flags
===============================================================================
MhStandby                              : true
===============================================================================
```

EVPN route handling in dual BGP-instance VPLSs with I-ES

The configuration of I-ESs on DCGWs with two BGP instances has the following impact on the advertisement and process of the BGP-EVPN routes:

- EVPN MAC/IP routes:
  - MAC/IP routes received on the EVPN-MPLS BGP instance will be re-advertised to the EVPN-VXLAN BGP instance with the ESI set to zero in SR OS Release 21.2.R2.
  - EVPN-VXLAN PE/NVEs (Network Virtual Edge devices) in the DC will receive the same MAC address from two (or more) different MAC/IP routes from the DCGWs. The EVPN-VXLAN PE/NVEs will perform regular EVPN MAC/IP route selection.
  - MAC/IP routes received on the EVPN-VXLAN BGP instance will be re-advertised to the EVPN-MPLS BGP instance with the configured non-zero I-ESI value, assuming the VXLAN instance is not in the MhStandby operational state. MAC/IP routes received on the EVPN-VXLAN BGP instance will be dropped if the VXLAN instance is in the MhStandby state.
  - EVPN-MPLS PEs in the WAN will receive the same MAC address from two (or more) DCGWs, set with the same ESI. EVPN-MPLS PEs will perform regular aliasing and backup functions.
- ES routes are exchanged for the I-ES. They should be sent only to the MPLS network and not to the VXLAN side. This can be achieved by using router policies. In any case, because ES routes use an ES-import route-target extended community, they should not be imported by VXLAN PEs.

- Auto-discover per ES (AD per-ES) and AD per-EVI routes are also advertised for the I-ES. They should be sent only to the MPLS network and not to the VXLAN network. As for ES routes, router policies can be used to prevent AD routes being sent to VXLAN peers.

Required BGP policies to avoid control plane loops

Usually, the use of router policies is required when I-ESs are used for redundancy, to avoid control plane loops with MAC/IP routes. The control plane loops to be avoided are as follows:

1. Loops created by remote MAC addresses (learned on remote PE SAPs):
   a. Remote EVPN-MPLS MAC/IP routes are re-advertised into EVPN-VXLAN with a Site of Origin (SOO) extended community (added by a BGP peer or vsi-export policy) identifying the DCGW pair. The other DCGW in the pair will drop EVPN-VXLAN MAC routes tagged with the self SOO. Router policies to add SOO and drop routes received with self SOO are needed.
   b. Also, when remote EVPN-VXLAN MAC/IP routes are re-advertised into EVPN-MPLS, the DCGWs will automatically drop EVPN-MPLS MAC/IP routes received with their own non-zero I-ESI. No router policies are needed for this.
2. Loops created by local SAP MAC addresses:
   a. Local SAP MACs are learned and MAC/IP routes are advertised into both BGP instances. The MAC/IP routes advertised in the EVPN-VXLAN instance will be dropped by the peer based on the SOO router policies, as described in (1a) above, and DCGW local MACs will always be learned over the EVPN-MPLS destinations between the DCGWs.
   b. Because only EVPN-MPLS destinations exist between the DCGWs, EVPN-VXLAN MAC/IP and IMET routes exchanged between the DCGWs will be discarded and EVPN-VXLAN destinations will not be created between them.

As an example, the following BGP peer policies on PE-2 and PE-3 achieve the goals described above (similar policies would be configured on PE-4 and PE-5) and summarized as follows:

- Avoid sending service VXLAN routes to MPLS peers, and service MPLS routes to VXLAN peers.
- Avoid sending AD and ES routes to VXLAN peers.
- Add SOO to VXLAN routes to be sent to the ES peer.
- Drop VXLAN routes received from the ES peer.

```
# on PE-2, PE-3:
configure
    router Base
        policy-options
            begin
            community "mpls"
                members "bgp-tunnel-encap:MPLS"
            exit
            community "vxlan"
                members "bgp-tunnel-encap:VXLAN"
            exit
            community "SOO-DCGW-23"
                members "origin:64500:23"
            exit
```

The following policy prevents the router from sending service VXLAN routes to MPLS peers:

```
            policy-statement "allow only mpls"
```

```
                entry 10
                    from
                        community "vxlan"
                        family evpn
                    exit
                    action drop
                    exit
                exit
            exit
```

The following policy makes sure the router exports only routes that include the VXLAN encapsulation:

```
            policy-statement "allow only vxlan"
                entry 10
                    from
                        community "vxlan"
                        family evpn
                    exit
                    action accept
                    exit
                exit
                default-action drop
                exit
            exit
```

The following import policy avoids importing routes with self SOO:

```
            policy-statement "drop SOO-DCGW-23"
                entry 10
                    from
                        community "SOO-DCGW-23"
                        family evpn
                    exit
                    action drop
                    exit
                exit
            exit
```

The following export policy adds SOO but only to VXLAN routes. This allows the peer to drop routes based on the SOO, without affecting the MPLS routes.

```
            policy-statement "add SOO to vxlan routes"
                entry 10
                    from
                        community "vxlan"
                        family evpn
                    exit
                    action accept
                        community add "SOO-DCGW-23"
                    exit
                exit
                default-action accept
                exit
            exit
```

The BGP configuration for PE-2 and PE-3 is as follows:

```
# on PE-2:
configure
    router Base
        autonomous-system 64500
        router-id 192.0.2.2
        bgp
            family evpn
            vpn-apply-import
            vpn-apply-export
            rapid-withdrawal
            rapid-update evpn
            group "dc"
                type internal
                export "allow only vxlan"
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.3
                    import "drop SOO-DCGW-23"
                    export "add SOO to vxlan routes"
                exit
            exit
            group "wan"
                type internal
                export "allow only mpls"
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
            no shutdown
```

```
# on PE-3:
configure
    router Base
        autonomous-system 64500
        router-id 192.0.2.3
        bgp
            family evpn
            vpn-apply-import
            vpn-apply-export
            rapid-withdrawal
            rapid-update evpn
            group "dc"
                type internal
                export "allow only vxlan"
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.2
                    import "drop SOO-DCGW-23"
                    export "add SOO to vxlan routes"
                exit
            exit
            group "wan"
                type internal
                export "allow only mpls"
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
            no shutdown
```

Single-active multi-homing operation

When the I-ES is configured as **single-active** and **no shutdown** (assuming at least one service is associated), the DCGWs will send ES and AD routes as usual for any ES, and run DF election based on the ES routes, with the candidate list being pruned by the AD routes.

In Figure 90: VPLS service and association with I-ESs, PE-4 and PE-5 are configured with I-ES451, which is a single-active ES. The NDF for a service (PE-5 for VPLS 1 in the example) will perform the following tasks:

- The VXLAN instance on the NDF will enter the MhStandby state and will block ingress and egress traffic on the VXLAN destinations associated with the I-ES.

```
A:PE-5# show service id 1 vxlan instance 1 oper-flags

===============================================================================
VPLS VXLAN oper flags
===============================================================================
MhStandby                              : true
===============================================================================
```

- MAC/IP routes and FDB process:
  - Advertised MAC/IP routes that are associated with the VXLAN instance are withdrawn.
  - Advertised MAC/IP routes corresponding to local SAP MAC addresses or EVPN-MPLS binding MAC addresses are withdrawn if they were advertised to the EVPN-VXLAN instance.
  - Received MAC/IP routes associated with the VXLAN instance are not installed in FDB. The MAC routes will show as "used" in the **show router bgp routes evpn mac** commands; however, only the MAC addresses received from MPLS (in particular from the ES peer) will be programmed. As an example, the following CLI output shows how MAC address 00:ca:fe:ca:fe:06 is learned on PE-4 (DF) and associated with the VXLAN destination to PE-6, whereas the MAC address is installed associated with an MPLS destination (remote ES) on PE-5 (NDF).

```
*A:PE-4# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:ca:fe:ca:fe:01 eES:                    Evpn     05/03/21 13:10:58
                             00:23:23:23:23:23:23:00:00:01
1          00:ca:fe:ca:fe:06 vxlan-1:                Evpn     05/03/21 13:10:58
                             192.0.2.6:1
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

```
*A:PE-5# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
```

```
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:ca:fe:ca:fe:01 eES:                    Evpn     05/03/21 13:10:58
                             00:23:23:23:23:23:23:00:00:01
1          00:ca:fe:ca:fe:06 eES:                    Evpn     05/03/21 13:10:58
                             00:45:45:45:45:45:45:00:00:01
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

- Inclusive Multicast Ethernet Tag (IMET) routes process:
  - IMET-Assisted Replication with replicator role (IMET-AR-R) routes are withdrawn if the VXLAN instance enters the MhStandby state. Only the DF will advertise the IMET-AR-R routes. For more information on AR, see chapter Layer 2 Multicast Optimization for EVPN-VXLAN — Assisted Replication.
  - IMET-Ingress Replication advertisements (IMET-IR) routes, in case of NDF (or the MhStandby state), are controlled by the **config>service>vpls>bgp-evpn>vxlan# [no] send-imet-ir-on-ndf** command.
    - By default, the command is enabled and the router will advertise IMET-IR routes even if the PE is NDF (MhStandby). This will attract BUM traffic (even if the NDF ends up dropping it); however, attracting BUM traffic will also speed up convergence in case of DF switchover. The command works for single-active and all-active.
    - If disabled, the router will withdraw the IMET-IR routes when the PE is NDF and will not attract BUM traffic.

In spite of not sending BUM or unicast traffic, the NDF for a service still creates the VXLAN bindings; however, they are not associated with any MAC addresses and they are flagged as non-multicast capable, or "-" in the Mcast column of the following command:

```
*A:PE-5# show service id 1 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
Instance    VTEP Address                            Egress VNI  EvpnStatic Num
 Mcast       Oper State                              L2 PBR     SupBcasDom MACs
-------------------------------------------------------------------------------
1           192.0.2.6                               1           evpn       0
 -           Up                                      No          No
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================


===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

The I-ES DF PE for the service (PE-4) will continue advertising IMET and MAC/IP routes for the associated VXLAN instance. Forwarding will also happen as usual on the DF VXLAN bindings. When the DF PE receives BUM traffic from VXLAN, it will send it, adding the egress ESI label if needed.

All-active multi-homing operation

The same considerations as in single-active for ES and AD routes and DF election apply to all-active multi-homing. In Figure 90: VPLS service and association with I-ESs, PE-2 and PE-3 are configured with I-ES231, which is an all-active ES. The NDF PE for a service (PE-3 for VPLS 1, in the example) will show the following behavior:

- The VXLAN instance on the NDF will not enter the MhStandby state because it will still forward unicast traffic:

```
*A:PE-3# show service id 1 vxlan instance 1 oper-flags

===============================================================================
VPLS VXLAN oper flags
===============================================================================
MhStandby                               : false
===============================================================================
```

- MAC/IP routes and FDB process: MAC/IP routes are received, installed, and advertised as in the DF router.
- IMET routes process:
  - As in the single-active case, IMET-AR-R routes are withdrawn on the NDF. Only the DF will advertise the IMET-AR-R routes.
  - Also, as in the single-active case, IMET-IR advertisement from the NDF will be controlled by the **config>service>vpls>bgp-evpn>vxlan# [no] send-imet-ir-on-ndf** command. Advertising the IMET-IR route from the NDF will attract BUM traffic from the VXLAN PEs to the NDF, even though the unknown unicast traffic will be forwarded only when it is safe to do so. See section All-active multi-homing and unknown unicast forwarding on the NDF for more information about unknown unicast forwarding.

Contrary to the behavior in single-active multi-homing, in all-active, the NDF will forward unknown unicast to the VXLAN PEs as usual, but block broadcast and multicast in the upstream and downstream direction. In our example, the NDF for VPLS 1 (PE-3) will show the VXLAN destinations created as "U" (Unknown unicast) in the Mcast column of the **show service id 1 vxlan** command, as follows:

```
*A:PE-3# show service id 1 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
Instance    VTEP Address                             Egress VNI   EvpnStatic Num
 Mcast       Oper State                               L2 PBR      SupBcasDom MACs
-------------------------------------------------------------------------------
1           192.0.2.1                                1            evpn       1
 U           Up                                       No           No
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                        Num. Macs     Last Change
-------------------------------------------------------------------------------
No Matching Entries
```

```
===============================================================================
```

All-active multi-homing and unknown unicast forwarding on the NDF

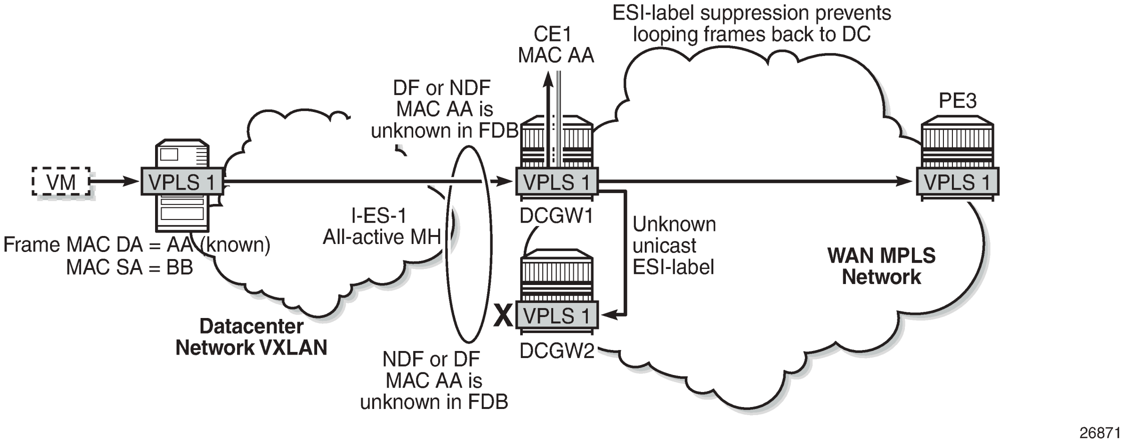
The unknown unicast traffic will be transmitted on the (all-active multi-homing) NDF in the upstream and downstream directions only in those cases where there is no risk of packet duplication. The router considers there is no risk when transmitting an unknown unicast packet on the NDF if:

- Unknown unicast packet arrives without an ESI label.
- Unknown unicast packet arrives without a BUM label (label advertised by an IMET route as opposed to a MAC/IP route).
- Unknown unicast packet passes a MAC Source Address (MAC SA) suppression (MAC SA lookup does not yield an entry associated with the I-ES).

The following examples show how unknown unicast traffic is handled in all-active I-ESs.

Figure 91: All-active multi-homing and unknown unicast example 1 shows an example with two DCGWs where (all-active) I-ES-1 is defined.

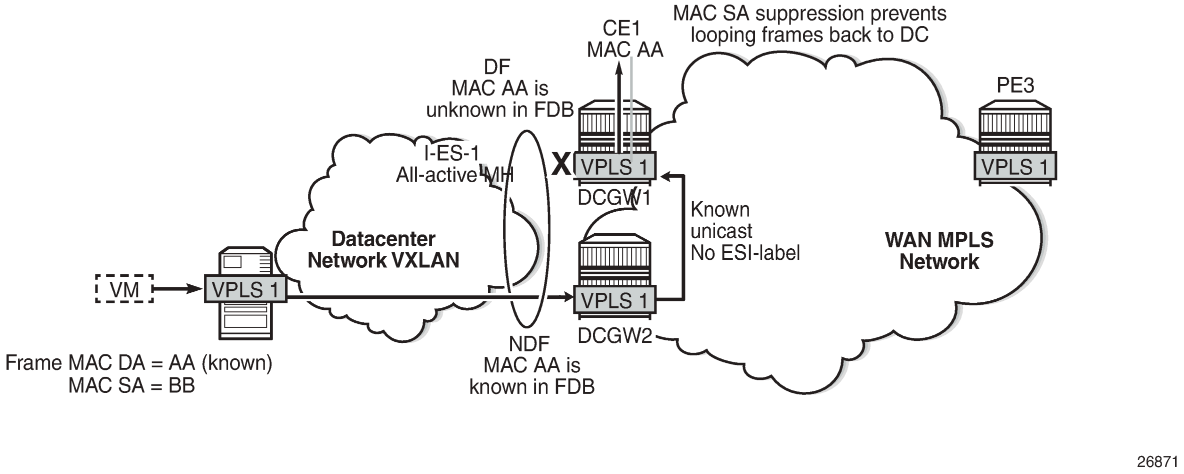
Figure 91: All-active multi-homing and unknown unicast example 1



The VXLAN PE/NVE transmits known unicast traffic, whereas DCGW1 has not learned the MAC address yet. Regardless of the DCGW1 being DF or NDF, it will accept unknown unicast and will flood to local SAPs and EVPN destinations. When sending to DCGW2, the router will send the ESI label identifying the I-ES. DCGW2 will not send unknown traffic back to the DC due to the ESI-label suppression on the I-ES.

Figure 92: All-active multi-homing and unknown unicast example 2 shows a similar example where the VXLAN node sends known unicast with MAC Destination Address (MAC DA) "AA" to DCGW2.

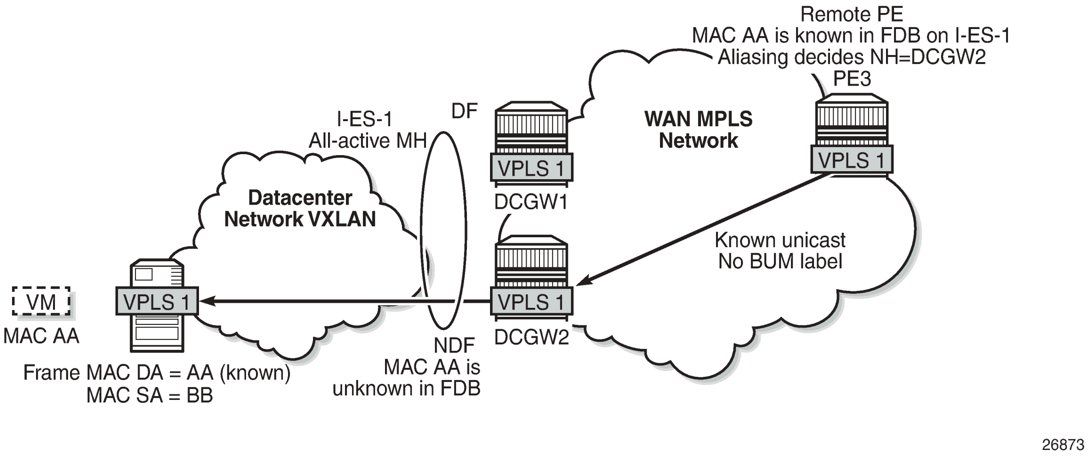
Figure 92: All-active multi-homing and unknown unicast example 2



DCGW2 does a MAC lookup and sends the frame as known unicast to DCGW1 via the EVPN-MPLS destination. However, MAC AA is unknown in DCGW1 for some reason (such as FDB limit exceeded, SAP failure, and so on). In this case, DCGW1 will flood the frame to CE1 and not to the VXLAN network. Even though the frame is not coming with an ESI label, the DCGW1 router does a MAC SA suppression and will not send unknown unicast frames to the I-ES. MAC SA suppression means that the router will do a MAC SA lookup on the FDB and will suppress the flooding to the I-ES if the MAC SA is learned on the I-ES (as in Figure 92: All-active multi-homing and unknown unicast example 2).

Figure 93: All-active multi-homing and unknown unicast example 3 shows an example in which the NDF forwards "no-risk" unknown unicast traffic to avoid black-holes.

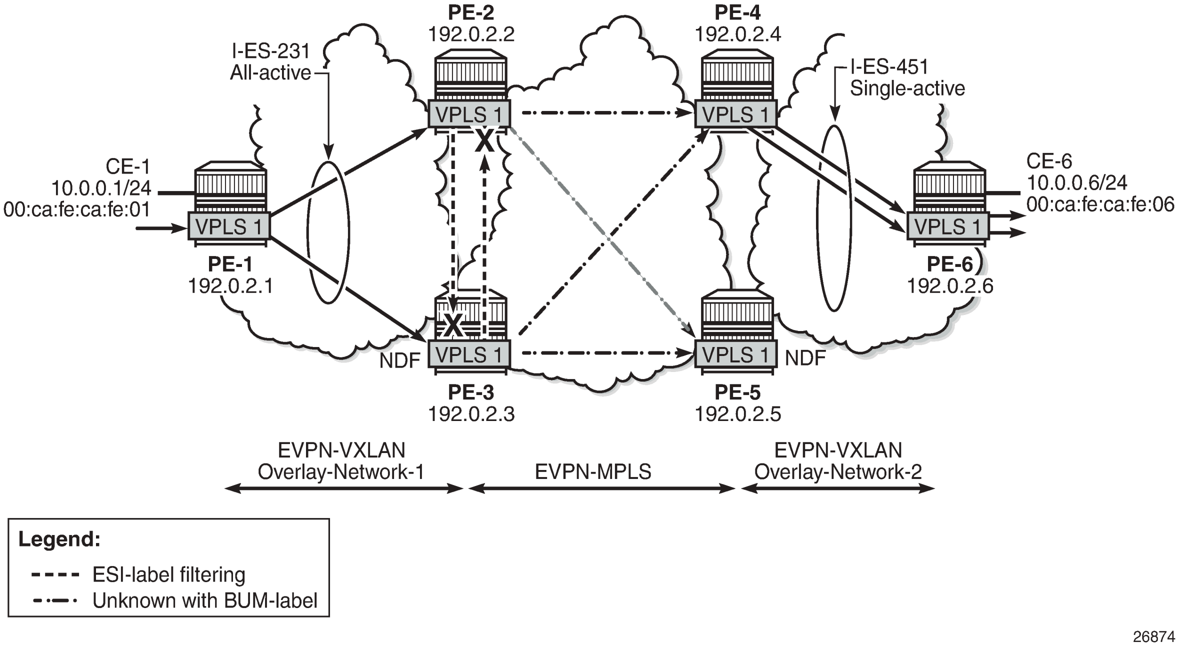
Figure 93: All-active multi-homing and unknown unicast example 3



PE3 receives unicast traffic with MAC DA = AA. The MAC address is known in the FDB and associated with I-ES-1; therefore, because PE3 is configured to do aliasing to DCGW1 and DCGW2 (bgp-evpn>mpls# ecmp 2), a packet hash determines that it has to be sent to DCGW2 (NDF). The packet arrives at DCGW2 with a unicast label. DCGW2 does a lookup and MAC AA is unknown for some reason (such as FDB limit exceeded, MAC not learned yet, and so on). In this case, DCGW2 will forward the packet to the I-ES VXLAN bindings, even if it is NDF. This behavior avoids black-hole periods in the network for unicast traffic.

Finally, in some cases, the unknown unicast forwarding behavior on the NDF may cause some transient packet duplication that can be avoided by configuring the **no send-imet-ir-on-ndf** command. The following example shows the use of this command to avoid transient packet duplication. Figure 94: All-active multi-homing and send-imet-ir-on-ndf shows how transient packet duplication may occur with the default setting **send-imet-ir-on-ndf**.

Figure 94: All-active multi-homing and send-imet-ir-on-ndf



Transient packet duplication may occur when sending unknown unicast from CE-1 to CE-6, if **send-imet-ir-on-ndf** is configured in PE-3 and PE-2. To show this, we clear the FDBs in all the PEs in the example as well as the ARP caches on the CEs.

The following command is executed in all the PEs and CEs:

```
*A:PE-1# clear service id 1 fdb all
*A:PE-1#
*A:PE-1# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

The following command clears the ARP table of the VPRN instance (defined in PE-1 using a loop) simulating CE-1:

```
*A:PE-1# clear router 300 arp all
*A:PE-1#
*A:PE-1# show router 300 arp
```

```
===============================================================================
ARP Table (Service: 300)
===============================================================================
IP Address      MAC Address       Expiry    Type   Interface
-------------------------------------------------------------------------------
10.0.0.1        00:ca:fe:ca:fe:01 00h00m00s Oth[I] local
-------------------------------------------------------------------------------
No. of ARP Entries: 1
===============================================================================
```

When ICMP traffic is sent from CE-1 to CE-6, a duplicate entry occurs on CE-1:

```
*A:PE-1# ping router 300 10.0.0.6
PING 10.0.0.6 56 data bytes
64 bytes from 10.0.0.6: icmp_seq=1 ttl=64 time=13.2ms.
64 bytes from 10.0.0.6: icmp_seq=1 ttl=64, duplicate.
64 bytes from 10.0.0.6: icmp_seq=2 ttl=64 time=5.27ms.
64 bytes from 10.0.0.6: icmp_seq=3 ttl=64 time=5.25ms.
64 bytes from 10.0.0.6: icmp_seq=4 ttl=64 time=4.73ms.
64 bytes from 10.0.0.6: icmp_seq=5 ttl=64 time=4.80ms.

---- 10.0.0.6 PING Statistics ----
5 packets transmitted, 5 packets received, 1 duplicate
round-trip min = 4.73ms, avg = 6.66ms, max = 13.2ms, stddev = 3.29ms
```

This duplicate entry occurs because the packet gets to CE-6 twice and CE-6 sends two unicast ICMP reply messages back. From the CE-1 packet walkthrough:

- PE-1 floods the packet to PE-2 and PE-3 because the CE-6 MAC DA is unknown and it has VXLAN multicast destinations to them.
- PE-2 floods the unknown unicast packet to all the remote PEs because it is DF for I-ES231. PE-2 will add an ESI label when sending to PE-3, and a BUM label when sending to all of them.
- PE-3 is NDF for I-ES231, but it floods the packet because the I-ES is all-active and the unknown unicast packet is considered low risk. The packet arrives with no ESI label, no BUM label (in VXLAN, VNIs are the same for unicast and BUM), and the MAC SA suppression passes because the packet is coming from the I-ES and not from MPLS. PE-3 uses a BUM label when flooding the packet and an ESI label when sending to PE-2.
- PE-4 receives two unknown unicast packets and forwards both to PE-6.
- PE-5 does not forward because it is NDF. This is true regardless of the I-ES being single-active or all-active (if all-active, the packet will not be forwarded because it arrives with a BUM label).

This packet duplication situation is transient and it will stop as soon as the two MAC addresses are learned on the PEs. However, if needed, this situation can be avoided by configuring **no send-imet-ir-on-ndf** (the BGP-EVPN VXLAN must be disabled first):

```
# on PE-2, PE-3:
configure
    service
        vpls "VPLS 1"
            bgp-evpn
                vxlan
                    shutdown
                    no send-imet-ir-on-ndf
                    no shutdown
```

This command will make the NDF (PE-3) withdraw the IMET-IR route; therefore, PE-1 will only flood unknown unicast packets to the DF (PE-2). The following IMET-IR routes are received on PE-1: one route sent by DF PE-2 for VPLS 1 and two routes for VPLS 101.

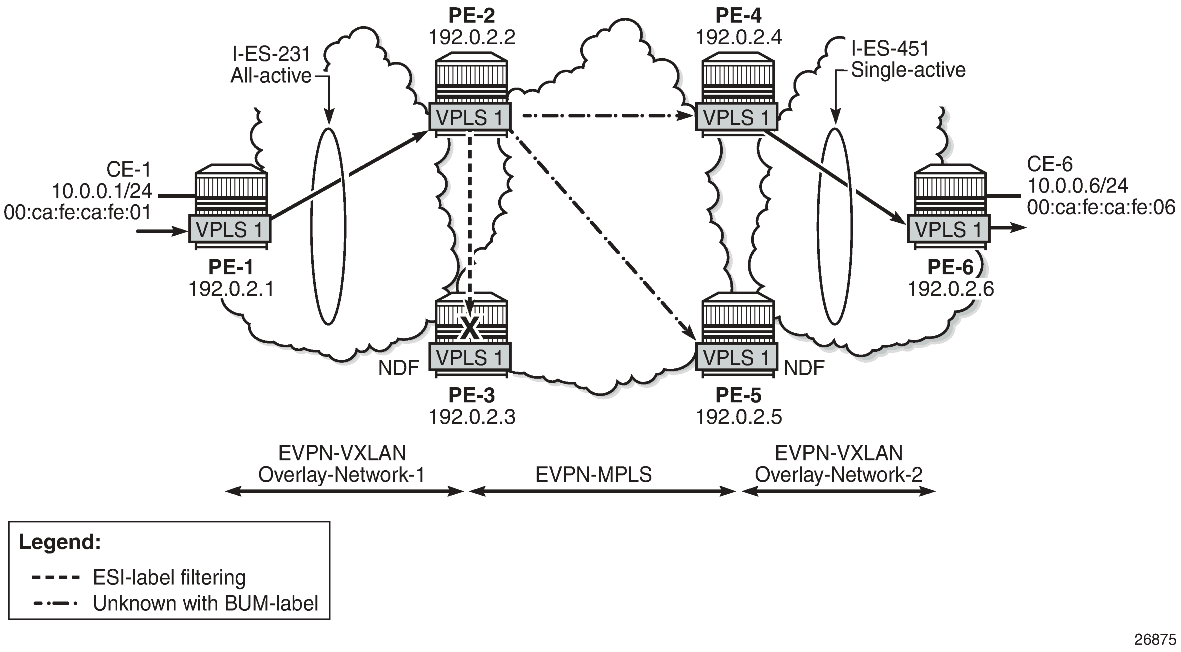
```
*A:PE-1# show router bgp routes evpn incl-mcast
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Flag  Route Dist.         OrigAddr
      Tag                 NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.2:1         192.0.2.2
      0                   192.0.2.2

u*>i  192.0.2.2:101       192.0.2.2
      0                   192.0.2.2

u*>i  192.0.2.3:101       192.0.2.3
      0                   192.0.2.3

-------------------------------------------------------------------------------
Routes : 3
===============================================================================
```

If a DF switchover occurs in the I-ES, the new DF would advertise the IMET-IR route and the new NDF would withdraw it.

After clearing FDBs and ARP caches again, the test is repeated with no packet duplication. Figure 95: All-active multi-homing and no send-imet-ir-on-ndf shows how PE-1 does not send unknown unicast to PE-3 (NDF) anymore and, therefore, there is no duplication.

```
*A:PE-1# ping router 300 10.0.0.6
PING 10.0.0.6 56 data bytes
64 bytes from 10.0.0.6: icmp_seq=1 ttl=64 time=15.3ms.
64 bytes from 10.0.0.6: icmp_seq=2 ttl=64 time=5.32ms.
64 bytes from 10.0.0.6: icmp_seq=3 ttl=64 time=5.33ms.
64 bytes from 10.0.0.6: icmp_seq=4 ttl=64 time=5.44ms.
64 bytes from 10.0.0.6: icmp_seq=5 ttl=64 time=4.98ms.

---- 10.0.0.6 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 4.98ms, avg = 7.26ms, max = 15.3ms, stddev = 4.00ms
```

Figure 95: All-active multi-homing and no send-imet-ir-on-ndf



Local SAPs and provider tunnels along with I-ES

As described in the Overview section, the main advantages of the I-ES solution over the anycast redundant solution for dual BGP-instance services are the support of local SAPs and P2MP mLDP trees without packet duplication. This section shows the configuration of local SAPs and provider tunnels along with I-ES in VPLS services. The local SAPs can, at the same time, belong to an ES or a vES.

As an example, VPLS 1 on PE-2 is reconfigured as follows (similar configuration on PE-3, with provider tunnel also configured on PE-4 and PE-5):

```
# on PE-2:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            vxlan instance 1 vni 1 create
            exit
            bgp
                route-distinguisher 192.0.2.2:1
            exit
            bgp 2
                route-distinguisher 192.0.2.2:2
            exit
            bgp-evpn
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
                mpls bgp 2
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            provider-tunnel
                inclusive
                    owner bgp-evpn-mpls
                    root-and-leaf
                    mldp
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap lag-1:1 create
                no shutdown
            exit
            no shutdown
```

To have EVPN multi-homing from a CE locally connected to PE-2 and PE-3, an additional ES is configured on PE-2 and PE-3 that will include the local SAPs in VPLS 1, as follows:

```
# on PE-2:
configure
    service
        system
            bgp-evpn
                ethernet-segment "I-ES231" virtual create
                    esi 00:23:23:23:23:23:23:00:00:01
                    service-carving
                        mode manual
                        manual
                            preference non-revertive create
                                value 150
                            exit
                            evi 101 to 200
                        exit
                    exit
                    multi-homing all-active
                    network-interconnect-vxlan 1
                    service-id
                        service-range 1 to 100
                        service-range 101 to 200
                    exit
                    no shutdown
                exit
                ethernet-segment "vES232" virtual create
                    esi 00:23:23:23:23:23:23:00:00:02
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 1
                    dot1q
                        q-tag-range 1
                    exit
                    no shutdown
                exit
```

Troubleshooting and debugging

Common troubleshooting commands to operate dual BGP-instance VPLS services are in the corresponding section of EVPN-MPLS Interconnect for EVPN-VXLAN VPLS Services. Also, ES and virtual ES can be troubleshot by using the commands described in chapter EVPN for MPLS Tunnels.

As well, the following **show** commands are specific to the use of I-ES in the router:

```
*A:PE-2# show service id 1 vxlan instance 1 oper-flags

================================================================================
VPLS VXLAN oper flags
================================================================================
MhStandby                               : false
================================================================================
```

```
*A:PE-2# show service vxlan-instance-using ethernet-segment

================================================================================
VXLAN Ethernet-Segment Information
================================================================================
SvcId       VXLAN Instance      ES Name                           Status
--------------------------------------------------------------------------------
1           1                   I-ES231                           DF
101         1                   I-ES231                           NDF
================================================================================
```

```
*A:PE-2# show service vxlan-instance-using ethernet-segment "I-ES231"

================================================================================
VXLAN Ethernet-Segment Information
================================================================================
SvcId                            VXLAN Instance                   Status
--------------------------------------------------------------------------------
1                                1                                DF
101                              1                                NDF
================================================================================
```

Conclusion

Based on *draft-ietf-bess-dci-evpn-overlay*, SR OS supports the connectivity of Layer 2 EVPN-VXLAN services to an EVPN-MPLS network. This chapter complements the chapter EVPN-MPLS Interconnect for EVPN-VXLAN VPLS Services by describing how redundancy can be improved with the use of I-ES multi-homing, a concept standardized in *draft-ietf-bess-dci-evpn-overlay*.

EVPN Interconnect Ethernet Segments in Dual EVPN-VXLAN Instance VPLS Services

This chapter provides information about EVPN Interconnect Ethernet Segments in Dual EVPN-VXLAN Instance VPLS Services.

Topics in this chapter include:

Applicability

The information and configuration in this chapter are based on SR OS Release 21.7.R1. EVPN multi-homing on dual VXLAN instance VPLS services is supported on SR OS Release 19.10.R1, and later.

Overview

Some service providers are deploying large Data Centers (DCs) where SR OS routers are used as leaf switches in a VXLAN fabric. In those cases, all-active multi-homing can provide redundancy and maximize the bandwidth utilization.

SR OS supports Interconnect Ethernet Segments (I-ESs) for VXLAN as per RFC 9014. Chapter EVPN Interconnect Ethernet Segments (I-ESs) describes how I-ESs allow Data Center Gateways (DCGWs) with two BGP instances (one for EVPN-MPLS and one for EVPN-VXLAN) to handle redundancy in VXLAN access networks, as supported in SR OS 15.0.R4, and later.

This chapter describes similar scenarios with EVPN-VXLAN in the core network instead of EVPN-MPLS. The following scenarios are supported with I-ES in VXLAN instance 1:

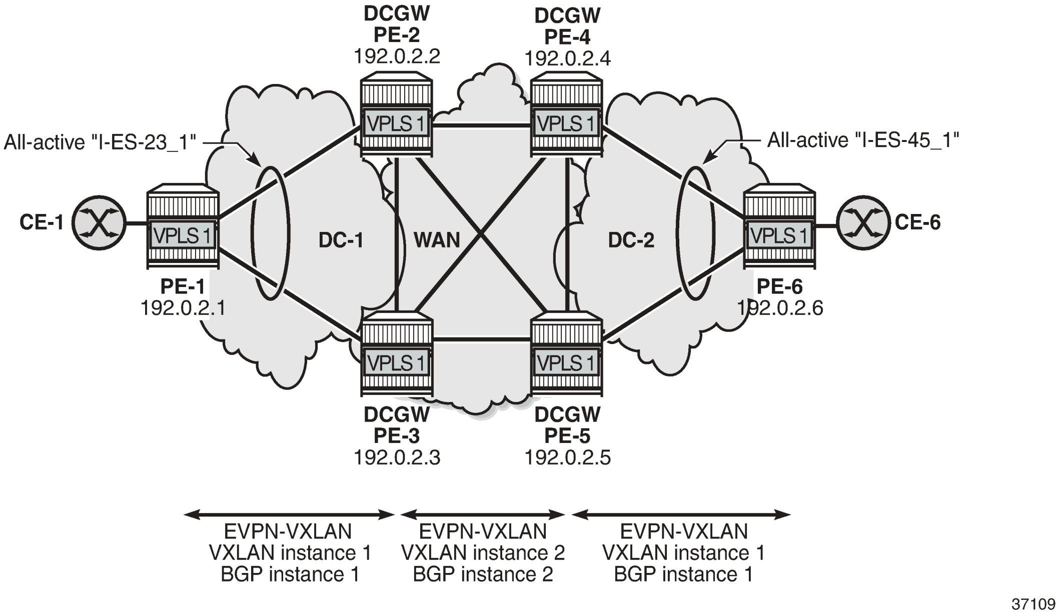
- dual instance VPLS with two EVPN-VXLAN instances
- dual instance VPLS with one EVPN-VXLAN instance and one static VXLAN instance
- dual instance VPLS with one EVPN-VXLAN instance and one EVPN-MPLS instance

The first two of these scenarios are described in this chapter.

CLI

Figure 96: Sample topology shows VPLS 1 with different EVPN-VXLAN instances: VXLAN instance 1 in DC 1 (and DC2) and VXLAN instance 2 in the WAN.

Figure 96: Sample topology



On DCGW PE-2, the following all-active I-ES is configured for VXLAN instance 1 and service id 1:

```
# on DCGW PE-2:
configure
    service
        system
            bgp-auto-rd-range 192.0.2.2 comm-val 1 to 1000
            bgp-evpn
                ethernet-segment "I-ES-23_1" virtual create
                    esi 00:23:23:23:23:23:23:00:00:01
                    service-carving
                        mode manual
                        manual
                            preference create
                                value 100
                            exit
                            evi 1
                        exit
                    exit
                    multi-homing all-active
                    network-interconnect-vxlan 1
                    service-id
                        service-range 1
                    exit
                    no shutdown
                exit
            exit
```

The following command configures VPLS 1 with dual EVPN-VXLAN instance. VXLAN instance 1 is a member of the I-ES and VXLAN instance 2 is configured with **mh-mode network** and **auto-disc-route-advertisement**:

```
# on DCGW PE-2:
configure
```

```
    service
        vpls 1 name "VPLS 1" customer 1 create
            vxlan instance 1 vni 11 create
                rx-discard-on-ndf bum
            exit
            vxlan instance 2 vni 12 create
            exit
            bgp
                route-distinguisher auto-rd
                route-target export target:64500:11 import target:64500:11
            exit
            bgp 2
                route-distinguisher auto-rd
                route-target export target:64500:12 import target:64500:12
            exit
            bgp-evpn
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    ecmp 2
                    default-route-tag 11
                    auto-disc-route-advertisement
                    no shutdown
                exit
                vxlan bgp 2 vxlan-instance 2
                    ecmp 2
                    default-route-tag 12
                    auto-disc-route-advertisement
                    mh-mode network
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/2/1:1 create
                no shutdown
            exit
            no shutdown
        exit
```

By default, the multi-homing mode for EVPN-VXLAN is access, but for VXLAN instance 2, it is modified to **mh-mode network**. The following error is raised when attempting to configure VXLAN instance 1—as a member of an I-ES—with **mh-mode network**:

```
*A:PE-2>config>service>vpls>bgp-evpn>vxlan# mh-mode network
MINOR: SVCMGR #7886 cannot modify evpn - not supported when vxlan instance is a member of an
 ethernet-segment
```

With **mh-mode network** configured, it is mandatory to configure **auto-disc-route-advertisement**; for **mh-mode access**, it is optional. When **auto-disc-route-advertisement** is enabled in an access instance associated to an I-ES, AD per-ES/EVI routes and MAC/IP routes are advertised for the I-ES.

The following AD per-EVI route is sent by DCGW PE-2:

```
13 2021/09/07 14:35:47.456 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 81
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-AD Len: 25 RD: 192.0.2.2:1 ESI: 00:23:23:23:23:23:23:00:00:01,
                     tag: 0 Label: 11 (Raw Label: 0xb) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        origin:64500:23
        target:64500:11
        bgp-tunnel-encap:VXLAN
"
```

For MAC routes and their ESI value for an access VXLAN instance, the following redistribution considerations apply.

- With **mh-mode access** and **auto-disc-route-advertisement** configured, MAC routes are redistributed from the instance network to the instance access with the I-ESI if present, regardless of the original ESI.
- With **mh-mode access** and **no auto-disc-route-advertisement**, MAC routes are redistributed with zero ESI, regardless of the original ESI.

The following EVPN-MAC route is sent by DCGW PE-2 with I-ESI 00:23:23:23:23:23:23:00:00:01 of "I-ES-23_1":

```
79 2021/09/07 14:36:18.380 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 89
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-MAC Len: 33 RD: 192.0.2.2:1 ESI: 00:23:23:23:23:23:23:00:00:01,
         tag: 0, mac len: 48 mac: 00:ca:fe:ca:fe:05, IP len: 0, IP: NULL, label1: 11
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        origin:64500:23
        target:64500:11
        bgp-tunnel-encap:VXLAN
"
```

The following ES route is sent by DCGW PE-2:

```
17 2021/09/07 14:35:47.456 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 71
    Flag: 0x90 Type: 14 Len: 34 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-ETH-SEG Len: 23 RD: 192.0.2.2:0
              ESI: 00:23:23:23:23:23:23:00:00:01, IP-Len: 4 Orig-IP-Addr: 192.0.2.2
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        df-election::DF-Type:Preference/DP:0/DF-Preference:100/AC:1
        target:23:23:23:23:23:23
"
```

The following commands are not supported when **mh-mode network** is configured:

- proxy-arp/nd
- assisted replication
- source-vtep-security

Attempting to enable these unsupported commands while a BGP-EVPN VXLAN instance has **mh-mode network** triggers error messages, as follows:

```
*A:PE-2>config>service>vpls# proxy-arp
MINOR: SVCMGR #8005 Cannot create proxy arp - not supported when a bgp-evpn vxlan instance has
 mh-mode network
```

```
*A:PE-2>config>service>vpls# proxy-nd
MINOR: SVCMGR #8008 Cannot create proxy nd - not supported when a bgp-evpn vxlan instance has
 mh-mode network
```

```
*A:PE-2>config>service>vpls>vxlan# assisted-replication replicator
MINOR: SVCMGR #8111 Cannot change assisted-replicated role - not supported when vxlan instance
 is in use by bgp-evpn with mh-mode network
```

```
*A:PE-2>config>service>vpls>vxlan# source-vtep-security
MINOR: SVCMGR #7897 Cannot modify vxlan instance - not supported when vxlan instance is in use
 by bgp-evpn with mh-mode network
```

Local bias

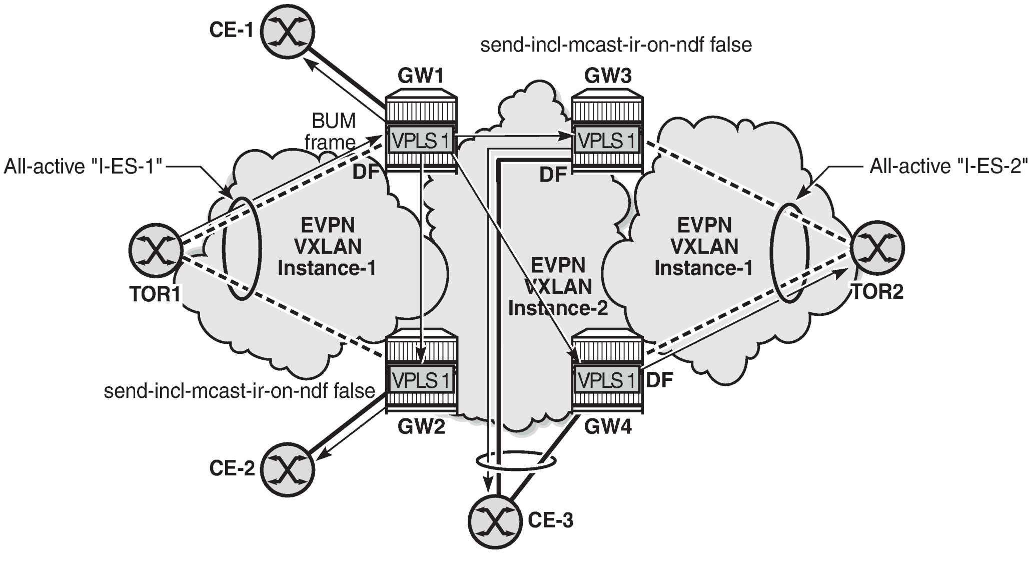
When EVPN-VXLAN is used in the instance network of a dual-instance VPLS service, local bias—as described in RFC 8365—is used for split horizon in all-active I-ESs. In VXLAN, there is no multicast label or multicast BMAC, so BUM traffic is identified by the MAC destination address. The modified forwarding rules for the I-ES-sourced BUM traffic for ingress PE and egress PE are as follows:

- ingress PE
  - The Non-Designated Forwarder (NDF) must discard BUM traffic, so one of the following two commands must be configured in VXLAN instance 1.
    - **no send-imet-ir-on-ndf**
    - **rx-discard-on-ndf bum**
  - BUM frames received on any SAP or I-ES VXLAN binding are flooded to:
    - local non-ES and single-active DF ES SAPs
    - local all-active ES SAPs (DF and NDF)
    - EVPN-VXLAN destinations (BUM frames received on an I-ES VXLAN binding follow SHG rules, so they can only be forwarded to EVPN-VXLAN destinations belonging to the other VXLAN instance.)
- egress PE
  - Look up source VTEP for BUM frames received on EVPN-VXLAN.

- If the source VTEP matches a PE with which the local PE shares an ES and a VXLAN service, then the local PE does not forward to the shared local ESs (this includes port, lag, and network-interconnect-VXLAN ESs).
- The local PE forwards to local ESs that are not shared but only when in DF state.

Figure 97: EVPN-VXLAN network interconnect VXLAN multi-homing and local bias shows the BUM forwarding with local bias procedures in multi-instance VPLS services.

Figure 97: EVPN-VXLAN network interconnect VXLAN multi-homing and local bias



In the example, GW1 and GW2 are configured with **no send-imet-ir-on-ndf**. TOR1 generates BUM traffic that will only reach DF GW1 and is forwarded as follows.

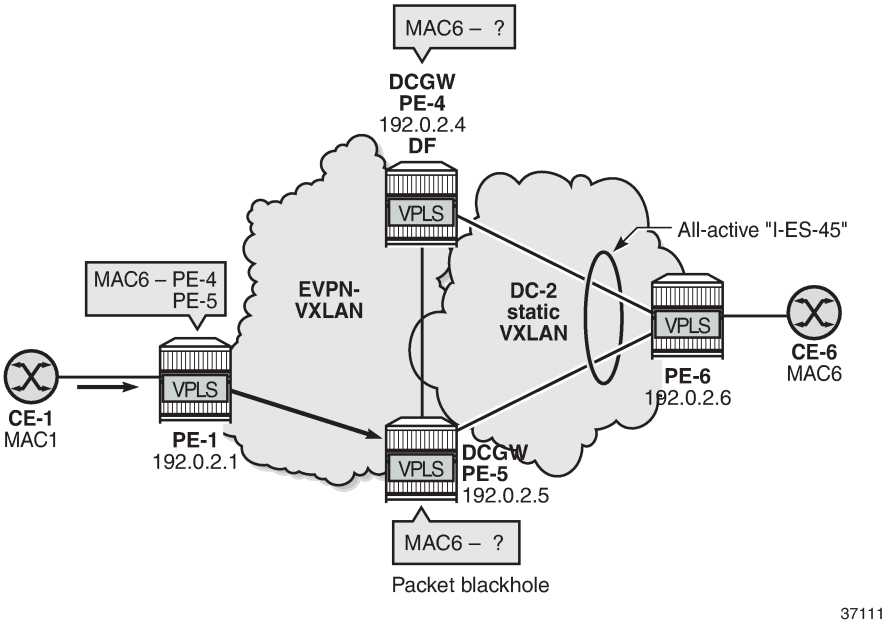
- Ingress PE GW1 forwards to CE-1 and EVPN-VXLAN destinations GW2, GW3, and GW4.
- Egress PE GW2 identifies the source VTEP as a PE with which I-ES-1 is shared, so it does not forward the BUM frames to the local I-ES. PE GW2 forwards only to the non-shared ES and local SAPs, in this case, to CE-2.
- Egress PE GW3 receives the BUM traffic with a source VTEP that does not match any PE with which GW3 shares an ES, so it forwards to all ESs that are DF, in this case, to CE-3.
- Egress PE GW4 receives the BUM traffic with a source VTEP that does not match any PE with which GW4 shares an ES, so it forwards to all ESs that are DF, in this case, to TOR2 through I-ES-2.

Local bias with static VXLAN on I-ES

When a static VXLAN instance coexists with an EVPN-VXLAN instance in the same VPLS service, traffic blackholes may occur when the static VXLAN instance is associated to an all-active I-ES. This is because, when multi-homing is used with an EVPN-VXLAN network instance, the NDF PE always discards unknown unicast traffic to the static VXLAN instance (this is not the case with EVPN-MPLS if the unknown traffic has a BUM label).

Figure 98: All-active I-ES NDF PE-5 drops unknown unicast traffic shows the packet blackhole for unknown unicast traffic at all-active I-ES NDF PE-5.

Figure 98: All-active I-ES NDF PE-5 drops unknown unicast traffic



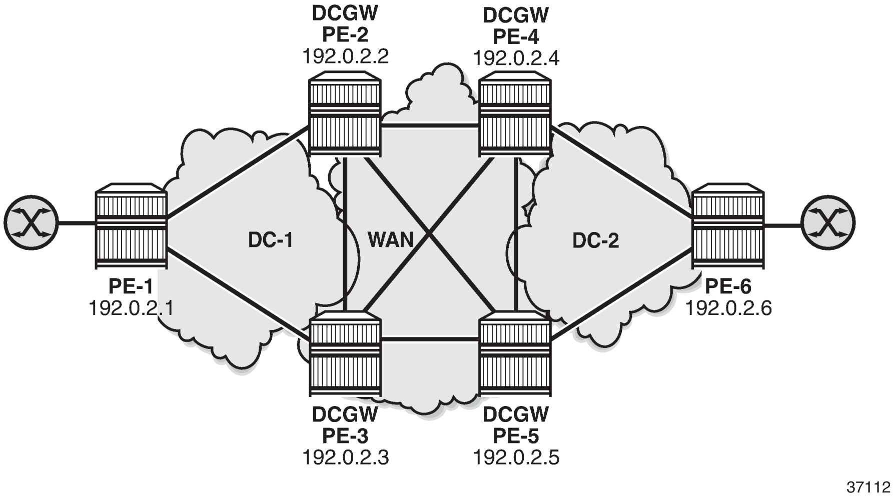
In the event that the remote PE-1 has learned the destination MAC address MAC6 via I-ES-45 EVPN destination, but the DCGWs PE-4 and PE-5 do not know MAC6, regular aliasing procedures allow that PE-1 sends unicast traffic with destination MAC6 to the NDF PE-5, which does not know MAC6 and drops all unknown unicast traffic, creating a blackhole for the flow.

When a static VXLAN instance coexists with an EVPN-VXLAN instance in the same VPLS service, Nokia recommends using a single-active I-ES or an anycast solution without I-ES instead of an all-active I-ES.

Configuration

Figure 99: Sample topology shows the sample topology with six SR OS nodes:

Figure 99: Sample topology



The initial configuration includes:

- Cards, MDAs, and ports
- Router interfaces
- IS-IS on all interfaces (level 1 in the DCs; level 2 in the WAN)

BGP is configured for the EVPN address family. PE-1 acts as Route Reflector (RR) in DC 1 and PE-6 as RR in DC 2; no RR is used in the WAN.

The BGP configuration on RR PE-1 in DC 1 is as follows. The BGP configuration on RR PE-6 in DC2 is similar.

```
# on PE-1:
configure
    router Base
        autonomous-system 64500
        bgp
            family evpn
            vpn-apply-import
            vpn-apply-export
            cluster 192.0.2.1
            rapid-withdrawal
            rapid-update evpn
            group "DC"
                type internal
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
```

On DCGWs PE-2 and PE-3, BGP is configured as follows. The policies are explained in the next section.

```
# on PE-2, PE-3:
configure
    router Base
        autonomous-system 64500
        bgp
```

```
            family evpn
            vpn-apply-import
            vpn-apply-export
            rapid-withdrawal
            rapid-update evpn
            group "DC"
                type internal
                import "drop SOO-DCGW-23"
                export "export DC routes and add SOO"
                neighbor 192.0.2.1
                exit
            exit
            group "WAN"
                type internal
                export "export WAN routes only"
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
```

On DCGWs PE-4 and PE-5, BGP is configured as follows. The policies are explained in the next section.

```
# on PE-4, PE-5:
configure
    router
        autonomous-system 64500
        bgp
            family evpn
            vpn-apply-import
            vpn-apply-export
            rapid-withdrawal
            rapid-update evpn
            group "DC"
                type internal
                import "drop SOO-DCGW-45"
                export "export DC routes and add SOO"
                neighbor 192.0.2.6
                exit
            exit
            group "WAN"
                type internal
                export "export WAN routes only"
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
```

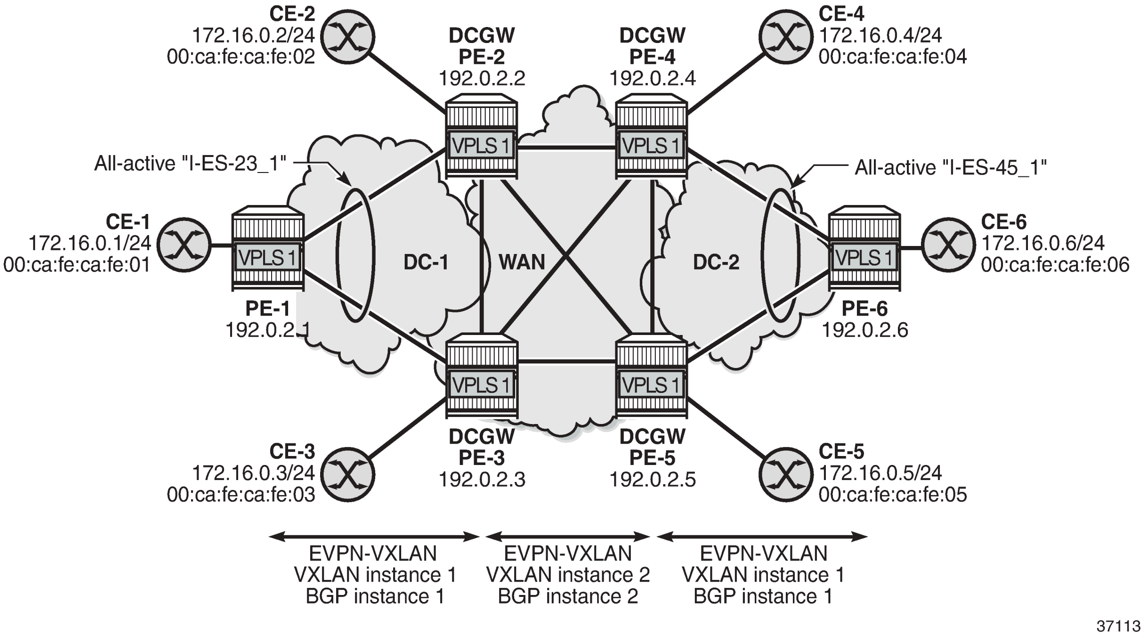
The following examples are configured:

- All-active multi-homing I-ESs in dual EVPN-VXLAN instance VPLS
- Single-active multi-homing I-ES when static VXLAN coexists with EVPN-VXLAN in the same VPLS

All-active multi-homing I-ESs in dual EVPN-VXLAN instance VPLS

Figure 100: All-active multi-homing for I-ESs shows the example topology with the service VPLS 1 on all nodes and two all-active I-ESs:

Figure 100: All-active multi-homing for I-ESs



On PE-1, VPLS 1 is configured as follows. The configuration on PE-6 is similar.

```
# on PE-1:
configure
    service
        system
            bgp-auto-rd-range 192.0.2.1 comm-val 1 to 1000     # on PE-6: 192.0.2.6
        exit
        vpls 1 name "VPLS 1" customer 1 create
            vxlan instance 1 vni 11 create
            exit
            bgp
                route-distinguisher auto-rd
                route-target export target:64500:11 import target:64500:11
            exit
            bgp-evpn
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    ecmp 2
                    no shutdown
                exit
            exit
            sap 1/2/1:1 create
            exit
            no shutdown
        exit
```

On DCGW PE-2, the following all-active multi-homing I-ES is configured for VXLAN instance 1 and service id 1. The configuration on DCGW PE-3 is similar, but the preference value is 150 instead of 100.

```
# on PE-2:
configure
    service
```

```
        system
            bgp-evpn
                ethernet-segment "I-ES-23_1" virtual create
                    esi 00:23:23:23:23:23:23:00:00:01
                    service-carving
                        mode manual
                        manual
                            preference create
                                value 100           # on PE-3: preference value 150
                            exit
                            evi 1
                        exit
                    exit
                    multi-homing all-active
                    network-interconnect-vxlan 1
                    service-id
                        service-range 1
                    exit
                    no shutdown
                exit
            exit
```

On DCGWs PE-4 and PE-5, the following I-ES is configured:

```
# on PE-4, PE-5:
configure
    service
        system
            bgp-evpn
                ethernet-segment "I-ES-45_1" virtual create
                    esi 00:45:45:45:45:45:45:00:00:01
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    network-interconnect-vxlan 1
                    service-id
                        service-range 1
                    exit
                    no shutdown
                exit
            exit
        exit
```

On DCGWs PE-2, PE-3, PE-4, and PE-5, VPLS 1 is configured as follows. The **rx-discard-on-ndf bum** command makes the NDF drop any BUM traffic in VXLAN instance 1. VXLAN instance 2 is configured with **mh-mode network and auto-disc-route-advertisement**.

```
# on PE-2, PE-3, PE-4, PE-5:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            vxlan instance 1 vni 11 create
                rx-discard-on-ndf bum
            exit
            vxlan instance 2 vni 12 create
            exit
            bgp
                route-distinguisher auto-rd
                route-target export target:64500:11 import target:64500:11
            exit
            bgp 2
```

```
                route-distinguisher auto-rd
                route-target export target:64500:12 import target:64500:12
            exit
            bgp-evpn
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    ecmp 2
                    default-route-tag 11
                    auto-disc-route-advertisement
                    no shutdown
                exit
                vxlan bgp 2 vxlan-instance 2
                    ecmp 2
                    default-route-tag 12
                    auto-disc-route-advertisement
                    mh-mode network
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/2/1:1 create
                no shutdown
            exit
            no shutdown
        exit
```

On PE-2 and PE-3, the following policies are configured.

- The import policy "drop SOO-DCGW-23" in group "DC" is used to drop all VXLAN instance 1 routes between PE-2 and PE-3.
- The export policy "export WAN routes only" in group "WAN" is applied to avoid sending VXLAN instance 1 routes to the WAN PEs.
- The export policy "export DC routes and add SOO" in group "DC" is used to tag VXLAN instance 1 routes with community "SOO-23".

```
# on PE-2, PE-3:
configure
    router Base
        policy-options
            begin
            community "SOO-23"
                members "origin:64500:23"
            exit
            policy-statement "drop SOO-DCGW-23"         # import in group "DC"
                entry 10
                    from
                        community "SOO-23"
                        family evpn
                    exit
                    action drop
                    exit
                exit
                default-action accept
                exit
            exit
            policy-statement "export WAN routes only"   # export in group "WAN"
                entry 10
                    from
                        tag 11
                        family evpn
                    exit
                    action drop
                    exit
                exit
                default-action accept
                exit
            exit
            policy-statement "export DC routes and add SOO"   # export in group "DC"
                entry 10
                    from
                        tag 11
                        family evpn
                    exit
                    action accept
                        community add "SOO-23"
                    exit
                exit
                default-action accept
                exit
            exit
            commit
```

On PE-4 and PE-5, the following policies are configured:

```
# on PE-4, PE-5:
configure
    router Base
        policy-options
            begin
            community "SOO-45"
                members "origin:64500:45"
            exit
            policy-statement "drop SOO-DCGW-45"         # import in group "DC"
                entry 10
                    from
                        community "SOO-45"
                        family evpn
                    exit
                    action drop
                    exit
                exit
                default-action accept
                exit
            exit
            policy-statement "export WAN routes only"   # export in group "WAN"
                entry 10
                    from
                        tag 11
                        family evpn
                    exit
                    action drop
                    exit
                exit
                default-action accept
                exit
            exit
            policy-statement "export DC routes and add SOO"   # export in group "DC"
                entry 10
                    from
                        tag 11
                        family evpn
                    exit
                    action accept
                        community add "SOO-45"
                    exit
                exit
                default-action accept
                exit
            exit
            commit
```

For VPLS 1, PE-2 is DF and PE-3 is NDF in the I-ES "I-ES-23_1":

```
*A:PE-2# show service id 1 ethernet-segment
No sap entries
No sdp entries

===============================================================================
VXLAN Ethernet-Segment Information
===============================================================================
VXLAN Instance        Eth-Seg                         Status
-------------------------------------------------------------------------------
1                     I-ES-23_1                       DF
===============================================================================
```

```
*A:PE-3# show service id 1 ethernet-segment
No sap entries
No sdp entries

===============================================================================
VXLAN Ethernet-Segment Information
===============================================================================
VXLAN Instance        Eth-Seg                         Status
-------------------------------------------------------------------------------
1                     I-ES-23_1                       NDF
===============================================================================
```

PE-4 is NDF and PE-5 is DF in the I-ES "I-ES-45_1":

```
*A:PE-4# show service vxlan-instance-using ethernet-segment

===============================================================================
VXLAN Ethernet-Segment Information
===============================================================================
SvcId       VXLAN Instance      ES Name                         Status
-------------------------------------------------------------------------------
1           1                   I-ES-45_1                       NDF
===============================================================================
```

```
*A:PE-5# show service vxlan-instance-using ethernet-segment

===============================================================================
VXLAN Ethernet-Segment Information
===============================================================================
SvcId       VXLAN Instance      ES Name                         Status
-------------------------------------------------------------------------------
1           1                   I-ES-45_1                       DF
===============================================================================
```

On leaf PE-1, the VXLAN destinations in VXLAN instance 1 are the following:

```
*A:PE-1# show service id 1 vxlan destinations
```

```
===============================================================================
Egress VTEP, VNI
===============================================================================
Instance    VTEP Address                            Egress VNI  EvpnStatic Num
 Mcast       Oper State                              L2 PBR     SupBcasDom MACs
-------------------------------------------------------------------------------
1           192.0.2.2                               11          evpn       0
 BUM         Up                                      No          No
1           192.0.2.3                               11          evpn       0
 BUM         Up                                      No          No
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 2
-------------------------------------------------------------------------------
===============================================================================
===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
1         00:23:23:23:23:23:23:00:00:01   5             09/07/2021 14:35:57
-------------------------------------------------------------------------------
Number of entries: 1
-------------------------------------------------------------------------------
===============================================================================
```

On DCGW PE-2, the VXLAN destinations in VXLAN instances 1 and 2 are the following:

```
*A:PE-2# show service id 1 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
Instance    VTEP Address                            Egress VNI  EvpnStatic Num
 Mcast       Oper State                              L2 PBR     SupBcasDom MACs
-------------------------------------------------------------------------------
1           192.0.2.1                               11          evpn       1
 BUM         Up                                      No          No
2           192.0.2.3                               12          evpn       1
 BUM         Up                                      No          No
2           192.0.2.4                               12          evpn       1
 BUM         Up                                      No          No
2           192.0.2.5                               12          evpn       1
 BUM         Up                                      No          No
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 4
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
2         00:45:45:45:45:45:45:00:00:01   1             09/07/2021 14:36:31
-------------------------------------------------------------------------------
Number of entries: 1
-------------------------------------------------------------------------------
===============================================================================
```

ECMP 2 is configured, so aliasing is used. PE-1 can reach the I-ES "I-ES-23_1" in VXLAN instance 1 via PE-2 and PE-3:

```
*A:PE-1# show service id 1 vxlan esi 00:23:23:23:23:23:23:00:00:01

===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
1         00:23:23:23:23:23:23:00:00:01   5             09/07/2021 14:35:57
-------------------------------------------------------------------------------
Number of entries: 1
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-VXLAN Dest TEP Info
===============================================================================
Instance  TEP Address                   Egr VNI             Last Change
-------------------------------------------------------------------------------
1         192.0.2.2                     11                  09/07/2021 14:35:47
1         192.0.2.3                     11                  09/07/2021 14:35:57
-------------------------------------------------------------------------------
Number of entries : 2
-------------------------------------------------------------------------------
===============================================================================
```

In a similar way, PE-4 can reach the I-ES "I-ES-23_1" via PE-2 and PE-3 in VXLAN instance 2:

```
*A:PE-4# show service id 1 vxlan esi 00:23:23:23:23:23:23:00:00:01

===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
2         00:23:23:23:23:23:23:00:00:01   1             09/07/2021 14:36:10
-------------------------------------------------------------------------------
Number of entries: 1
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-VXLAN Dest TEP Info
===============================================================================
Instance  TEP Address                   Egr VNI             Last Change
-------------------------------------------------------------------------------
2         192.0.2.2                     12                  09/07/2021 14:36:10
2         192.0.2.3                     12                  09/07/2021 14:36:10
-------------------------------------------------------------------------------
Number of entries : 2
-------------------------------------------------------------------------------
===============================================================================
```

The following command on PE-2 shows the ES information for "I-ES-23_1": DF status, DF candidate list, VXLAN instance service range, and so on:

```
*A:PE-2# show service system bgp-evpn ethernet-segment name "I-ES-23_1" all

===============================================================================
Service Ethernet Segment
```

```
===============================================================================
Name                    : I-ES-23_1
Eth Seg Type            : Virtual
Admin State             : Enabled            Oper State         : Up
ESI                     : 00:23:23:23:23:23:23:00:00:01
Oper ESI                : 00:23:23:23:23:23:23:00:00:01
Auto-ESI Type           : None
AC DF Capability        : Include
Multi-homing            : allActive          Oper Multi-homing  : allActive
ES SHG Label            : 524287
Source BMAC LSB         : None
VXLAN Instance Id       : 1
ES Activation Timer     : 3 secs (default)
Oper Group              : (Not Specified)
Svc Carving             : manual             Oper Svc Carving   : manual
Cfg Range Type          : lowest-pref

-------------------------------------------------------------------------------
DF Pref Election Information
-------------------------------------------------------------------------------
Preference     Preference     Last Admin Change        Oper Pref      Do No
Mode           Value                                   Value          Preempt
-------------------------------------------------------------------------------
revertive      100            09/07/2021 14:35:47      100            Disabled
-------------------------------------------------------------------------------

-------------------------------------------------------------------------------
EVI Ranges
-------------------------------------------------------------------------------
From                                    To
-------------------------------------------------------------------------------
1                                       1
-------------------------------------------------------------------------------
ISID Ranges: <none>
===============================================================================

===============================================================================
EVI Information
===============================================================================
EVI                 SvcId               Actv Timer Rem      DF
-------------------------------------------------------------------------------
1                   1                   0                   yes
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================

-------------------------------------------------------------------------------
DF Candidate list
-------------------------------------------------------------------------------
EVI                                     DF Address
-------------------------------------------------------------------------------
1                                       192.0.2.2
1                                       192.0.2.3
-------------------------------------------------------------------------------
Number of entries: 2
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------

===============================================================================

---snip---

===============================================================================
Vxlan Instance Service Ranges
```

```
===============================================================================
Svc Range Start          Svc Range End            Last Changed
-------------------------------------------------------------------------------
1                        1                        09/07/2021 14:35:47
-------------------------------------------------------------------------------
Number of Entries: 1
===============================================================================
```

When traffic is sent between CE-1 and CE-6, the EVPN-MAC routes are sent with I-ESI. The FDB for VPLS 1 on PE-1 shows I-ESI 00:23:23:23:23:23:23:00:00:01 of "I-ES-23_1" as source identifier for MAC address 00:ca:fe:ca:fe:06 of CE-6:

```
*A:PE-1# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:ca:fe:ca:fe:01 sap:1/2/1:1             L/127    09/07/21 14:48:43
1          00:ca:fe:ca:fe:06 eES:                    Evpn     09/07/21 14:48:43
                             00:23:23:23:23:23:23:00:00:01
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

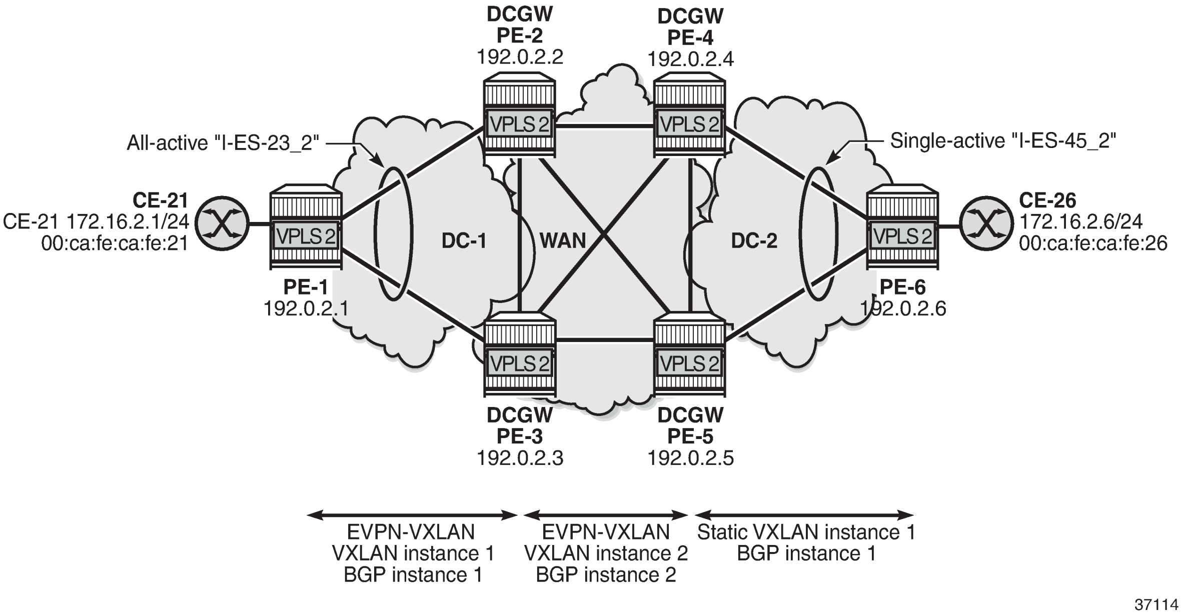
On PE-2, the FDB for VPLS 1 shows I-ESI 00:45:45:45:45:45:45:00:00:01 of "I-ES-45_1" as source identifier for MAC address 00:ca:fe:ca:fe:06 of CE-6:

```
*A:PE-2# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:ca:fe:ca:fe:01 vxlan-1:                Evpn     09/07/21 14:48:43
                             192.0.2.1:11
1          00:ca:fe:ca:fe:06 eES:                    Evpn     09/07/21 14:48:43
                             00:45:45:45:45:45:45:00:00:01
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

Single-active multi-homing I-ES when static VXLAN coexists with EVPN-VXLAN in the same VPLS

Figure 101: I-ES with EVPN-VXLAN in DC 1 and static VXLAN in DC2 shows the sample topology for VPLS 2 with static VXLAN in DC 2 and the single-active "I-ES-45_2" on PE-4 and PE-5.

Figure 101: I-ES with EVPN-VXLAN in DC 1 and static VXLAN in DC2



The configuration for VPLS 2 on PE-1, PE-2, and PE-3 is similar to the configuration for VPLS 1, so only the configuration on PE-4, PE-5, and PE-6 is shown.

On PE-6, VPLS 2 is configured with static VXLAN using non-anycast VTEP addresses:

```
# on PE-6:
configure
    service
        system
            bgp-auto-rd-range 192.0.2.6 comm-val 1 to 1000
        exit
        vpls 2 name "VPLS 2" customer 1 create
            vxlan instance 1 vni 21 create
                egr-vtep 192.0.2.4 create
                exit
                egr-vtep 192.0.2.5 create
                exit
            exit
            sap 1/2/1:2 create
                no shutdown
            exit
            no shutdown
        exit
```

To avoid blackholes, the I-ES between DCGWs PE-4 and PE-5 must not be all-active.

On PE-4 and PE-5, the single-active I-ES "I-ES-45_2" is configured as follows:

```
# on PE-4, PE-5:
configure
    service
        system
            bgp-evpn
                ethernet-segment "I-ES-45_2" virtual create
```

```
                    esi 00:45:45:45:45:45:45:00:00:02
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    network-interconnect-vxlan 1
                    service-id
                        service-range 2
                    exit
                    no shutdown
                exit
            exit
```

On PE-4 and PE-5, VPLS 2 is configured as follows:

```
# on PE-4, PE-5:
configure
    service
       vpls 2 name "VPLS 2" customer 1 create
            vxlan instance 1 vni 21 create
                egr-vtep 192.0.2.6 create
                exit
            exit
            vxlan instance 2 vni 22 create
            exit
            bgp 2
                route-distinguisher auto-rd
                route-target export target:64500:22 import target:64500:22
            exit
            bgp-evpn
                evi 2
                vxlan bgp 2 vxlan-instance 2
                    ecmp 2
                    default-route-tag 22
                    auto-disc-route-advertisement
                    mh-mode network
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/2/1:2 create        # optional SAP toward local CE
                no shutdown
            exit
            no shutdown
        exit
```

The policies on all DCGWs must be modified with tag 21 for VXLAN instance 1 in VPLS 2, as follows:

```
# on PE-2, PE-3:
configure
    router Base
        policy-options
            begin
            policy-statement "export WAN routes only"
                entry 20
                    from
                        tag 21
                        family evpn
                    exit
                    action drop
                    exit
                exit
                default-action accept
                exit
            exit
            policy-statement "export DC routes and add SOO"
                entry 20
                    from
                        tag 21
                        family evpn
                    exit
                    action accept
                        community add "SOO-23"
                    exit
                exit
                default-action accept
                exit
            exit
            commit
```

DCGW PE-5 is NDF for "I-ES-45_2":

```
*A:PE-5# show service id 2 ethernet-segment
No sap entries
No sdp entries

===============================================================================
VXLAN Ethernet-Segment Information
===============================================================================
VXLAN Instance        Eth-Seg                         Status
-------------------------------------------------------------------------------
1                     I-ES-45_2                       NDF
===============================================================================
```

On PE-5, the status of VXLAN instance 1 in VPLS 2 is mhStandby, as follows:

```
*A:PE-5# show service id 2 vxlan
===============================================================================
VPLS VXLAN
===============================================================================
Vxlan Src Vtep IP: N/A

===============================================================================
Vxlan Instance
===============================================================================
VXLAN Instance                VNI         AR          Oper-flags   VTEP
                                                                   security
-------------------------------------------------------------------------------
1                             21          none        mhStandby    disabled
2                             22          none        none         disabled
-------------------------------------------------------------------------------
Number of Entries : 2
-------------------------------------------------------------------------------
===============================================================================
```

The VXLAN destinations in VPLS 2 on PE-5 are the following:

```
*A:PE-5# show service id 2 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
Instance    VTEP Address                                Egress VNI  EvpnStatic Num
```

```
 Mcast       Oper State                              L2 PBR     SupBcasDom MACs
-------------------------------------------------------------------------------
1           192.0.2.6                               21          static     0
 -           Up                                      No          No
2           192.0.2.2                               22          evpn       1
 BUM         Up                                      No          No
2           192.0.2.3                               22          evpn       1
 BUM         Up                                      No          No
2           192.0.2.4                               22          evpn       1
 BUM         Up                                      No          No
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 4
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
2         00:23:23:23:23:23:23:00:00:02   1             09/07/2021 14:53:05
-------------------------------------------------------------------------------
Number of entries: 1
-------------------------------------------------------------------------------
===============================================================================
```

Note:
An anycast solution without I-ES can also be configured when an EVPN-VXLAN coexists with a static VXLAN.

Conclusion

Service providers can use I-ESs for better bandwidth utilization and redundancy in large DCs. EVPN all-active multi-homing I-ESs can be used in dual EVPN-VXLAN instance VPLS services. However, when a static VXLAN instance coexists with EVPN-VXLAN in the same VPLS, a single-active multi-homing I-ES (or an anycast solution without I-ES) is required to avoid blackholes.

EVPN IP-VRF-to-IP-VRF Models

This chapter provides information about EVPN IP-VRF-to-IP-VRF models.

Topics in this chapter include:

Applicability

This chapter was initially written based on SR OS Release 16.0.R3, but the CLI in the current edition corresponds to SR OS Release 23.7.R2. SR OS supports the three EVPN IP-VRF-to-IP-VRF models described in *draft-ietf-bess-evpn-prefix-advertisement*.

Overview

EVPN is considered the standard for Data Centers (DCs) and DC Interconnect (DCI) for layer 2 and layer 3 services. *Draft-ietf-bess-evpn-prefix-advertisement* describes the following three IP-VRF-to-IP-VRF models:

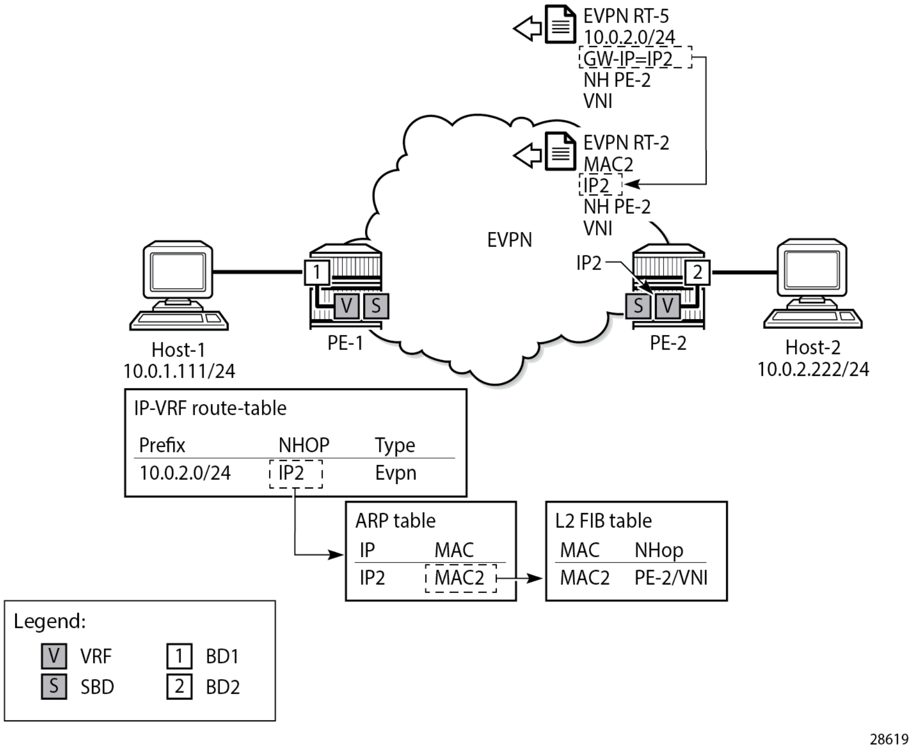
- Interface-less model (mandatory)
- Interface-ful model with Supplementary Broadcast Domain (SBD) Interworking Routing and Bridging (IRB) (mandatory)
- Interface-ful model with unnumbered SBD IRB (optional)

In standard terminology, SBD is the Broadcast Domain (BD) that joins two IP-VRFs. In SR OS, the SBD is a "backhaul" R-VPLS service that connects two PEs attached to VPRNs of the same VPN. For IP prefix advertisement in the SBD, IP route advertisement needs to be enabled in the BGP-EVPN context, whereas MAC advertisement is enabled by default. BGP-EVPN IP prefix route type 5 (RT-5) updates are used in all models; MAC/IP routes (RT-2) are used in the interface-ful models only. In the interface-less model, MAC advertisement must be disabled.

Figure 102: Interface-ful SBD IRB and Figure 103: Interface-ful unnumbered SBD IRB show the two interface-ful IP-VRF-to-IP-VRF models: SBD IRB and unnumbered SBD IRB. Both interface-ful SBD IRB models require BGP-EVPN IP prefix routes (RT-5) with recursive lookup to MAC/IP routes (RT-2). Host 1 is located in broadcast domain 1 (BD1 corresponds to an R-VPLS) linked to the VRF in PE-1 and host 2 is located in BD2 linked to the VRF in PE-2. The VRFs correspond to VPRNs that are linked to an SBD, which is a backhaul R-VPLS.

The following examples are based on EVPN-VXLAN, but IP-VRF-to-IP-VRF also works for EVPN-MPLS. Instead of the VNI, the MPLS label is then included in the RT-5 and RT-2 updates.

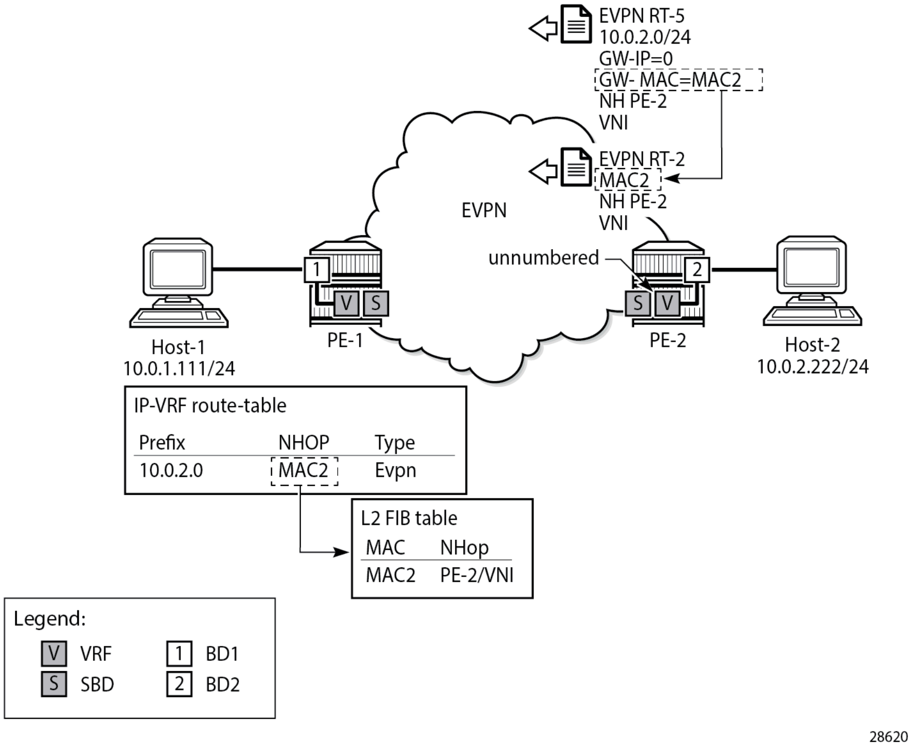
Figure 102: Interface-ful SBD IRB



The interface-ful SBD IRB model requires an IP address on the VPRN interface for the SBD (IP2 on PE-2); no EVPN tunnel can be used. Both PEs will send BGP-EVPN RT-5 (IP prefix) and BGP-EVPN RT-2 (MAC/IP) updates. PE-2 sends an RT-5 update for IP prefix 10.0.2.0/24 with GW IP address IP2 and an RT-2 update for GW IP address IP2 with MAC2 and next-hop PE-2. On PE-1, the prefix 10.0.2.0/24 appears in the VRF route table as an EVPN route with next-hop GW IP2. The ARP table for the VRF contains the corresponding MAC address MAC2 for the GW IP address IP2. The FDB of the SBD includes an EVPN entry for GW MAC address MAC2 with next-hop PE-2.

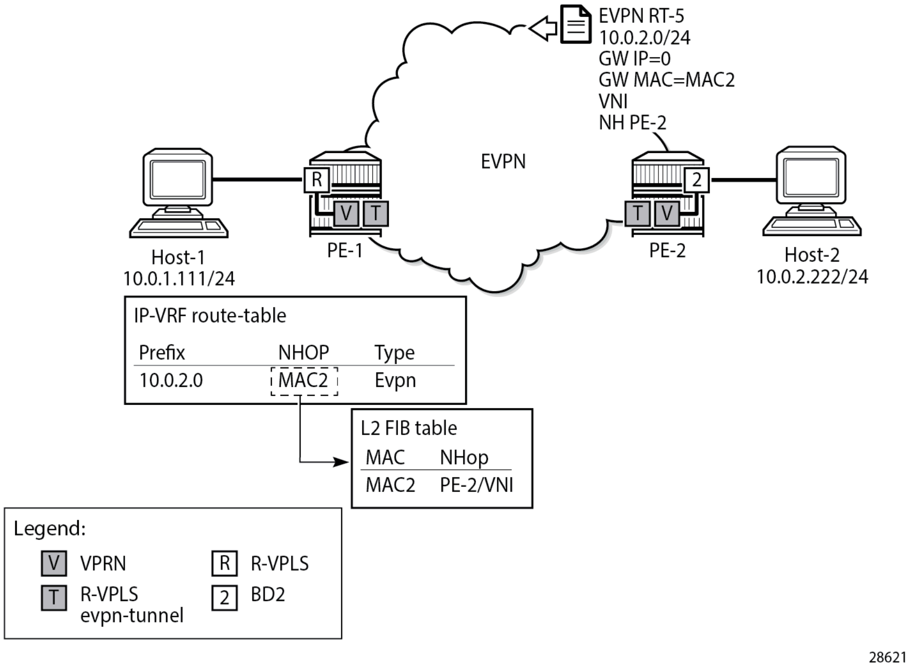
When the VPRN is configured toward the SBD with an EVPN tunnel rather than a numbered IP interface, the RT-5 update will contain the GW MAC address MAC2 instead of the GW IP address IP2. Figure 103: Interface-ful unnumbered SBD IRB shows that PE-2 sends an RT-5 update for IP prefix 10.0.2.0/24 with GW MAC address MAC2 and an RT-2 update for GW MAC address MAC2 with next-hop PE-2. Again, a recursive lookup is done.

Figure 103: Interface-ful unnumbered SBD IRB



Finally, in the interface-less IP-VRF-to-IP-VRF model, MAC advertisement is disabled in the BGP-EVPN context of the backhaul R-VPLS. BGP-EVPN RT-5 updates will contain the GW MAC address, and no RT-2 updates will be sent; therefore, the number of BGP-EVPN updates is reduced and no recursive lookup is done on PE-1. PE-1 adds an entry in its FDB based on an RT-5 route instead of an RT-2 route from PE-2. Figure 104: Interface-less IP-VRF-to-IP-VRF model shows the interface-less IP-VRF-to-IP-VRF model where PE-2 sends an RT-5 update with GW MAC address MAC2.

Figure 104: Interface-less IP-VRF-to-IP-VRF model



Note:
Other vendors do not use a service context as the R-VPLS EVPN tunnel shown in Figure 104: Interface-less IP-VRF-to-IP-VRF model, and they configure the route targets used for the RT-5 updates in the VPRN (or VRF) instances. When interoperating with those vendors, ensure that the R-VPLS route targets match the route targets in the VRF of the third-party router.

The standard specification *draft-ietf-bess-evpn-ip-prefix* supports two variants of the interface-less model that are not interoperable with each other:

- EVPN interface-less (EVPN IFL) for Ethernet Network Virtualization Overlay (NVO) tunnels

  Ethernet NVO indicates that the EVPN packets contain an inner Ethernet header. The ingress PE uses the received router's MAC extended community address in the IP prefix route as the inner destination MAC address for the EVPN packets sent to the prefix. This corresponds to the scenario described in Figure 104: Interface-less IP-VRF-to-IP-VRF model.

- EVPN IFL for IP NVO tunnels

  IP NVO indicates that the EVPN packets contain an inner IP packet, but no Ethernet header. This is similar to the IP-VPN packets exchanged between PEs. In this scenario, the IP prefix route does not contain any GW (IP or MAC) address. The IP packets are directly encapsulated with an EVPN service label and the transport labels. This model is described further in Interface-less model in EVPN-MPLS with IP encapsulation.

EVPN MAC selection criteria

In the EVPN IFL for Ethernet NVO scenario, the MAC address entry in the R-VPLS FDB that is required to forward packets to the remote PE is obtained from an internal MAC/IP route. This internal route is obtained from the router MAC extended community in the BGP-EVPN RT-5 update. In case the same MAC address is received in multiple ways, the following MAC selection criteria apply. Beginning with criterion (1), the MAC is selected if the criterion is met, or the next criterion is applied. As indicated in (8), a MAC received from an RT-2 has higher priority than a MAC populated by the router MAC extended community in an RT-5 update.

1. Conditional static MAC addresses (locally protected MAC addresses)
2. Auto-learned protected MAC addresses (locally learned MAC addresses on SAPs or SDP-bindings due to the configuration of **auto-learn-mac-protect**)
3. EVPN ES PBR MAC addresses
4. EVPN static MAC addresses (remotely protected MAC addresses)
5. Data plane learned MAC addresses (regular learning on SAPs or SDP-bindings)
6. EVPN MAC routes with a higher sequence number
7. EVPN E-Tree root MAC addresses
8. EVPN non-RT-5 MAC addresses (this tie-breaking rule is only applied if the selection algorithm is comparing received MAC routes (RT-2) and internal MAC routes derived from the MAC addresses in IP-prefix routes, such as RT-5 MACs)
9. Lowest IP address for the next-hop of the EVPN NLRI
10. Lowest Ethernet tag (that will be zero for MPLS and might be different from zero for VXLAN)
11. Lowest route distinguisher
12. Lowest BGP instance (this tie-breaking rule is only applied if the preceding rules fail to select a unique MAC address and the service has two BGP instances of the same encapsulation)

EVPN IP-VRF-to-IP-VRF model comparison

Each model has its advantages. Table 6: EVPN IP-VRF-to-IP-VRF model comparison compares the three IP-VRF-to-IP-VRF models.

Table 6: EVPN IP-VRF-to-IP-VRF model comparison

Advantage	Model 1 Interface-less	Model 2 Interface-ful SBD IRB	Model 3 Interface-ful unnumbered SBD IRB
Reduced number of EVPN routes	Yes	No	No
Ease of provisioning (no IP address on core IRB)	Yes	No	Yes
Mass withdrawal due to recursive resolution	No	Yes	Yes

Configuration

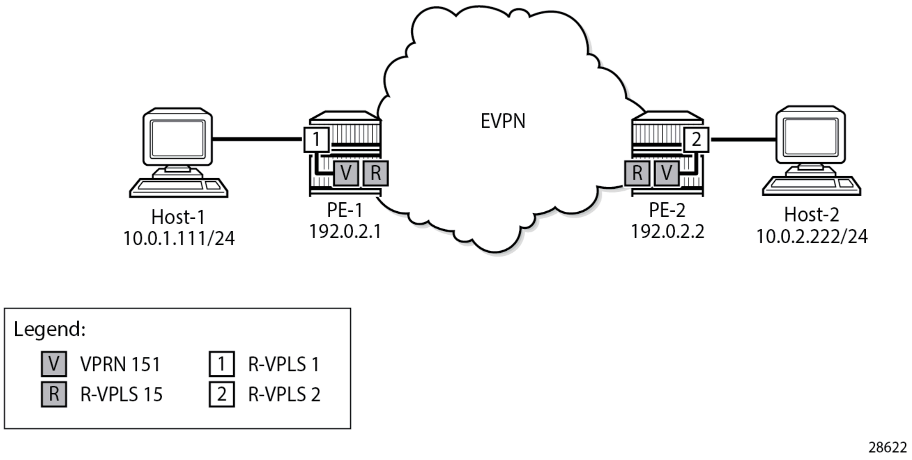
The following use cases are documented in this chapter:

- IP-VRF-to-IP-VRF Models in EVPN-VXLAN
  - Interface-ful model with SBD IRB in EVPN-VXLAN
  - Interface-ful model with unnumbered SBD IRB in EVPN-VXLAN
  - Interoperable interface-less model in EVPN-VXLAN
- IP-VRF-to-IP-VRF Models in EVPN-MPLS
  - Interface-ful model with SBD IRB in EVPN-MPLS
  - Interface-ful model with unnumbered SBD IRB in EVPN-MPLS
  - Interface-less model in EVPN-MPLS
    - Interoperable interface-less model with Ethernet encapsulation
    - Interface-less model with IP encapsulation for MPLS tunnels

IP-VRF-to-IP-VRF model in EVPN-VXLAN

Figure 105: Example topology with services - EVPN-VXLAN shows the example topology with two PEs. Hosts 1 and 2—emulated through VPRNs—are attached to R-VPLS 1 and 2 respectively.

Figure 105: Example topology with services - EVPN-VXLAN



The initial configuration on the PEs includes the following:

- Cards, MDAs, ports
- Router interfaces
- IS-IS (alternatively, OSPF can be used)
- BGP for address family EVPN

On PE-1, the BGP configuration is as follows. The BGP configuration on PE-2 is similar.

```
# on PE-1:
configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            rapid-withdrawal
            rapid-update evpn
            group "dc"
                family evpn
                type internal
                neighbor 192.0.2.2
                exit
            exit
        exit
```

Interface-ful model with SBD IRB in EVPN-VXLAN

The service configuration on PE-1 includes the SBD R-VPLS "sbd-15", VPRN "ip-vrf-151", and R-VPLS "bd-1". The service configuration on PE-2 is similar, but R-VPLS "bd-2" is configured instead of R-VPLS "bd-1".

On PE-1, SBD R-VPLS "sbd-15" is configured with VNI 15, as follows. MAC advertisement is enabled by default, but IP route advertisement must be enabled explicitly. Only one BGP instance and one VXLAN instance are configured.

```
# on PE-1:
configure
    service
        vpls 15 name "sbd-15" customer 1 create
            description "backhaul R-VPLS 15"
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 15 create
            exit
            bgp
            exit
            bgp-evpn
                ip-route-advertisement
                evi 15
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            no shutdown
            stp
                shutdown
            exit
        exit
```

VPRN "ip-vrf-151" has two interfaces: one toward the SBD R-VPLS "sbd-15" and one toward BD R-VPLS "bd-1". The interface toward the SBD has GW IP address 172.16.151.1/24 and MAC address 00:00:00:01:51:01. The interface toward R-VPLS 1 has IP address 10.0.1.1/24 and MAC address 00:00:00:1e:01:01. VRRP is configured in passive mode, so PE-1 uses the backup IP address as an anycast gateway. The backup IP address is 10.0.1.254 and the auto-derived virtual MAC address is 00:00:5e:00:00:01 for VRID 1. On PE-1, VPRN "ip-vrf-151" is configured as follows:

```
# on PE-1:
configure
    service
        vprn 151 name "ip-vrf-151" customer 1 create
            ecmp 2
            interface "int-sbd-15" create
                address 172.16.151.1/24
                mac 00:00:00:01:51:01
                vpls "sbd-15"
                exit
            exit
            interface "int-bd-1" create
                address 10.0.1.1/24
                mac 00:00:00:1e:01:01
                vrrp 1 passive
                    backup 10.0.1.254
                    ping-reply
                    traceroute-reply
                exit
                vpls "bd-1"
                exit
            exit
            no shutdown
        exit
```

On PE-1, R-VPLS "bd-1" is configured as follows. Host 1 is attached to the SAP.

```
# on PE-1:
configure
    service
        vpls 1 name "bd-1" customer 1 create
            description "R-VPLS 1 - BD 1"
            allow-ip-int-bind
            exit
            stp
                shutdown
            exit
            sap pxc-10.a:1 create
                no shutdown
            exit
            no shutdown
        exit
```

In this example, host 1 is simulated by VPRN "host1", as follows. The default route has next-hop 10.0.1.254, which is the VRRP backup address in VPRN "ip-vrf-151".

```
# on PE-1:
configure
    service
        vprn 11 name "host1" customer 1 create
            description "Host-1 attached to R-VPLS 1"
            interface "local" create
                address 10.0.1.111/24
                mac 00:00:00:10:11:01
                sap pxc-10.b:1 create
                exit
            exit
            static-route-entry 0.0.0.0/0
```

```
                next-hop 10.0.1.254
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

The service configuration on PE-2 is similar, with R-VPLS "bd-2" instead of R-VPLS "bd-1" and VPRN "host2" instead of VPRN "host1". The GW IP address on PE-2 is 172.16.151.2/24, interface "int-bd-2" in VPRN "ip-vrf-151" has IP address 10.0.2.2/24, and host "host2" has IP address 10.0.2.222/24.

PE-1 receives a BGP-EVPN RT-5 update from PE-2 for IP prefix 10.0.2.0/24, as follows. The GW address is IP address 172.16.151.2 and the next-hop is PE-2.

```
*A:PE-1# show router bgp routes evpn ip-prefix
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
Flag  Route Dist.         Prefix
      Tag                 Gw Address
                          NextHop
                          Label
                          ESI
-------------------------------------------------------------------------------
u*>i  192.0.2.2:15        10.0.2.0/24
      0                   172.16.151.2
                          192.0.2.2
                          VNI 15
                          ESI-0

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

PE-1 receives the following BGP-EVPN MAC update for MAC address 00:00:00:01:51:02, which corresponds to GW IP 172.16.151.2:

```
*A:PE-1# show router bgp routes evpn mac
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
```

```
u*>i  192.0.2.2:15        00:00:00:01:51:02 ESI-0
      0                   Static            VNI 15
                          172.16.151.2
                          192.0.2.2

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

The following traceroute on PE-1 from host 1 to host 2 shows that the first hop is 10.0.1.1 (interface "int-bd-1" in VPRN "ip-vrf-151" on PE-1), the second hop is the IP GW address 172.16.151.2 (interface "int-sbd-15" in VPRN "ip-vrf-151" on PE-2), and the third hop is host 2 with IP address 10.0.2.222:

```
*A:PE-1# traceroute router-instance "host1" 10.0.2.222 source 10.0.1.111
traceroute to 10.0.2.222 from 10.0.1.111, 30 hops max, 40 byte packets
  1  10.0.1.1 (10.0.1.1)    4.96 ms  2.59 ms  2.07 ms
  2  172.16.151.2 (172.16.151.2)    4.04 ms  3.88 ms  5.15 ms
  3  10.0.2.222 (10.0.2.222)    11.5 ms  4.06 ms  4.73 ms
```

On PE-1, the following route table for VPRN "ip-vrf-151" contains a EVPN interface-ful (EVPN IFF) route for IP prefix 10.0.2.0/24 with next-hop 172.16.151.2 and preference 169 (whereas BGP-VPN routes for IP-VPN have a preference of 170):

```
*A:PE-1# show router service-name "ip-vrf-151" route-table

===============================================================================
Route Table (Service: 151)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.1.0/24                                   Local   Local     00h02m35s  0
       int-bd-1                                                     0
10.0.2.0/24                                   Remote  EVPN-IFF  00h02m28s  169
       172.16.151.2                                                 0
172.16.151.0/24                               Local   Local     00h02m35s  0
       int-sbd-15                                                   0
-------------------------------------------------------------------------------
No. of Routes: 3
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

On PE-1, the following ARP table of VPRN "ip-vrf-151" contains an EVPN entry for GW IP address 172.16.151.2:

```
*A:PE-1# show service id "ip-vrf-151" arp

===============================================================================
ARP Table
===============================================================================
IP Address      MAC Address       Type    Expiry    Interface         SAP
-------------------------------------------------------------------------------
172.16.151.1    00:00:00:01:51:01 Other   00h00m00s int-sbd-15        rvpls
172.16.151.2    00:00:00:01:51:02 EVPN    00h00m00s int-sbd-15        rvpls
10.0.1.1        00:00:00:1e:01:01 Other   00h00m00s int-bd-1          rvpls
10.0.1.111      00:00:00:10:11:01 Dynamic 03h59m20s int-bd-1          rvpls
10.0.1.254      00:00:5e:00:01:01 Other   00h00m00s int-bd-1          rvpls
```

```
===============================================================================
```

The following FDB on PE-1 shows a static and protected EVPN entry for MAC address 00:00:00:01:51:02:

```
*A:PE-1# show service id "sbd-15" fdb detail

===============================================================================
Forwarding Database, Service 15
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
15         00:00:00:01:51:01 cpm                     Intf     10/25/23 07:35:43
15         00:00:00:01:51:02 vxlan-1:                EvpnS:P  10/25/23 07:35:50
                             192.0.2.2:15
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

Interface-ful model with unnumbered SBD IRB in EVPN-VXLAN

On both PEs, the GW IP addresses 172.16.151.x/24 are removed from interface "int-sbd-15" in VPRN "ip-vrf-151" and an EVPN tunnel is configured instead. The changes in the configuration of VPRN "ip-vrf-151" on PE-1 are the following:

```
# on PE-1:
configure
    service
        vprn "ip-vrf-151"
            interface "int-sbd-15"
                no address 172.16.151.1/24
                vpls "sbd-15"
                    evpn-tunnel
                exit
            exit
```

Similarly, the following is configured in VPRN "ip-vrf-151" on PE-2:

```
*A:PE-2#
configure
    service
        vprn "ip-vrf-151"
            interface "int-sbd-15"
                no address 172.16.151.2/24
                vpls "sbd-15"
                    evpn-tunnel
                exit
            exit
```

The configuration of VPRN "ip-vrf-151" on PE-1 is as follows:

```
*A:PE-1>config>service>vprn# info
----------------------------------------------
            ecmp 2
            interface "int-sbd-15" create
                mac 00:00:00:01:51:01
                vpls "sbd-15"
                    evpn-tunnel
                exit
            exit
            interface "int-bd-1" create
                address 10.0.1.1/24
                mac 00:00:00:1e:01:01
                vrrp 1 passive
                    backup 10.0.1.254
                    ping-reply
                    traceroute-reply
                exit
                vpls "bd-1"
                exit
            exit
            no shutdown
---------------------------------------------
```

The provisioning is easier with unnumbered SBD IRB because no IRB IP addresses need to be configured in the VPRN.

PE-1 receives the following RT-5 update for IP prefix 10.0.2.0/24 with GW MAC address 00:00:00:01:51:02, because there is no GW IP address. The GW MAC address is used in the VPRN route table, where the EVPN tunnel leads toward this GW MAC address.

```
*A:PE-1# show router bgp routes evpn ip-prefix
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
Flag  Route Dist.         Prefix
      Tag                 Gw Address
                          NextHop
                          Label
                          ESI
-------------------------------------------------------------------------------
u*>i  192.0.2.2:15        10.0.2.0/24
      0                   00:00:00:01:51:02
                          192.0.2.2
                          VNI 15
                          ESI-0

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

MAC advertisement is by default enabled, so PE-1 also receives the following RT-2 update for the GW MAC address. The interface is unnumbered, so there is no corresponding IP address.

```
*A:PE-1# show router bgp routes evpn mac
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
```

```
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.2:15        00:00:00:01:51:02 ESI-0
      0                   Static            VNI 15
                          n/a
                          192.0.2.2

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

The following traceroute from host 1 to host 2 shows that the second hop now is 10.0.2.2, which corresponds to the "bd-2" interface in VPRN "ip-vrf-151" on PE-2. The other hops remain the same as in the preceding case.

```
*A:PE-1# traceroute router-instance "host1" 10.0.2.222 source 10.0.1.111
traceroute to 10.0.2.222 from 10.0.1.111, 30 hops max, 40 byte packets
  1  10.0.1.1 (10.0.1.1)    2.22 ms  2.47 ms  2.56 ms
  2  10.0.2.2 (10.0.2.2)    3.75 ms  4.20 ms  3.91 ms
  3  10.0.2.222 (10.0.2.222)    4.60 ms  4.40 ms  4.63 ms
```

The following route table of VPRN "ip-vrf-151" on PE-1 shows a EVPN IFF route for IP prefix 10.0.2.0/24 with EVPN tunnel (ET) to GW MAC address 00:00:00:01:51:02 in VPRN "ip-vrf-151" on PE-2.

```
*A:PE-1# show router service-name "ip-vrf-151" route-table

===============================================================================
Route Table (Service: 151)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.1.0/24                                   Local   Local     00h06m07s  0
       int-bd-1                                                     0
10.0.2.0/24                                   Remote  EVPN-IFF  00h01m48s  169
       int-sbd-15 (ET-00:00:00:01:51:02)                            0
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The following ARP table for VPRN "ip-vrf-151" does not contain any entries for the unnumbered interface "int-sbd-15":

```
*A:PE-1# show service id "ip-vrf-151" arp

===============================================================================
ARP Table
===============================================================================
IP Address      MAC Address       Type    Expiry    Interface         SAP
-------------------------------------------------------------------------------
10.0.1.1        00:00:00:1e:01:01 Other   00h00m00s int-bd-1          rvpls
10.0.1.111      00:00:00:10:11:01 Dynamic 03h55m43s int-bd-1          rvpls
10.0.1.254      00:00:5e:00:01:01 Other   00h00m00s int-bd-1          rvpls
===============================================================================
```

However, internally, ARP entries are created. The following command shows that the same number of ARP entries are consumed as in the preceding use case with the numbered interface "int-sbd-15". The BGP-EVPN ARP entry corresponds to the GW interface "int-sbd-15" on the BGP peer.

```
*A:PE-1# show router service-name "ip-vrf-151" arp summary

============================================================
ARP Table Summary (Service: 151)
============================================================
Local ARP Entries    : 3
Static ARP Entries   : 0
Dynamic ARP Entries  : 1
Managed ARP Entries  : 0
Internal ARP Entries : 0
BGP-EVPN ARP Entries : 1
------------------------------------------------------------
No. of ARP Entries   : 5
============================================================
```

The FDB for R-VPLS "ip-vrf-151" on PE-1 is as follows:

```
*A:PE-1# show service id "sbd-15" fdb detail

===============================================================================
Forwarding Database, Service 15
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
15         00:00:00:01:51:01 cpm                     Intf     10/25/23 07:35:43
15         00:00:00:01:51:02 vxlan-1:                EvpnS:P  10/25/23 07:40:02
                             192.0.2.2:15
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

Interoperable interface-less model in EVPN-VXLAN

This model is interface-less because no SBD is required to connect the VPRNs and no recursive resolution is required upon receiving an IP prefix route. The next-hop of the IP prefix route is directly resolved to an EVPN tunnel, without the need for any other route.

The only difference from the preceding configuration is that MAC route advertisement is disabled in the backhaul R-VPLS on both PEs, as follows:

```
# on PE-1, PE-2:
configure
    service
        vpls "sbd-15"
            bgp-evpn
```

```
                no mac-advertisement
            exit
```

The configuration of the backhaul R-VPLS is as follows:

```
*A:PE-2# configure service vpls 15
*A:PE-2>config>service>vpls# info
----------------------------------------------
            description "backhaul R-VPLS 15"
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 15 create
            exit
            bgp
            exit
            bgp-evpn
                no mac-advertisement
                ip-route-advertisement
                evi 15
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
----------------------------------------------
```

Again, the provisioning is easier with unnumbered SBD IRB because no IRB IP addresses need to be configured in the VPRN.

PE-1 receives the following BGP-EVPN RT-5 update for IP prefix 10.0.2.0/24 with GW MAC address 00:00:00:01:51:02, which is the same as in the preceding use case:

```
*A:PE-1# show router bgp routes evpn ip-prefix
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
Flag  Route Dist.         Prefix
      Tag                 Gw Address
                          NextHop
                          Label
                          ESI
-------------------------------------------------------------------------------
u*>i  192.0.2.2:15        10.0.2.0/24
      0                   00:00:00:01:51:02
                          192.0.2.2
                          VNI 15
                          ESI-0

-------------------------------------------------------------------------------
Routes : 1
```

```
===============================================================================
```

PE-1 does not receive any BGP-EVPN RT-2 updates because PE-2 does not advertise any MAC addresses in the backhaul R-VPLS, as follows:

```
*A:PE-1# show router bgp routes evpn mac
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
No Matching Entries Found.
===============================================================================
```

The following traceroute from host 1 to host 2 shows that the second hop is the IP address of the "int-bd-2" interface in VPRN "ip-vrf-151" on PE-2, as in the preceding use case:

```
*A:PE-1# traceroute router-instance "host1" 10.0.2.222 source 10.0.1.111
traceroute to 10.0.2.222 from 10.0.1.111, 30 hops max, 40 byte packets
  1  10.0.1.1 (10.0.1.1)    2.66 ms  2.08 ms  2.38 ms
  2  10.0.2.2 (10.0.2.2)    3.77 ms  3.95 ms  3.74 ms
  3  10.0.2.222 (10.0.2.222)    5.40 ms  4.38 ms  4.94 ms
```

The following route table for VPRN "ip-vrf-151" on PE-1 shows an EVPN IFF route for IP prefix 10.0.2.0/24 with EVPN tunnel:

```
*A:PE-1# show router service-name "ip-vrf-151" route-table

===============================================================================
Route Table (Service: 151)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.1.0/24                                   Local   Local     00h09m04s  0
       int-bd-1                                                     0
10.0.2.0/24                                   Remote  EVPN-IFF  00h04m45s  169
       int-sbd-15 (ET-00:00:00:01:51:02)                            0
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

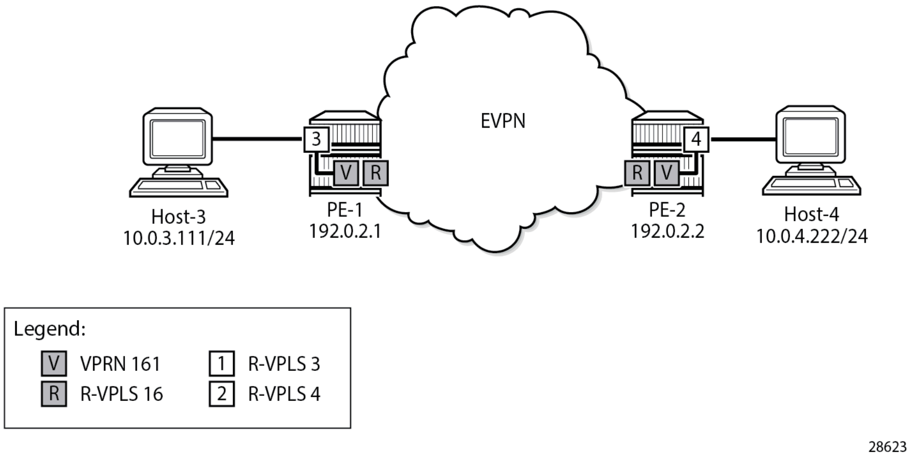
The following FDB in the backhaul R-VPLS on PE-1 shows an EVPN entry for GW MAC address 00:00:00:01:51:02, which is created out of the RT-5 GW MAC (router MAC extended community):

```
*A:PE-1# show service id "sbd-15" fdb detail

===============================================================================
Forwarding Database, Service 15
===============================================================================
ServId     MAC               Source-Identifier        Type     Last Change
            Transport:Tnl-Id                          Age
-------------------------------------------------------------------------------
15         00:00:00:01:51:01 cpm                      Intf     10/25/23 07:35:43
15         00:00:00:01:51:02 vxlan-1:                 Evpn     10/25/23 07:43:28
                             192.0.2.2:15
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

IP-VRF-to-IP-VRF models in EVPN-MPLS

The three IP-VRF-to-IP-VRF models are also supported in EVPN-MPLS. Figure 106: Example topology with services - EVPN-MPLS shows the example topology with the services R-VPLS "sbd-16", VPRN "ip-vrf-161", R-VPLS "bd-3" (or "bd-4"), and VPRN "host3" for host 3 (or VPRN "host4" for host 4).

Figure 106: Example topology with services - EVPN-MPLS



For MPLS, LDP is configured on the interface between PE-1 and PE-2.

Interface-ful model with SBD IRB in EVPN-MPLS

The following services are configured on PE-1 and PE-2:

- Backhaul R-VPLS "sbd-16"
- VPRN "ip-vrf-161"

- R-VPLS "bd-3" on PE-1; R-VPLS "bd-4" on PE-2
- VPRN "host3" on PE-1; VPRN "host4" on PE-2

The service configuration on PE-1 is as follows. MAC route advertisement is enabled by default. The configuration on PE-2 is similar.

```
# on PE-1:
configure
    service
        vpls 16 name "sbd-16" customer 1 create
            description "backhaul EVPN-MPLS R-VPLS 16"
            allow-ip-int-bind
            exit
            bgp
            exit
            bgp-evpn
                ip-route-advertisement          # mac-advertisement is by default enabled
                evi 16
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
        vprn 161 name "ip-vrf-161" customer 1 create
            ecmp 2
            interface "int-sbd-16" create
                address 172.16.161.1/24
                mac 00:00:00:01:61:01
                vpls "sbd-16"
                exit
            exit
            interface "int-bd-3" create
                address 10.0.3.1/24
                mac 00:00:00:3e:03:01
                vrrp 1 passive
                    backup 10.0.3.254
                    ping-reply
                    traceroute-reply
                exit
                vpls "bd-3"
                exit
            exit
            no shutdown
        exit
        vpls 3 name "bd-3" customer 1 create
            description "R-VPLS 3 - BD 3"
            allow-ip-int-bind
            exit
            stp
                shutdown
            exit
            sap pxc-10.a:3 create
                no shutdown
            exit
            no shutdown
        exit
        vprn 31 name "host3" customer 1 create
            description "Host-3 attached to R-VPLS 3"
            interface "local" create
                address 10.0.3.111/24
                mac 00:00:00:30:11:01
                sap pxc-10.b:3 create
                exit
            exit
            static-route-entry 0.0.0.0/0
                next-hop 10.0.3.254
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

PE-1 receives the following BGP-EVPN IP prefix route for prefix 10.0.4.0/24:

```
*A:PE-1# show router bgp routes evpn ip-prefix
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
Flag  Route Dist.         Prefix
      Tag                 Gw Address
                          NextHop
                          Label
                          ESI
-------------------------------------------------------------------------------
u*>i  192.0.2.2:16        10.0.4.0/24
      0                   172.16.161.2
                          192.0.2.2
                          LABEL 524286
                          ESI-0

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

The GW address is the IP address 172.16.161.2. The following BGP-EVPN MAC route advertises the corresponding MAC address 00:00:00:01:61:02:

```
*A:PE-1# show router bgp routes evpn mac
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
```

```
      Tag                   Mac Mobility      Label1
                            Ip Address
                            NextHop
--------------------------------------------------------------------------------
u*>i  192.0.2.2:16          00:00:00:01:61:02 ESI-0
      0                     Static            LABEL 524286
                            172.16.161.2
                            192.0.2.2

--------------------------------------------------------------------------------
Routes : 1
================================================================================
```

The following traceroute from host 3 to host 4 shows that the GW IP address is the second hop:

```
*A:PE-1# traceroute router-instance "host3" 10.0.4.222 source 10.0.3.111
traceroute to 10.0.4.222 from 10.0.3.111, 30 hops max, 40 byte packets
  1  10.0.3.1 (10.0.3.1)    5.60 ms  2.41 ms  2.39 ms
  2  172.16.161.2 (172.16.161.2)    3.99 ms  3.61 ms  3.88 ms
  3  10.0.4.222 (10.0.4.222)    11.4 ms  4.95 ms  4.89 ms
```

The route table and ARP table in VPRN 161 and the FDB in R-VPLS 16 are similar to the ones in theInterface-ful model with SBD IRB in EVPN-VXLAN section.

Interface-ful model with unnumbered SBD IRB in EVPN-MPLS

The GW IP addresses are removed from the "int-sbd-16" interface in VPRN "ip-vrf-161" and an EVPN tunnel is configured instead. On PE-1, VPRN "ip-vrf-161" is configured as follows:

```
*A:PE-1>config>service>vprn# info
----------------------------------------------
            ecmp 2
            interface "int-sbd-16" create
                mac 00:00:00:01:61:01
                vpls "sbd-16"
                    evpn-tunnel
                exit
            exit
            interface "int-bd-3" create
                address 10.0.3.1/24
                mac 00:00:00:3e:03:01
                vrrp 1 passive
                    backup 10.0.3.254
                    ping-reply
                    traceroute-reply
                exit
                vpls "bd-3"
                exit
            exit
            no shutdown
----------------------------------------------
```

The route table in VPRN "ip-vrf-161" and the FDB in R-VPLS "sbd-16" are similar to the ones in the Interface-ful model with unnumbered SBD IRB in EVPN-VXLAN section.

Interoperable interface-less model in EVPN-MPLS with Ethernet encapsulation

In the EVPN interface-less (EVPN IFL) model, the next hop of the IP prefix route is directly resolved to an EVPN tunnel, without the need for any other route.

MAC route advertisement is disabled in backhaul R-VPLS "sbd-16", as follows:

```
*A:PE-1# configure service vpls 16
*A:PE-1>config>service>vpls# info
----------------------------------------------
            description "backhaul EVPN-MPLS R-VPLS 16"
            allow-ip-int-bind
            exit
            bgp
            exit
            bgp-evpn
                no mac-advertisement
                ip-route-advertisement
                evi 16
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
----------------------------------------------
```

The following route table for VPRN "ip-vrf-161" contains a EVPN IFF entry for prefix 10.0.4.0/24 with an EVPN tunnel to GW MAC address 00:00:00:01:61:02:

```
*A:PE-1# show router service-name "ip-vrf-161" route-table

===============================================================================
Route Table (Service: 161)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.3.0/24                                   Local   Local     00h03m58s  0
       int-bd-3                                                     0
10.0.4.0/24                                   Remote  EVPN-IFF  00h01m55s  169
       int-sbd-16 (ET-00:00:00:01:61:02)                            0
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The following FDB for VPLS "sbd-16" contains an EVPN entry for GW MAC address 00:00:00:01:61:02. This information is retrieved from a BGP-EVPN IP prefix route.

```
*A:PE-1# show service id "sbd-16" fdb detail
```

```
===============================================================================
Forwarding Database, Service 16
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
16         00:00:00:01:61:01 cpm                     Intf     10/25/23 07:47:18
16         00:00:00:01:61:02 mpls-1:                 Evpn     10/25/23 07:50:40
                             192.0.2.2:524286
           ldp:65537
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

The IP prefix route for prefix 10.0.4.0/24 has GW MAC address 00:00:00:01:61:02, as follows:

```
*A:PE-1# show router bgp routes evpn ip-prefix
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
Flag  Route Dist.         Prefix
      Tag                 Gw Address
                          NextHop
                          Label
                          ESI
-------------------------------------------------------------------------------
u*>i  192.0.2.2:16        10.0.4.0/24
      0                   00:00:00:01:61:02
                          192.0.2.2
                          LABEL 524286
                          ESI-0

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

However, no EVPN MAC routes were received for R-VPLS 16, as follows:

```
*A:PE-1# show router bgp routes evpn mac
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
No Matching Entries Found.
===============================================================================
```

The interoperable interface-less model in EVPN-MPLS with Ethernet encapsulation is interface-ful although compatible with EVPN interface-less.

Interface-less model in EVPN-MPLS with IP encapsulation

In this IP NVO model, the ingress PE no longer pushes an inner Ethernet header, but the IP packet is directly encapsulated with an EVPN service label and the transport labels.

The PEs advertise IP prefixes without router MAC extended community. The route lookup in the VPRN does not point at an SBD R-VPLS, but rather to an MPLS tunnel terminated in the other PE. The packets are sent with an EVPN service label that was received in the IP prefix route.

The configuration of VPRN "ip-vrf-161" is modified: the interface "int-sbd-16" is removed and a BGP-EVPN context is added with route distinguisher, VRF target, and auto-bind tunnel. VPLS "sbd-16" is not used at all. The following shows the configuration of VPRN "ip-vrf-161" on PE-1:

```
*A:PE-1>config>service>vprn# info
----------------------------------------------
            ecmp 2
            interface "int-bd-3" create
                address 10.0.3.1/24
                mac 00:00:00:3e:03:01
                vrrp 1 passive
                    backup 10.0.3.254
                    ping-reply
                    traceroute-reply
                exit
                vpls "bd-3"
                exit
            exit
            bgp-evpn
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    route-distinguisher 192.0.2.1:161
                    vrf-target target:64500:161
                    no shutdown
                exit
            exit
            no shutdown
----------------------------------------------
```

The configuration on PE-2 is similar.

The following route table shows that the EVPN route is interface-less, the next hop is the IP address of PE-2, and the tunnel is an MPLS (LDP) tunnel instead of an EVPN tunnel:

```
*A:PE-1# show router service-name "ip-vrf-161" route-table

===============================================================================
Route Table (Service: 161)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
```

```
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.3.0/24                                   Local   Local     00h06m31s  0
       int-bd-3                                                     0
10.0.4.0/24                                   Remote  EVPN-IFL  00h03m57s  170
       192.0.2.2 (tunneled)                                          10
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The following EVPN IP prefix does not have any GW address:

```
*A:PE-1# show router bgp routes evpn ip-prefix
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
Flag  Route Dist.         Prefix
      Tag                 Gw Address
                          NextHop
                          Label
                          ESI
-------------------------------------------------------------------------------
u*>i  192.0.2.2:161       10.0.4.0/24
      0                   00:00:00:00:00:00
                          192.0.2.2
                          LABEL 524284
                          ESI-0

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

Conclusion

The three EVPN IP-VRF-to-IP-VRF models each have advantages. Different vendors have chosen different models in the first phases of their EVPN implementations. SR OS supports all three EVPN IP-VRF-to-IP-VRF models, so they can be deployed in all environments where third-party vendors are deployed already.

EVPN Multi-Homing for VXLAN VPLS Services

This chapter provides information about EVPN Multi-Homing for VXLAN VPLS Services.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

The information and configuration in this chapter are based on SR OS Release 21.7.R1.

EVPN multi-homing has been supported in SR OS for EVPN-MPLS and PBB-EVPN in SR OS Release 13.0.R4 and later. SR OS Release 16.0 introduced EVPN multi-homing for EVPN-VXLAN on Epipe services. EVPN-VXLAN multi-homing in a single VXLAN instance VPLS or R-VPLS service—as specified in RFC 8365—is supported in SR OS Release 19.5.R1, and later.

Before you read this chapter, ensure you are familiar with the concepts in the EVPN for VXLAN Tunnels (Layer 2) chapter.

Overview

Some Service Providers are deploying large Telco cloud Data Centers (DCs) where SR OS nodes are used as leaf switches in a VXLAN fabric. In those cases, all-active multi-homing can provide redundancy and maximize the bandwidth use.

The multi-homing procedures consist of three components:

- Designated Forwarder (DF) election
  - The PEs attached to the same Ethernet Segment (ES) elect a single PE as DF to:
    - forward all traffic, in case of single-active mode
    - forward all Broadcast, Unknown unicast, Multicast (BUM) traffic, in case of all-active mode
- split-horizon
  - BUM traffic received from a peer ES PE is filtered so that it is not looped back to the CE that first transmitted the frame.
  - in EVPN-VXLAN services, split-horizon is only used with all-active mode and makes use of the local bias procedure described in RFC 8365.
- aliasing

- PEs that are not attached to the ES can process non-zero Ethernet Segment Identifier (ESI) MAC/IP routes and AD routes and create ES destinations to which per-flow Equal Cost Multi-Path (ECMP) can be applied.
- Aliasing only applies to all-active mode.

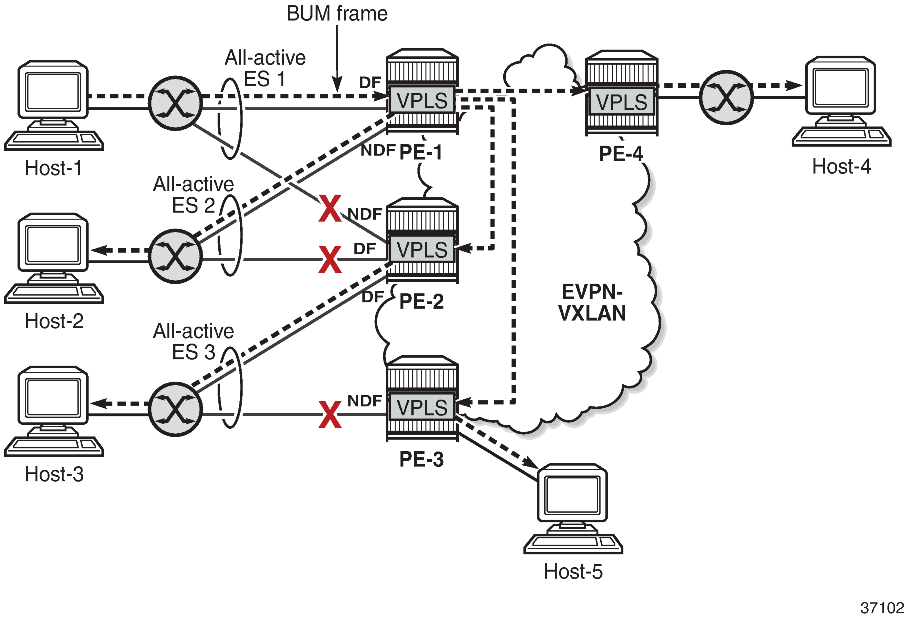
Split-horizon using local bias

In EVPN-MPLS services, split-horizon filtering uses ESI labels. VXLAN does not support ESI labels or MPLS labels. In EVPN-VXLAN services, the split-horizon filtering is based on the tunnel source IP address. In RFC 8365, this forwarding is referred to as local bias. Local bias works as follows:

- Every PE knows the IP addresses associated with the other PEs with which it has shared multi-homed ESs.
- The ingress PE replicates locally to all directly attached ESs, regardless of the DF state, for all flooded traffic coming from the access interfaces. BUM frames received on any SAP are flooded to:
  - local non-ES SAPs and non-ES SDP bindings
  - local all-active ES SAPs (DF and NDF)
  - local single-active ES SDP bindings and SAPs (DF only)
  - EVPN-VXLAN destinations
- When an egress PE receives a BUM frame from a VXLAN binding, it looks up the source IP address in the tunnel header and filters out the frame on all local interfaces connected to ESs that are shared with the ingress PE. The following rules apply to egress PE forwarding for EVPN-VXLAN services.
  1. The source VTEP is looked up for BUM frames received on EVPN-VXLAN.
  2. The router checks if the source VTEP matches one of the PEs with which the egress PE shared both an ES and a VXLAN service.
     - If there is a match, the egress PE is not forwarding to the shared ES local SAPs.
     - If there is no match, the egress PE forwards to ES SAPs in DF state (as usual).

Figure 107: Split-horizon filtering based on tunnel source IP address shows an example of local bias forwarding for BUM frames.

Figure 107: Split-horizon filtering based on tunnel source IP address



In this example, BUM frames sent by Host-1 are treated as follows.

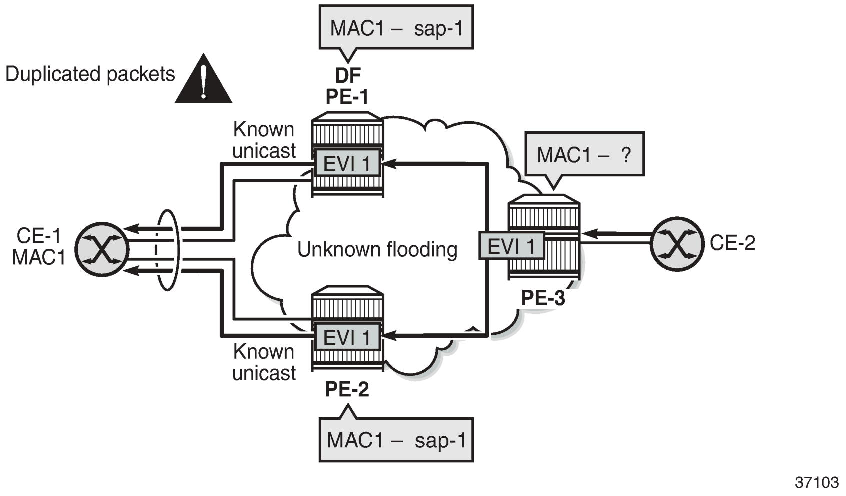
- Ingress node PE-1 receives BUM frames from Host-1 and forwards them to the other PEs (EVPN-VXLAN destinations) and the local all-active ES SAP toward Host-2, even though the SAP is in NDF state.
- Egress node PE-2 receives BUM frames on VXLAN. PE-2 identifies the source VTEP as a PE with which two all-active ESs are shared, so it does not forward the BUM frames to the two shared ESs. PE-2 forwards the BUM frames to the non-shared ES toward Host-3 because it is in DF state.
- Egress node PE-3 receives BUM traffic from PE-1, with which it does not share any ESs, so it forwards the BUM frames based on normal rules: it does not forward them toward Host-3, because the ES SAP is in NDF state. PE-3 only forwards toward Host-5.
- PE-4 does not share any ESs with PE-1, so the normal rules apply. PE-4 forwards the BUM frames toward Host-4.

Known limitations for local bias

In VXLAN, there are no BUM labels or any tunnel indication that can identify BUM traffic. The egress PE must solely rely on the Customer MAC (CMAC) destination address and this may create transient issues.

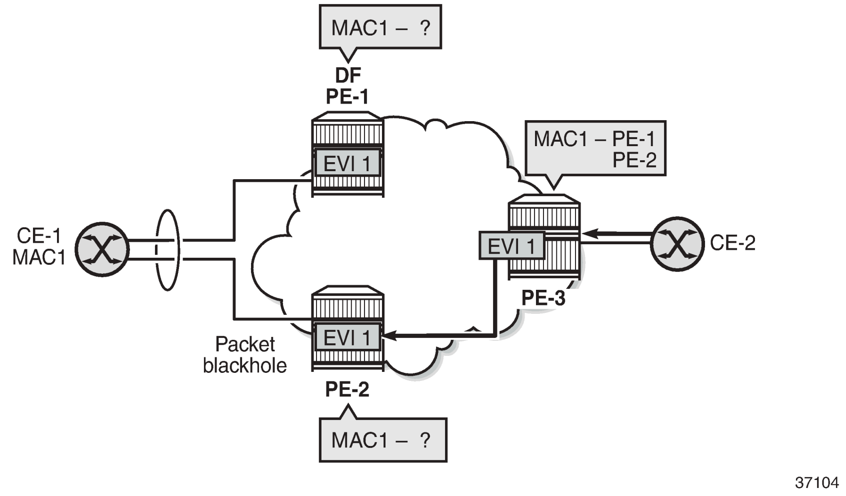
- Duplicate unicast traffic may occur when the CMAC destination address MAC1 is unknown on the ingress PE-3, while known on the egress PEs (PE-1 and PE-2). Figure 108: Duplicate unicast packets when MAC1 is unknown on PE-3 only shows that a packet with destination MAC1 arrives at PE-3, where it is flooded via ingress replication to PE-1 and PE-2, where MAC1 is known. PE-1 and PE-2 both forward the packets with CMAC destination MAC1 to CE-1, so multiple copies are sent to CE-1.

Figure 108: Duplicate unicast packets when MAC1 is unknown on PE-3 only



- A blackhole may occur when the CMAC destination address MAC1 is known on PE-3, but unknown on PE-1 and PE-2 and the aliasing hashing on PE-3 picks up the path to the NDF, where unknown unicast traffic is dropped, as shown in Figure 109: Packet blackhole for traffic on NDF PE-2 when MAC1 is known on PE-3 only. When the path to the DF is picked, no problem occurs, because the DF forwards BUM traffic.

Figure 109: Packet blackhole for traffic on NDF PE-2 when MAC1 is known on PE-3 only

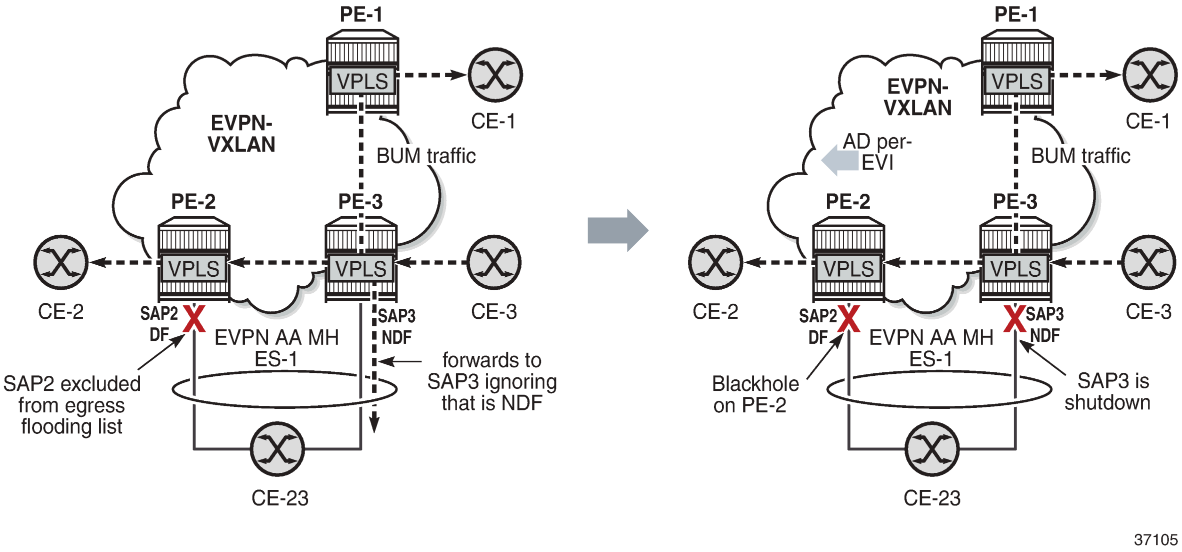


- A blackhole can be created when a remote SAP is disabled (**shutdown**), as shown in Figure 110: Blackhole created when a remote SAP is disabled.

  Under normal circumstances, when CE-3 sends BUM traffic to ingress node PE-3, the local bias mechanism on PE-3 forwards the BUM packets to SAP3, even though it is NDF for the ES. The BUM traffic is also flooded to PE-2, where it is forwarded to CE-2, but not to SAP2, because the ES is shared with PE-3.

When SAP3 is manually disabled, PE-3 withdraws the AD per-EVI route corresponding to SAP3. This does not change the local bias filtering for SAP2 on PE-2, so when CE-3 sends BUM traffic, it can neither be forwarded to CE-23 via SAP3 nor by PE-2.

Figure 110: Blackhole created when a remote SAP is disabled



CLI

The multi-homing capabilities are enabled in all the PEs attached to the VPLS service by configuring the options **auto-disc-route-advertisement** and **mh-mode network** in the **vpls bgp-evpn vxlan** context.

The **auto-disc-route-advertisement** option is by default disabled, but it can be enabled as follows:

```
*A:PE-2>config>service>vpls>bgp-evpn>vxlan$ auto-disc-route-advertisement
```

This **auto-disc-route-advertisement** command is only configurable for EVPN-VXLAN VPLS services and is implicitly enabled on all instances where it is not configurable. **Auto-disc-route-advertisement** is required in nodes with local ESs and remote ESs to process and enable the creation of ES destinations.

When auto-disc-route-advertisement is enabled, BGP-EVPN:

- processes Auto-Discovery per EVPN instance (AD per-EVI) routes and AD per-ES routes
- processes MAC/IP routes with non-zero Ethernet Segment Identifier (ESI) — without resetting the ESI to zero
- creates ES destinations upon receiving MAC/IP routes and AD per-ES/EVI routes with non-zero ESI

The **mh-mode** option can be configured with the values **access** or **network**. For EVPN-VXLAN services, the default value is **access**. The following command configures **mh-mode network**:

```
*A:PE-2>config>service>vpls>bgp-evpn>vxlan$ mh-mode network
```

When **mh-mode network** is configured, BGP-EVPN:

- activates multi-homing for the local ES SAPs or SDP-bindings and creates ES associations and related processes, such as:
  - the local bias mode allowing the system to add all-active SAPs to the flooding list regardless of the DF state
  - the source VTEP lookup mode
- runs DF election for the ESs associated to the service
- triggers the advertisement of AD per-ES routes, AD per-EVI routes, and non-zero MAC/IP routes for the ESs in the service

Configuration

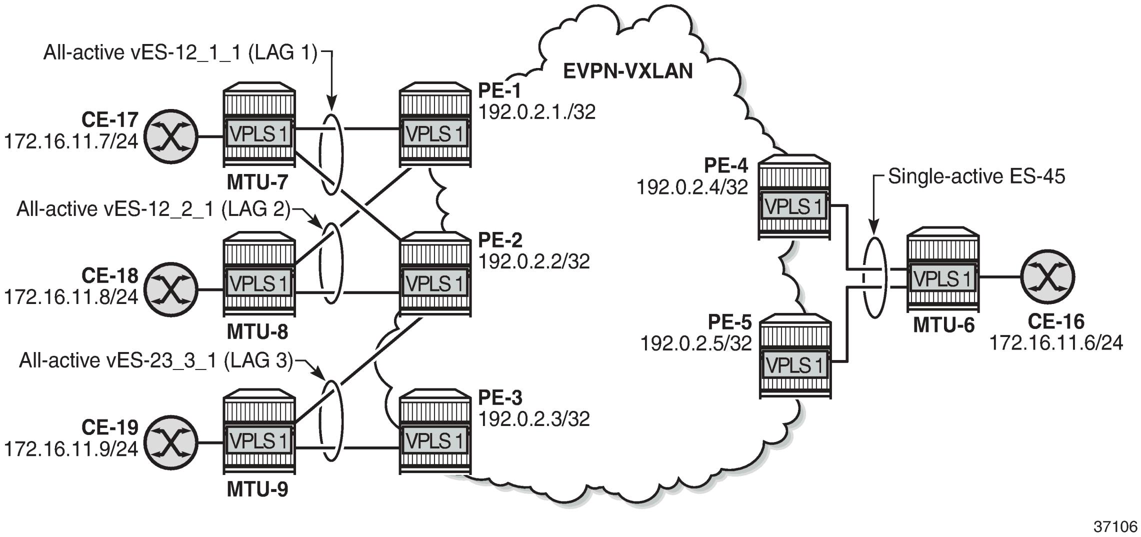
The following examples are configured:

- EVPN-VXLAN multi-homing with system IPv4 VTEP addresses
- EVPN-VXLAN multi-homing with non-system IPv4 VTEP addresses
- EVPN-VXLAN multi-homing with non-system IPv6 VTEP addresses

EVPN-VXLAN multi-homing with system IPv4 VTEP addresses

Figure 111: Example topology shows the topology with three all-active multi-homing ESs and one single-active multi-homing ES. This example shows the configuration for virtual Ethernet Segments, as described in the Virtual Ethernet Segments chapter, but non-virtual ES can also be used.

Figure 111: Example topology



The initial configuration on the PEs includes:

- cards, MDAs, ports
- LAG 1 on MTU-7, PE-1, PE-2

LAG 2 on MTU-8, PE-1, PE-2

LAG 3 on MTU-9, PE-2, PE-3

- router interfaces
- IS-IS between the PEs
- SR-ISIS between PE-4 and MTU-6 and between PE-5 and MTU-6 (and TLDP for SDP signaling)

BGP is configured between the PEs for the EVPN address family. PE-1 acts as route reflector, as follows:

```
# on RR PE-1:
configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            rapid-update evpn
            group "internal"
                family evpn
                cluster 192.0.2.1
                peer-as 64500
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
        exit
```

ES configuration

The all-active ESs "vES-12_1_1" and "vES-12_2_1" are configured on PE-1 and PE-2. The configuration on PE-1 is as follows. The configuration on PE-2 is similar, but with different preference values.

```
# on PE-1:
configure
    service
        system
            bgp-evpn
                ethernet-segment "vES-12_1_1" virtual create
                    esi 00:12:12:12:12:12:12:00:01:01
                    service-carving
                        mode manual
                        manual
                            preference create
                                value 100            # on PE-2: preference value 150
                            exit
                            evi 1
                        exit
                    exit
                    multi-homing all-active
                    lag 1
                    dot1q
```

```
                        q-tag-range 1
                    exit
                    no shutdown
                exit
                ethernet-segment "vES-12_2_1" virtual create
                    esi 00:12:12:12:12:12:12:00:02:01
                    service-carving
                        mode manual
                        manual
                            preference create
                                value 150            # on PE-2: preference value 100
                            exit
                            evi 1
                        exit
                    exit
                    multi-homing all-active
                    lag 2
                    dot1q
                        q-tag-range 1
                    exit
                    no shutdown
                exit
            exit
        exit
```

On PE-2 and PE-3, the all-active ES "vES-23_3_1" is configured in a similar way:

```
# on PE-2:
configure
    service
        system
            bgp-evpn
                ethernet-segment "vES-23_3_1" virtual create
                    esi 00:23:23:23:23:23:23:00:03:01
                    service-carving
                        mode manual
                        manual
                            preference create
                                value 100            # on PE-3: preference value 150
                            exit
                            evi 1
                        exit
                    exit
                    multi-homing all-active
                    lag 3
                    dot1q
                        q-tag-range 1
                    exit
                    no shutdown
                exit
```

On PE-4 and PE-5, the single-active ES "ES-45" is configured, as follows:

```
# on PE-4:
configure
    service
        sdp 46 mpls create                           # on PE-5: sdp 56
            far-end 192.0.2.6
            sr-isis
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        system
            bgp-evpn
                ethernet-segment "ES-45" create
                    esi 00:45:45:45:45:45:45:00:00:01
                    service-carving
                        mode manual
                        manual
                            preference create
                                value 100            # on PE-5: preference value 150
                            exit
                            evi 1
                        exit
                    exit
                    multi-homing single-active
                    sdp 46                           # on PE-5: sdp 56
                    no shutdown
                exit
            exit
        exit
```

VPLS configuration

VPLS 1 is configured on PE-2 as follows. The configuration is similar on PE-1 and PE-3.

```
# on PE-2:
configure
    service
        system
            bgp-auto-rd-range 192.0.2.2 comm-val 1 to 1000   # different values on PEs
        exit
        vpls 1 name "VPLS 1" customer 1 create
            vxlan instance 1 vni 1 create
            exit
            bgp
                route-distinguisher auto-rd
                route-target export target:64500:1 import target:64500:1
            exit
            bgp-evpn
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    ecmp 2
                    auto-disc-route-advertisement
                    mh-mode network
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap lag-1:1 create                       # LAG 1 also on PE-1, not on PE-3
                no shutdown
            exit
            sap lag-2:1 create                       # LAG 2 also on PE-1, not on PE-3
                no shutdown
            exit
            sap lag-3:1 create                       # LAG 3 also on PE-3, not on PE-1
                no shutdown
            exit
            no shutdown
        exit
```

The EVPN-VXLAN multi-homing capabilities are enabled in the PEs attached to VPLS 1 by the commands **auto-disc-route-advertisement** and **mh-mode network**. The **auto-disc-route-advertisement** command enables the advertisement and processing of multi-homing routes, and the **mh-mode network** command activates the DF election procedures.

ECMP is required for per-flow load balancing for VXLAN ES destinations with two or more next hops. In this example, ECMP is configured with a value of 2.

On PE-4, VPLS 1 is configured as follows. The configuration on PE-5 is similar.

```
# on PE-4:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            vxlan instance 1 vni 1 create
            exit
            bgp
                route-distinguisher auto-rd
                route-target export target:64500:1 import target:64500:1
            exit
            bgp-evpn
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    ecmp 2
                    auto-disc-route-advertisement
                    mh-mode network
                    no shutdown
                exit
            exit
            spoke-sdp 46:1 create                    # on PE-5: spoke-sdp 56:1
            exit
            no shutdown
        exit
```

Show commands

The following command shows that the commands **mh-mode network** and **auto-disc-route-advertisement** are enabled:

```
*A:PE-2# show service id 1 bgp-evpn

===============================================================================
BGP EVPN Table
===============================================================================
MAC Advertisement  : Enabled            Unknown MAC Route  : Disabled
CFM MAC Advertise  : Disabled
Creation Origin    : manual
MAC Dup Detn Moves : 5                  MAC Dup Detn Window: 3
MAC Dup Detn Retry : 9                  Number of Dup MACs : 0
MAC Dup Detn BH    : Disabled
IP Route Advert    : Disabled
Sel Mcast Advert   : Disabled

EVI                : 1
Ing Rep Inc McastAd: Enabled
Accept IVPLS Flush : Disabled

-------------------------------------------------------------------------------
Detected Duplicate MAC Addresses             Time Detected
-------------------------------------------------------------------------------
```

```
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN VXLAN Information
===============================================================================
Admin Status       : Enabled            Bgp Instance       : 1
Vxlan Instance     : 1
Max Ecmp Routes    : 2
Default Route Tag  : none
Send EVPN Encap    : Enabled
Imet-Ir routes     : Enabled
MH Mode            : network
Auto Disc Route Adv: Enabled
Oper Group         :
===============================================================================
```

The following command shows that PE-1 is DF for the all-active ES vES-12_1_1 and NDF for the all-active ES vES-12_2_1:

```
*A:PE-1# show service id 1 ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
lag-1:1               vES-12_1_1                       DF
lag-2:1               vES-12_2_1                       NDF
===============================================================================
No sdp entries
No vxlan instance entries
```

The following command shows that PE-2 is NDF for the all-active ES vES-12_1_1 and DF for the other two all-active ESs:

```
*A:PE-2# show service id 1 ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
lag-1:1               vES-12_1_1                       NDF
lag-2:1               vES-12_2_1                       DF
lag-3:1               vES-23_3_1                       DF
===============================================================================
No sdp entries
No vxlan instance entries
```

PE-3 is NDF for the all-active multi-homing ES vES-23_3_1:

```
*A:PE-3# show service id 1 ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
lag-3:1               vES-23_3_1                       NDF
===============================================================================
No sdp entries
No vxlan instance entries
```

PE-4 is DF for the single-active multi-homing ES ES-45:

```
*A:PE-4# show service id 1 ethernet-segment
No sap entries

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
46:1                  ES-45                            DF
===============================================================================
No vxlan instance entries
```

PE-5 is NDF for the single-active multi-homing ES ES-45:

```
*A:PE-5# show service id 1 ethernet-segment
No sap entries

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
56:1                  ES-45                            NDF
===============================================================================
No vxlan instance entries
```

The following command shows the VXLAN destinations for VPLS 1 on PE-3; the system addresses of the other PEs act as destination VTEP addresses.

```
*A:PE-3# show service id 1 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
Instance    VTEP Address                            Egress VNI  EvpnStatic Num
 Mcast       Oper State                              L2 PBR     SupBcasDom MACs
-------------------------------------------------------------------------------
1           192.0.2.1                               1           evpn       0
 BUM         Up                                      No          No
1           192.0.2.2                               1           evpn       0
 BUM         Up                                      No          No
1           192.0.2.4                               1           evpn       0
 BUM         Up                                      No          No
1           192.0.2.5                               1           evpn       0
 BUM         Up                                      No          No
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 4
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
1         00:12:12:12:12:12:12:00:01:01   1             08/26/2021 07:17:08
1         00:12:12:12:12:12:12:00:02:01   1             08/26/2021 07:17:18
```

```
1         00:45:45:45:45:45:45:00:00:01    1             08/26/2021 07:17:19
--------------------------------------------------------------------------------
Number of entries: 3
--------------------------------------------------------------------------------
================================================================================
```

The following command on PE-3 shows the EVPN-VXLAN destination next hops (192.0.2.1 and 192.0.2.2) for alias ESI 00:12:12:12:12:12:12:00:01:01. The VTEP addresses 192.0.2.1 and 192.0.2.2 are the system addresses of PE-1 and PE-2.

```
*A:PE-3# show service id 1 vxlan esi 00:12:12:12:12:12:12:00:01:01

================================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
================================================================================
Instance  Eth SegId                       Num. Macs     Last Change
--------------------------------------------------------------------------------
1         00:12:12:12:12:12:12:00:01:01    1             08/26/2021 07:17:18
--------------------------------------------------------------------------------
Number of entries: 1
--------------------------------------------------------------------------------
================================================================================

================================================================================
BGP EVPN-VXLAN Dest TEP Info
================================================================================
Instance  TEP Address                   Egr VNI             Last Change
--------------------------------------------------------------------------------
1         192.0.2.1                     1                   08/26/2021 07:17:18
1         192.0.2.2                     1                   08/26/2021 07:17:18
--------------------------------------------------------------------------------
Number of entries : 2
--------------------------------------------------------------------------------
================================================================================
```

Tools command to check local bias

The following **tools** command on PE-2 checks whether local bias is enabled for the peers in ES "vES-12_1_1". The output lists the PEs that are in the candidate DF election list for the ES and whether local bias procedures are enabled on them. In this case, only peer 192.0.2.1 is in the list and local bias is enabled. The output is similar for ES "vES-12_2_1".

```
*A:PE-2# tools dump service system bgp-evpn ethernet-segment "vES-12_1_1" local-bias
--------------------------------------------------------------------------------
[08/26/2021 07:20:48] Vxlan Local Bias Information
------------------------------------------------------------------------+-------
Peer                                                                    | Enabled
------------------------------------------------------------------------+-------
192.0.2.1                                                               | Yes
--------------------------------------------------------------------------------
```

The PE can only enable local bias procedures on a maximum of three PEs that are attached to the same ES and use multi-homed VXLAN services. If more than three PEs exist, the PEs are ordered by preference or IP address and only the top three PEs are considered for local bias. The order is as follows:

- lowest IP address (automatic service-carving)
- lowest preference (manual service-carving with configured EVI)

- highest preference (manual service-carving without configured EVI)

The following **tools** command on PE-2 shows that local bias is enabled for peer 192.0.2.3 in ES "vES-23_3_1":

```
*A:PE-2# tools dump service system bgp-evpn ethernet-segment "vES-23_3_1" local-bias
-------------------------------------------------------------------------------
[08/26/2021 07:20:48] Vxlan Local Bias Information
----------------------------------------------------------------------+--------
Peer                                                                  | Enabled
----------------------------------------------------------------------+--------
192.0.2.3                                                             | Yes
-------------------------------------------------------------------------------
```

Verify local bias for BUM traffic in all-active multi-homing ESs

Unknown unicast traffic is generated on MTU-7. This traffic is received in ingress queue 11 for SAP lag-1:1 on ingress node PE-1, as follows:

```
*A:PE-1# monitor service id 1 sap lag-1:1

===============================================================================
Monitor statistics for Service 1 SAP lag-1:1
===============================================================================
---snip---
-------------------------------------------------------------------------------
Sap per Queue Stats
-------------------------------------------------------------------------------
                        Packets                 Octets

Ingress Queue 1 (Unicast) (Priority)
Off. HiPrio           : 0                       0
Off. LowPrio          : 0                       0
Dro. HiPrio           : 0                       0
Dro. LowPrio          : 0                       0
For. InProf           : 0                       0
For. OutProf          : 0                       0

Ingress Queue 11 (Multipoint) (Priority)
Off. Combined         : 6                       408
Off. Managed          : 0                       0
Dro. HiPrio           : 0                       0
Dro. LowPrio          : 0                       0
For. InProf           : 0                       0
For. OutProf          : 6                       408

Egress Queue 1
For. In/InplusProf    : 0                       0
For. Out/ExcProf      : 0                       0
Dro. In/InplusProf    : 0                       0
Dro. Out/ExcProf      : 0                       0

===============================================================================
```

On the ingress node PE-1, the local bias mechanism forwards this BUM traffic toward EVPN-VXLAN destinations, and also to the local SAPs of all-active ESs, regardless of the DF state. In this case, the local bias mechanism forwards the BUM traffic to lag-2:1 toward MTU-8, even though PE-1 is NDF in ES "vES-12_2_1".

```
*A:PE-1# monitor service id 1 sap lag-2:1

===============================================================================
Monitor statistics for Service 1 SAP lag-2:1
===============================================================================
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
Sap Statistics
-------------------------------------------------------------------------------
Last Cleared Time     : N/A

                        Packets                 Octets
CPM Ingress           : 0                       0
Forwarding Engine Stats
Dropped               : 0                       0
Received Valid        : 0                       0
Off. HiPrio           : 0                       0
Off. LowPrio          : 0                       0
Off. Uncolor          : 0                       0
Off. Managed          : 0                       0

Queueing Stats(Ingress QoS Policy 1)
Dro. HiPrio           : 0                       0
Dro. LowPrio          : 0                       0
For. InProf           : 0                       0
For. OutProf          : 0                       0

Queueing Stats(Egress QoS Policy 1)
Dro. In/InplusProf    : 0                       0
Dro. Out/ExcProf      : 0                       0
For. In/InplusProf    : 0                       0
For. Out/ExcProf      : 6                       408
-------------------------------------------------------------------------------
```

The egress PEs PE-2 and PE-3 receive the BUM traffic on the EVPN-VXLAN terminations. On egress PEs, the local bias mechanism filters BUM traffic based on the source IP address 192.0.2.1 of PE-1. PE-2 does not forward the traffic to the local SAPs lag-1:1 and lag-2:1, because PE-2 shares the all-active ESs "vES-12_1_1" and "vES-12_2_1" with PE-1. However, PE-2 forwards the BUM traffic to the non-shared ES "vES-23_3_1" because it is DF.

The following **monitor** commands show that PE-2 does not send any traffic toward SAP lag-1:1 or SAP lag-2:1.

```
*A:PE-2# monitor service id 1 sap lag-1:1
---snip---

Queueing Stats(Egress QoS Policy 1)
Dro. In/InplusProf    : 0                       0
Dro. Out/ExcProf      : 0                       0
For. In/InplusProf    : 0                       0
For. Out/ExcProf      : 0                       0
---snip---
```

```
*A:PE-2# monitor service id 1 sap lag-2:1
---snip---

Queueing Stats(Egress QoS Policy 1)
```

```
Dro. In/InplusProf    : 0                       0
Dro. Out/ExcProf      : 0                       0
For. In/InplusProf    : 0                       0
For. Out/ExcProf      : 0                       0
---snip---
```

The following **monitor** command shows that PE-2 forwards the traffic to SAP lag-3:1 toward MTU-9:

```
*A:PE-2# monitor service id 1 sap lag-3:1
---snip---

Queueing Stats(Egress QoS Policy 1)
Dro. In/InplusProf    : 0                       0
Dro. Out/ExcProf      : 0                       0
For. In/InplusProf    : 0                       0
For. Out/ExcProf      : 6                       408
---snip---
```

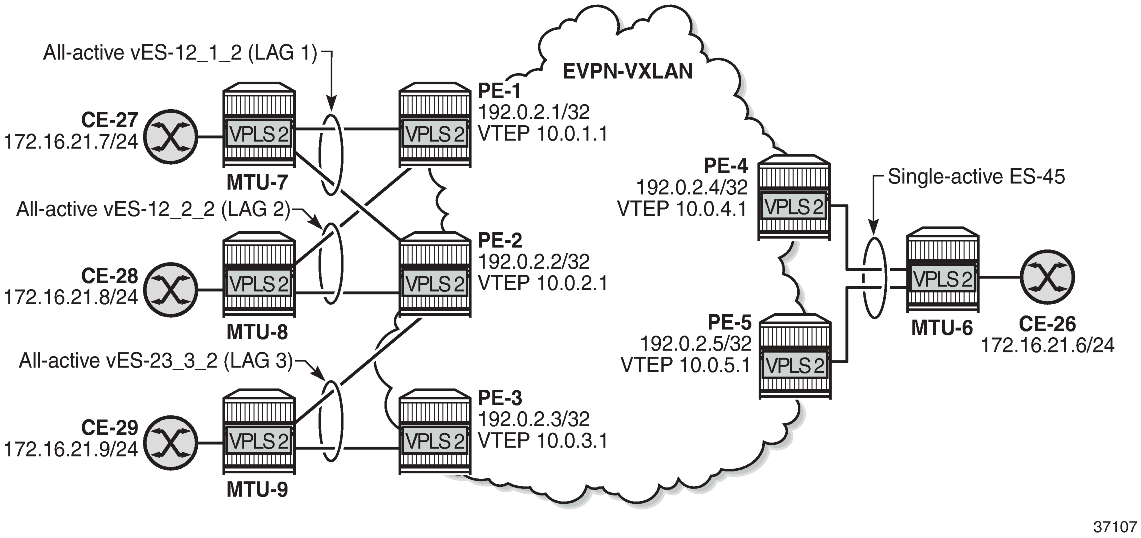
Egress node PE-3 receives BUM traffic on VXLAN and filters on IP address 192.0.2.1, but there are no shared ESs with PE-1. PE-3 is NDF for the non-shared ES vES-23_3_1, so it does not forward the traffic to SAP lag-3:1, as follows:

```
*A:PE-3# monitor service id 1 sap lag-3:1
---snip---

Queueing Stats(Egress QoS Policy 1)
Dro. In/InplusProf    : 0                       0
Dro. Out/ExcProf      : 0                       0
For. In/InplusProf    : 0                       0
For. Out/ExcProf      : 0                       0
---snip---
```

EVPN-VXLAN multi-homing with non-system IPv4 VTEP addresses

Figure 112: Non-system IPv4 VTEP multi-homing for VXLAN VPLS 2 shows the non-system IPv4 addresses to be used as VTEP addresses.

Figure 112: Non-system IPv4 VTEP multi-homing for VXLAN VPLS 2



Forwarding Path Extension (FPE), as described in the VXLAN Forwarding Path Extension chapter, is configured on all PEs. The configuration on PE-1 is as follows:

```
# on PE-1:
configure
    port-xc
        pxc 1 create
            port 1/2/6
            no shutdown
        exit
    exit
    port pxc-1.a
        ethernet
            encap-type dot1q
        exit
        no shutdown
    exit
    port pxc-1.b
        ethernet
            encap-type dot1q
        exit
        no shutdown
    exit
    port 1/2/6
        no shutdown
    exit
    fwd-path-ext
        sdp-id-range from 10000 to 10127
        fpe 1 create
            path pxc 1
            vxlan-termination
        exit
    exit
    router Base
        interface "loopback1"
            address 10.0.1.0/31
            loopback
            ipv6
```

```
                address 2001:db8::10:0/127
            exit
            no shutdown
        exit
        isis 0
            interface "loopback1"
                passive
                no shutdown
            exit
        exit
    exit
    service
        system
            vxlan
                tunnel-termination 10.0.1.1 fpe 1 create
                tunnel-termination 2001:db8::10:1 fpe 1 create
            exit
        exit
    exit
```

The configuration on the other PEs is similar but with different IP addresses, for example, 10.0.2.1 on PE-2, 10.0.3.1 on PE-3, and so on.

The non-system IP address in each of the PEs in the ES must match in the following three commands for the local PE to be considered suitable for DF election:

- **es-orig-ip** 10.0.x.1 (ES)

  The **es-orig-ip** command modifies the originating IP address in the ES routes advertised for the ES and makes the system use this IP address when adding the local PE as DF candidate.

- **route-next-hop** 10.0.x.1 (ES)

  The **route-next-hop** command changes the next hop of the ES routes and AD per-ES routes to the configured address.

- **vxlan-src-vtep** 10.0.x.1 (VPLS)

  The **vxlan-src-vtep** command makes the router use the configured IP address as the VXLAN tunnel source IP address (source VTEP) for originating VXLAN-encapsulated frames for the service. The source VTEP is also used to set the BGP NLRI next hop in EVPN route advertisements for the services.

The following all-active multi-homing ESs are configured on PE-2 with non-system IPv4 address 10.0.2.1:

```
# on PE-2:
configure
    service
        system
            bgp-evpn
                ethernet-segment "vES-12_1_2" virtual create
                    esi 00:12:12:12:12:12:12:00:01:02
                    es-orig-ip 10.0.2.1
                    route-next-hop 10.0.2.1
                    service-carving
                        mode manual
                        manual
                            preference create
                                value 150
                            exit
                        exit
                    exit
                    multi-homing all-active
                    lag 1
                    dot1q
```

```
                        q-tag-range 2
                    exit
                    no shutdown
                exit
                ethernet-segment "vES-12_2_2" virtual create
                    esi 00:12:12:12:12:12:12:00:02:02
                    es-orig-ip 10.0.2.1
                    route-next-hop 10.0.2.1
                    service-carving
                        mode manual
                        manual
                            preference create
                                value 100
                            exit
                        exit
                    exit
                    multi-homing all-active
                    lag 2
                    dot1q
                        q-tag-range 2
                    exit
                    no shutdown
                exit
                ethernet-segment "vES-23_3_2" virtual create
                    esi 00:23:23:23:23:23:23:00:03:02
                    es-orig-ip 10.0.2.1
                    route-next-hop 10.0.2.1
                    service-carving
                        mode manual
                        manual
                            preference create
                                value 100
                            exit
                        exit
                    exit
                    multi-homing all-active
                    lag 3
                    dot1q
                        q-tag-range 2
                    exit
                    no shutdown
                exit
            exit
        exit
```

The ES configuration on the other PEs is similar, but with different IP addresses and preference values.

VPLS 2 is configured with source VTEP 10.0.2.1 on PE-2:

```
# on PE-2:
configure
    service
        vpls 2 name "VPLS 2" customer 1 create
            vxlan-src-vtep 10.0.2.1                  # different IP address on different PEs
            vxlan instance 1 vni 2 create
            exit
            bgp
                route-distinguisher auto-rd
                route-target export target:64500:2 import target:64500:2
            exit
            bgp-evpn
                evi 2
                vxlan bgp 1 vxlan-instance 1
                    ecmp 2
                    auto-disc-route-advertisement
                    mh-mode network
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap lag-1:2 create                       # lag-1 is shared with PE-1
                no shutdown
            exit
            sap lag-2:2 create                       # lag-2 is shared with PE-1
                no shutdown
            exit
            sap lag-3:2 create                       # lag-3 is shared with PE-3
                no shutdown
            exit
            no shutdown
        exit
```

The configuration on the other PEs is similar.

Verification

The following command shows the DF status for the different ESs in VPLS 2 on PE-1:

```
*A:PE-1# show service id 2 ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                           Status
-------------------------------------------------------------------------------
lag-1:2               vES-12_1_2                        NDF
lag-2:2               vES-12_2_2                        DF
===============================================================================
No sdp entries
No vxlan instance entries
```

The following command on PE-1 shows that the source VTEP for VPLS 2 is 10.0.1.1:

```
*A:PE-1# show service id 2 vxlan
===============================================================================
VPLS VXLAN
===============================================================================
Vxlan Src Vtep IP: 10.0.1.1

===============================================================================
Vxlan Instance
===============================================================================
VXLAN Instance                VNI         AR          Oper-flags   VTEP
                                                                   security
-------------------------------------------------------------------------------
1                             2           none        none         disabled
-------------------------------------------------------------------------------
Number of Entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

The following command on PE-1 shows the (non-system) VXLAN destinations for VPLS 2:

```
*A:PE-1# show service id 2 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
Instance    VTEP Address                            Egress VNI  EvpnStatic Num
 Mcast       Oper State                              L2 PBR     SupBcasDom MACs
-------------------------------------------------------------------------------
1           10.0.2.1                                2           evpn       0
 BUM         Up                                      No          No
1           10.0.3.1                                2           evpn       0
 BUM         Up                                      No          No
1           10.0.4.1                                2           evpn       0
 BUM         Up                                      No          No
1           10.0.5.1                                2           evpn       0
 BUM         Up                                      No          No
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 4
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
1         00:23:23:23:23:23:23:00:03:02   1             08/26/2021 07:35:03
1         00:45:45:45:45:45:45:00:00:02   1             08/26/2021 07:34:38
-------------------------------------------------------------------------------
Number of entries: 2
-------------------------------------------------------------------------------
===============================================================================
```

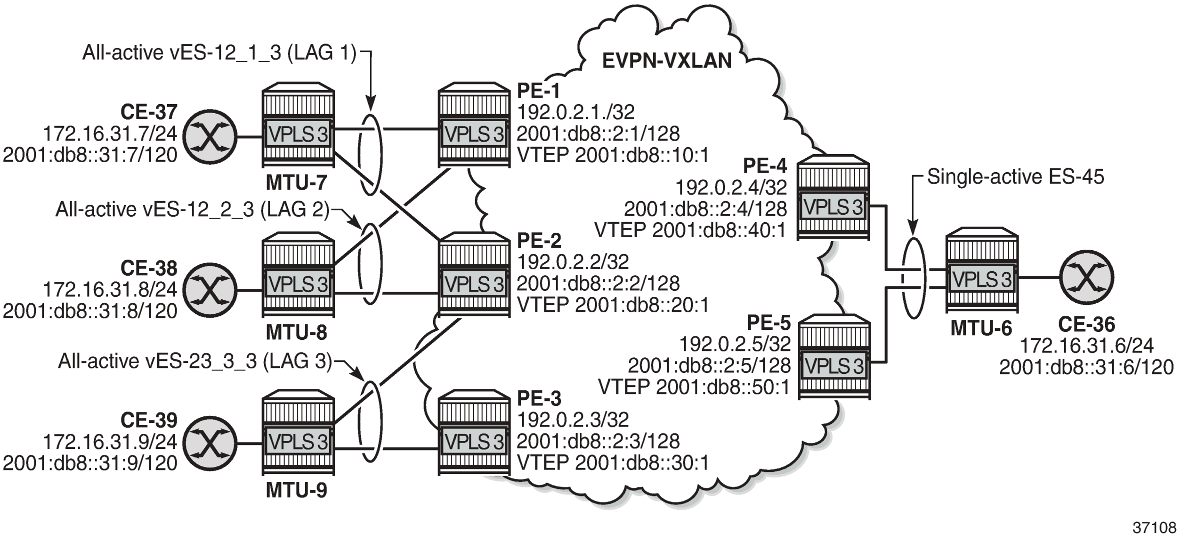
The non-system VTEP addresses in the all-active multi-homing ES with ESI 00:23:23:23:23:23:23:00:03:02 are 10.0.2.1 and 10.0.3.1, as follows:

```
*A:PE-1# show service id 2 vxlan esi 00:23:23:23:23:23:23:00:03:02

===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
1         00:23:23:23:23:23:23:00:03:02   1             08/26/2021 07:35:03
-------------------------------------------------------------------------------
Number of entries: 1
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-VXLAN Dest TEP Info
===============================================================================
Instance  TEP Address                   Egr VNI            Last Change
-------------------------------------------------------------------------------
1         10.0.2.1                      2                  08/26/2021 07:35:03
1         10.0.3.1                      2                  08/26/2021 07:35:03
-------------------------------------------------------------------------------
Number of entries : 2
-------------------------------------------------------------------------------
===============================================================================
```

EVPN-VXLAN multi-homing with non-system IPv6 VTEP addresses

Figure 113: Non-system IPv6 VTEP multi-homing for VXLAN VPLS 2 shows the non-system IPv6 addresses to be used as VTEP addresses.

Figure 113: Non-system IPv6 VTEP multi-homing for VXLAN VPLS 2



Between the PEs, the router interfaces have IPv6 addresses as well as IPv4 addresses, and **ipv6-routing native** is configured in IS-IS on the PEs. FPE is configured with VXLAN termination 2001:db8::x0:1 on PE-x.

The following all-active multi-homing ESs with non-system IPv6 addresses are configured on PE-2:

```
# on PE-2:
configure
    service
        system
            bgp-evpn
                ethernet-segment "vES-12_1_3" virtual create        # same ES on PE-1
                    esi 00:12:12:12:12:12:12:00:01:03
                    es-orig-ip 2001:db8::20:1
                    route-next-hop 2001:db8::20:1
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 1
                    dot1q
                        q-tag-range 3
                    exit
                    no shutdown
                exit
                ethernet-segment "vES-12_2_3" virtual create        # same ES on PE-1
                    esi 00:12:12:12:12:12:12:00:02:03
                    es-orig-ip 2001:db8::20:1
                    route-next-hop 2001:db8::20:1
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 2
                    dot1q
                        q-tag-range 3
                    exit
                    no shutdown
                exit
                ethernet-segment "vES-23_3_3" virtual create        # same ES on PE-3
                    esi 00:23:23:23:23:23:23:00:03:03
                    es-orig-ip 2001:db8::20:1
                    route-next-hop 2001:db8::20:1
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 3
                    dot1q
                        q-tag-range 3
                    exit
                    no shutdown
                exit
            exit
        exit
```

"VPLS 3" is configured with non-system source VTEP 2001:db8::x0:1, as follows:

```
# on PE-2:
configure
    service
        vpls 3 name "VPLS 3" customer 1 create
            vxlan-src-vtep 2001:db8::20:1
            vxlan instance 1 vni 3 create
            exit
            bgp
                route-distinguisher auto-rd
                route-target export target:64500:3 import target:64500:3
            exit
            bgp-evpn
                evi 3
                vxlan bgp 1 vxlan-instance 1
                    ecmp 2
                    auto-disc-route-advertisement
                    mh-mode network
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap lag-1:3 create                      # lag-1 shared with PE-1
                no shutdown
            exit
            sap lag-2:3 create                      # lag-2 shared with PE-1
                no shutdown
            exit
            sap lag-3:3 create                      # lag-3 shared with PE-3
                no shutdown
            exit
            no shutdown
        exit
```

Verification

The following command on PE-1 shows that the source VTEP is 2001:db8::10:1 for VPLS 3:

```
*A:PE-1# show service id 3 vxlan
================================================================================
VPLS VXLAN
================================================================================
Vxlan Src Vtep IP: 2001:db8::10:1

================================================================================
Vxlan Instance
================================================================================
VXLAN Instance                VNI         AR          Oper-flags   VTEP
                                                                   security
--------------------------------------------------------------------------------
1                             3           none        none         disabled
--------------------------------------------------------------------------------
Number of Entries : 1
--------------------------------------------------------------------------------
================================================================================
```

The following command on PE-1 shows the non-system IPv6 destination VTEPs for VPLS 3:

```
*A:PE-1# show service id 3 vxlan destinations

================================================================================
Egress VTEP, VNI
================================================================================
Instance    VTEP Address                            Egress VNI  EvpnStatic Num
 Mcast       Oper State                              L2 PBR     SupBcasDom MACs
--------------------------------------------------------------------------------
1           2001:db8::20:1                          3           evpn       0
 BUM         Up                                      No          No
1           2001:db8::30:1                          3           evpn       0
 BUM         Up                                      No          No
1           2001:db8::40:1                          3           evpn       0
 BUM         Up                                      No          No
1           2001:db8::50:1                          3           evpn       0
 BUM         Up                                      No          No
--------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 4
--------------------------------------------------------------------------------
================================================================================

================================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
================================================================================
Instance  Eth SegId                       Num. Macs     Last Change
--------------------------------------------------------------------------------
1         00:23:23:23:23:23:23:00:03:03   1             08/26/2021 07:41:20
1         00:45:45:45:45:45:45:00:00:03   1             08/26/2021 07:41:30
--------------------------------------------------------------------------------
Number of entries: 2
--------------------------------------------------------------------------------
================================================================================
```

The following command on PE-3 shows that VTEPs 2001:db8::10:1 and 2001:db8::20:1 are destinations in the all-active ES with ESI 00:12:12:12:12:12:12:00:01:03:

```
*A:PE-3# show service id 3 vxlan esi 00:12:12:12:12:12:12:00:01:03

===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
1         00:12:12:12:12:12:12:00:01:03   1             08/26/2021 07:41:04
-------------------------------------------------------------------------------
Number of entries: 1
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-VXLAN Dest TEP Info
===============================================================================
Instance  TEP Address                   Egr VNI             Last Change
-------------------------------------------------------------------------------
1         2001:db8::10:1                3                   08/26/2021 07:41:04
1         2001:db8::20:1                3                   08/26/2021 07:41:04
-------------------------------------------------------------------------------
Number of entries : 2
-------------------------------------------------------------------------------
===============================================================================
```

Debug

With debugging enabled for BGP updates, the following debug message on PE-3 shows that the NextHop value is changed in the EVPN-AD routes:

```
17 2021/08/26 07:40:54.081 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 85
    Flag: 0x90 Type: 14 Len: 48 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 16 Global NextHop 2001:db8::30:1
        Type: EVPN-AD Len: 25 RD: 192.0.2.3:3 ESI: 00:23:23:23:23:23:23:00:03:03,
                      tag: MAX-ET Label: 0 (Raw Label: 0x0) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:3
        esi-label:524285/All-Active
"
```

The following EVPN-ETH-SEG message on PE-3 shows that the NextHop value and Orig-IP-Addr is modified to the value 2001:db8::30:1.

```
20 2021/08/26 07:40:54.081 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 95
    Flag: 0x90 Type: 14 Len: 58 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 16 Global NextHop 2001:db8::30:1
        Type: EVPN-ETH-SEG Len: 35 RD: 192.0.2.3:0
         ESI: 00:23:23:23:23:23:23:00:03:03, IP-Len: 16 Orig-IP-Addr: 2001:db8::30:1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        df-election::DF-Type:Auto/DP:0/DF-Preference:0/AC:1
        target:23:23:23:23:23:23
"
```

Conclusion

All-active and single-active multi-homing can be configured for EVPN-VXLAN VPLSs. On all-active ESs, split-horizon for BUM traffic is based on local-bias, as described in RFC 8365.

EVPN R-VPLS Attached to IES

This chapter provides information about EVPN R-VPLS attached to IES.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written based on SR OS Release 16.0.R3, but the CLI configuration in the current edition corresponds to SR OS Release 23.10.R1.

Overview

R-VPLS services are often terminated on VPRN services. However, in some cases, R-VPLS services need to be terminated on IES services so that the traffic can be routed via the GRT. This is also supported for EVPN R-VPLS services.

Configuration

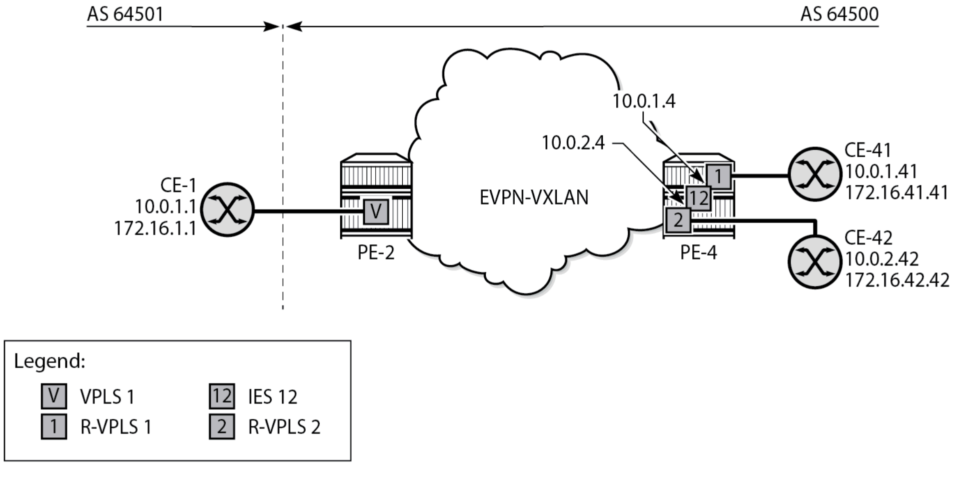
In this section, the following examples are configured:

- EVPN-VXLAN R-VPLS attached to IES without multi-homing
- EVPN-MPLS R-VPLS attached to IES with all-active and single-active multi-homing

EVPN-VXLAN R-VPLS attached to IES

Figure 114: EVPN-VXLAN R-VPLS attached to IES shows the example topology with EVPN-VXLAN configured on PE-2 and PE-4 and EVPN-VXLAN R-VPLSs 1 and 2 attached to IES 12 on PE-4.

Figure 114: EVPN-VXLAN R-VPLS attached to IES



CE-1 is in Autonomous System (AS) 64501 and the other nodes are in AS 64500.

The initial configuration includes the following:

- Cards, MDAs, ports
- Router interfaces
- IS-IS between PE-2 and PE-4

Configuration on PE-2

On PE-2, BGP is configured for the EVPN address family, as follows:

```
# on PE-2:
configure
    router Base
        autonomous-system 64500
        bgp
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal-evpn"
                family evpn
                type internal
                peer-as 64500
                neighbor 192.0.2.4
                exit
            exit
            no shutdown
        exit
```

EVPN-VXLAN VPLS 1 is an ordinary VPLS on PE-2, not an R-VPLS, and configured as follows. CE-1 is attached to SAP 1/1/c2/1:1 on PE-2.

```
# on PE-2:
configure
    service
```

```
        vpls 1 name "VPLS-1" customer 1 create
            vxlan instance 1 vni 1 create
            exit
            bgp
            exit
            bgp-evpn
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/1/c2/1:1 create
                no shutdown
            exit
            no shutdown
        exit
```

Configuration on PE-4

On PE-4, R-VPLS "evi-1" is configured as follows. CE-41 is attached to the SAP. The configuration of R-VPLS "evi-2" is similar.

```
# on PE-4:
configure
    service
        vpls 1 name "evi-1" customer 1 create
            description "EVPN-VXLAN R-VPLS 1"
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 1 create
            exit
            bgp
            exit
            bgp-evpn
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap pxc-1.a:1 create
                no shutdown
            exit
            no shutdown
        exit
```

Both R-VPLSs are attached to IES 12, which is configured as follows. Interface "int-evi-1" gets IP address 10.0.1.4/24 and interface "int-evi-2" gets IP address 10.0.2.4/24; these addresses are used as next-hop in default static routes on CE-1, CE-41, and CE-42.

```
# on PE-4:
configure
    service
        ies 12 name "IES-12" customer 1 create
            interface "int-evi-1" create
```

```
                address 10.0.1.4/24
                mac 00:00:00:00:01:04
                vpls "evi-1"
                exit
            exit
            interface "int-evi-2" create
                address 10.0.2.4/24
                mac 00:00:00:00:02:04
                vpls "evi-2"
                exit
            exit
            no shutdown
        exit
```

The BGP configuration on PE-4 includes an internal EVPN session with PE-2 (neighbor 192.0.2.2), an internal IPv4 session with CE-42 (neighbor 10.0.2.42), and an external IPv4 session with CE-1 (neighbor 10.0.1.1), as follows:

```
# on PE-4:
configure
    router Base
        bgp
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "external-ipv4"
                family ipv4
                type external
                local-as 64500
                peer-as 64501
                neighbor 10.0.1.1
                exit
            exit
            group "internal-evpn"
                family evpn
                type internal
                neighbor 192.0.2.2
                exit
            exit
            group "internal-ipv4"
                family ipv4
                type internal
                neighbor 10.0.2.42
                exit
            exit
            no shutdown
        exit
```

In this example, CE-41 is emulated as VPRN "CE-41" on PE-4. CE-41 is attached via port cross-connect (PXC) to R-VPLS "evi-1". The default static route has next-hop 10.0.1.4 on interface "int-evi-1" in IES 12. CE-41 has an EBGP-IPv4 session configured with neighbor CE-1 (10.0.1.1); CE-41 exports prefix 172.16.41.0/24 to CE-1. The configuration of VPRN "CE-41" on PE-4 is as follows:

```
# on PE-4:
configure
    service
        vprn 41 name "CE-41" customer 1 create
            description "CE-41 attached to R-VPLS evi-1 on PE-4"
            autonomous-system 64500
            interface "int-1_41" create
```

```
                address 10.0.1.41/24
                mac 00:00:00:00:01:41
                sap pxc-1.b:1 create
                exit
            exit
            interface "lo1" create
                address 172.16.41.41/24
                mac 00:00:00:04:41:41
                loopback
            exit
            static-route-entry 0.0.0.0/0
                next-hop 10.0.1.4
                    no shutdown
                exit
            exit
            bgp
                router-id 10.0.1.41
                enable-peer-tracking
                rapid-withdrawal
                split-horizon
                group "external"
                    family ipv4
                    type external
                    export "export-bgp-ipv4-41"
                    local-as 64500
                    peer-as 64501
                    neighbor 10.0.1.1
                    exit
                exit
            exit
            no shutdown
        exit
```

CE-42 is emulated as VPRN "CE-42" on PE-4. CE-42 is attached via PXC to R-VPLS "evi-2". The default static route has next-hop equal to 10.0.2.4 on interface "int-evi-2" in IES 12. An IBGP-IPv4 session is configured to this IES interface (neighbor 10.0.2.4). CE-42 exports prefix 172.16.42.0/24 to this IES interface on PE-4. The configuration of VPRN "CE-42" on PE-4 is as follows:

```
# on PE-4:
configure
    service
        vprn 42 name "CE-42" customer 1 create
            description "CE-42 attached to R-VPLS evi-2 on PE-4"
            autonomous-system 64500
            interface "int-2_42" create
                address 10.0.2.42/24
                mac 00:00:00:00:02:42
                sap pxc-1.b:2 create
                exit
            exit
            interface "int-test42" create
                address 172.16.42.42/24
                mac 00:00:00:04:42:42
                sap pxc-1.b:42 create
                exit
            exit
            static-route-entry 0.0.0.0/0
                next-hop 10.0.2.4
                    no shutdown
                exit
            exit
            bgp
                router-id 10.0.2.42
                enable-peer-tracking
                rapid-withdrawal
                split-horizon
                group "internal-ipv4"
                    family ipv4
                    type internal
                    export "export-bgp-ipv4-42"
                    neighbor 10.0.2.4
                    exit
                exit
                no shutdown
            exit
            no shutdown
        exit
```

The export policies are configured as follows:

```
# on PE-4:
configure
    router Base
        policy-options
            begin
            prefix-list "172.16.41.x"
                prefix 172.16.41.0/24 exact
            exit
            prefix-list "172.16.42.x"
                prefix 172.16.42.0/24 exact
            exit
            policy-statement "export-bgp-ipv4-41"
                entry 10
                    from
                        prefix-list "172.16.41.x"
                    exit
                    action accept
                    exit
                exit
            exit
            policy-statement "export-bgp-ipv4-42"
                entry 10
                    from
                        prefix-list "172.16.42.x"
                    exit
                    action accept
                    exit
                exit
            exit
            commit
        exit
```

Configuration on CE-1

On CE-1, the following static route is configured with next-hop 10.0.1. 4, which is the address on the interface "int-evi-1" in IES 12 on PE-4:

```
# on CE-1:
configure
    router Base
        static-route-entry 0.0.0.0/0
            next-hop 10.0.1.4
                no shutdown
            exit
        exit
```

The following loopback address is configured on CE-1 for test purposes:

```
# on CE-1:
configure
    router Base
        interface "lo1"
            address 172.16.1.1/24
            loopback
            no shutdown
        exit
```

On CE-1, EBGP-IPv4 sessions are configured to the IES interface "int-evi-1" on PE-4 (neighbor 10.0.1.4) and to CE-41 (neighbor 10.0.1.41) for the IPv4 address family. CE-1 exports prefix 172.16.1.0/24 to its peers. The BGP configuration is as follows:

```
# on CE-1:
configure
    router Base
        policy-options
            begin
            prefix-list "172.16.1.x"
                prefix 172.16.1.0/24 exact
            exit
            policy-statement "export-bgp-ipv4"
                entry 10
                    from
                        prefix-list "172.16.1.x"
                    exit
                    action accept
                    exit
                exit
            exit
            commit
        exit
        bgp
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            group "external"
                family ipv4
                type external
                export "export-bgp-ipv4"
                local-as 64501
                peer-as 64500
                neighbor 10.0.1.4
                exit
                neighbor 10.0.1.41
                exit
            exit
            no shutdown
        exit
```

Verification

On PE-4, the following shows that five BGP sessions are established:

- EBGP-IPv4 session with neighbor 10.0.1.1 (CE-1) from the base router

- IBGP-IPv4 session with neighbor 10.0.2.42 (CE-42) from the base router
- IBGP-EVPN session with neighbor 192.0.2.2 (PE-2) from the base router
- EBGP-IPv4 session with neighbor 10.0.1.1 (CE-1) from VPRN "CE-41"
- IBGP-IPv4 session to IES interface "int-evi-2" (10.0.2.4) from VPRN "CE-42"

Routes have been exchanged between the peers. The EBGP-IPv4 sessions are established using R-VPLS "evi-1".

```
*A:PE-4# show router bgp summary all

===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
Description
ServiceId          AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
10.0.1.1
Def. Inst       64501       8    0 00h01m07s 2/1/1 (IPv4)
                            8    0
10.0.2.42
Def. Inst       64500       7    0 00h01m15s 1/1/1 (IPv4)
                            8    0
192.0.2.2
Def. Inst       64500      11    0 00h02m07s 2/2/7 (Evpn)
                           15    0

10.0.1.1
41              64501       8    0 00h01m15s 2/1/1 (IPv4)
                            7    0
10.0.2.4
42              64500       7    0 00h01m15s 1/1/1 (IPv4)
                            7    0
-------------------------------------------------------------------------------
```

On PE-4, the following route table includes the prefixes 10.0.1.0/24 of interface "int-evi-1" and 10.0.2.0/24 of "int-evi-2" in IES 12. Also, it includes the remote prefixes 172.16.1.0/24 and 172.16.42.0, which are received as BGP IPv4 routes from CE-1 and CE-42.

```
*A:PE-4# show router route-table

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.1.0/24                                   Local   Local     00h03m23s  0
       int-evi-1                                                    0
10.0.2.0/24                                   Local   Local     00h03m23s  0
       int-evi-2                                                    0
172.16.1.0/24                                 Remote  BGP       00h02m15s  170
       10.0.1.1                                                     0
172.16.42.0/24                                Remote  BGP       00h02m20s  170
       10.0.2.42                                                    0
192.0.2.2/32                                  Remote  ISIS      00h04m02s  18
       192.168.24.1                                                 10
```

```
192.0.2.4/32                                  Local   Local     00h04m12s  0
       system                                                       0
192.168.24.0/30                               Local   Local     00h04m12s  0
       int-PE-4-PE-2                                                0
-------------------------------------------------------------------------------
No. of Routes: 7
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The following route table for CE-41 includes the remote prefix 172.16.1.0/24 received as BGP IPv4 route with next-hop 10.0.1.1. CE-1 and CE-41 are both in subnet 10.0.1.0/24.

```
*A:PE-4# show router service-name "CE-41" route-table

===============================================================================
Route Table (Service: 41)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
0.0.0.0/0                                     Remote  Static    00h03m23s  5
       10.0.1.4                                                     1
10.0.1.0/24                                   Local   Local     00h03m23s  0
       int-1_41                                                     0
172.16.1.0/24                                 Remote  BGP       00h02m20s  170
       10.0.1.1                                                     0
172.16.41.0/24                                Local   Local     00h03m23s  0
       lo1                                                          0
-------------------------------------------------------------------------------
No. of Routes: 4
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

Likewise, the following route table for CE-42 includes the remote prefix 172.16.1.0/24 received as BGP IPv4 route, but the next-hop is 10.0.2.4 instead of 10.0.1.1, because CE-42 is in subnet 10.0.2.0/24 whereas CE-1 is in subnet 10.0.1.0/24. Routing between the subnets 10.0.2.0/24 and 10.0.1.0/24 needs to be done in IES 12 on PE-4.

```
*A:PE-4# show router service-name "CE-42" route-table

===============================================================================
Route Table (Service: 42)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
0.0.0.0/0                                     Remote  Static    00h03m23s  5
       10.0.2.4                                                     1
10.0.2.0/24                                   Local   Local     00h03m23s  0
       int-2_42                                                     0
172.16.1.0/24                                 Remote  BGP       00h01m51s  170
       10.0.2.4                                                     1
172.16.42.0/24                                Local   Local     00h03m23s  0
       int-test42                                                   0
-------------------------------------------------------------------------------
No. of Routes: 4
```

```
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The following traceroute from CE-41 (172.16.41.41) to CE-1 (172.16.1.1) shows that no intermediate hops are required:

```
*A:PE-4# traceroute router-instance "CE-41" 172.16.1.1 source 172.16.41.41
traceroute to 172.16.1.1 from 172.16.41.41, 30 hops max, 40 byte packets
  1  172.16.1.1 (172.16.1.1)    4.76 ms  4.54 ms  4.39 ms
```

The following traceroute from CE-42 (172.16.42.42) to CE-1 (172.16.1.1) shows the IP address 10.0.2.4 on the interface "int-evi-2" in IES 12 as an intermediate hop:

```
*A:PE-4# traceroute router-instance "CE-42" 172.16.1.1 source 172.16.42.42
traceroute to 172.16.1.1 from 172.16.42.42, 30 hops max, 40 byte packets
  1  10.0.2.4 (10.0.2.4)    2.11 ms  2.74 ms  2.42 ms
  2  172.16.1.1 (172.16.1.1)    4.82 ms  4.92 ms  4.41 ms
```

The following ARP table on PE-4 includes entries for IP addresses in subnets 10.0.1.0/24 on interface "int-evi-1" and 10.0.2.0/24 on interface "int-evi-2":

```
*A:PE-4# show router arp

===============================================================================
ARP Table (Router: Base)
===============================================================================
IP Address      MAC Address       Expiry    Type   Interface
-------------------------------------------------------------------------------
192.0.2.4       00:04:fe:00:00:00 00h00m00s Oth    system
192.168.24.1    02:0e:01:01:00:01 03h53m51s Dyn[I] int-PE-4-PE-2
192.168.24.2    02:1a:01:01:00:0b 00h00m00s Oth[I] int-PE-4-PE-2
10.0.1.1        00:00:00:00:01:01 03h54m44s Dyn[I] int-evi-1
10.0.1.4        00:00:00:00:01:04 00h00m00s Oth[I] int-evi-1
10.0.1.41       00:00:00:00:01:41 03h59m30s Dyn[I] int-evi-1
10.0.2.4        00:00:00:00:02:04 00h00m00s Oth[I] int-evi-2
10.0.2.42       00:00:00:00:02:42 03h54m44s Dyn[I] int-evi-2
-------------------------------------------------------------------------------
No. of ARP Entries: 8
===============================================================================
```

The forwarding database (FDB) for R-VPLS 1 on PE-4 includes the MAC addresses corresponding to IP addresses 10.0.1.1, 10.0.1.4, and 10.0.1.41:

```
*A:PE-4# show service id "evi-1" fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:00:00:01:01 vxlan-1:                Evpn     11/08/23 07:24:14
                             192.0.2.2:1
1          00:00:00:00:01:04 cpm                     Intf     11/08/23 07:24:07
1          00:00:00:00:01:41 sap:pxc-1.a:1           LT/0     11/08/23 07:24:07
-------------------------------------------------------------------------------
No. of MAC Entries: 3
```

```
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

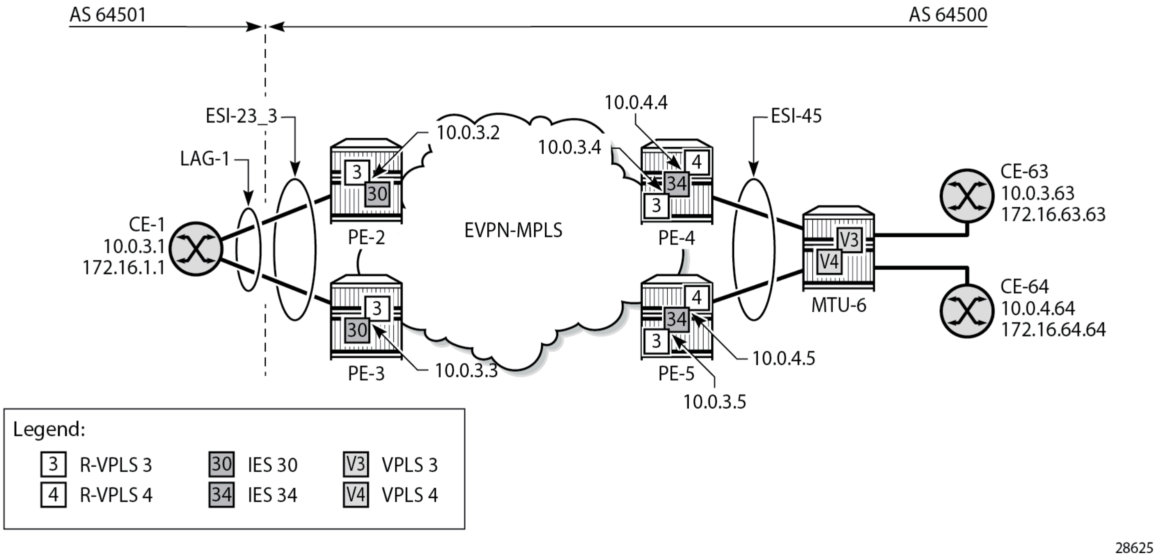
MAC address 00:00:00:00:01:01, which corresponds to IP address 10.0.1.1 on CE-1, is advertised in an EVPN MAC route by PE-2:

```
*A:PE-4# show router bgp routes evpn mac
===============================================================================
 BGP Router ID:192.0.2.4        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.2:1         00:00:00:00:01:01 ESI-0
      0                   Seq:0             VNI 1
                          n/a
                          192.0.2.2

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

EVPN-MPLS R-VPLS attached to IES

Figure 115: Example topology for EVPN-MPLS R-VPLS attached to IES shows the example topology for EVPN-MPLS R-VPLS attached to IES. All-active multi-homing (AA MH) is configured on PE-2 and PE-3, while single-active (SA) MH is configured on PE-4 and PE-5. R-VPLS "evi-3" is configured on all PEs. IES 30 is configured on PE-2 and PE-3, whereas IES 34 is configured on PE-4 and PE-5. On MTU-6, "VPLS-3" and "VPLS-4" are regular VPLSs, not routed.

Figure 115: Example topology for EVPN-MPLS R-VPLS attached to IES



The initial configuration on the nodes includes:

- Cards, MDAs, ports
- LAG "lag-1" on CE-1, PE-2, PE-3
- Router interfaces between the PEs and toward MTU-6
- IS-IS on these interfaces (alternatively, OSPF can be configured)
- LDP on these interfaces
- BGP configured for the EVPN address family on the PEs. PE-2 is the RR and has the following BGP configuration:

```
# on PE-2:
configure
    router Base
        bgp
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal-evpn"
                family evpn
                cluster 192.0.2.2
                peer-as 64500
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
            no shutdown
```

Configuration on PE-2 and PE-3

The service configuration on PE-2 and PE-3 is almost identical; only the IP address on the IES interface "int-evi-3" is different. The AA MH ES "ESI-23_3" is configured as follows, with LAG 1 and dot1q tag 3, so it is only applicable to VPLS "evi-3".

```
# on PE-2, PE-3:
configure
    service
        system
            bgp-evpn
                ethernet-segment "ESI-23_3" virtual create
                    esi 01:00:00:00:00:23:00:03:03:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 1
                    dot1q
                        q-tag-range 3
                    exit
                    no shutdown
                exit
```

R-VPLS "evi-3" has EVPN-MPLS enabled and is configured on PE-2 and PE-3, as follows. SAP lag-1:3 matches the configured LAG and the q-tag range for ESI-23_3.

```
# on PE-2, PE-3:
configure
    service
        vpls 3 name "evi-3" customer 1 create
            allow-ip-int-bind
            exit
            bgp
            exit
            bgp-evpn
                evi 3
                mpls bgp 1
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap lag-1:3 create
                no shutdown
            exit
            no shutdown
        exit
```

The following is the IES configuration on PE-2. In this example, IES 30 is only configured to demonstrate EVPN all-active multi-homing on R-VPLS with IES. If it were removed, everything still works and the connectivity between the CEs remains.

```
# on PE-2:
configure
    service
        ies 30 name "IES-30" customer 1 create
            interface "int-evi-3" create
                address 10.0.3.2/24
                mac 00:00:00:00:03:02
                vpls "evi-3"
                exit
            exit
            no shutdown
        exit
```

The IES configuration on PE-3 is similar, only using IP address 10.0.3.3/24.

Configuration on PE-4 and PE-5

On PE-4, SDP 46 is configured toward MTU-6. An SA MH ES "ESI-45" is configured using this SDP, as follows:

```
# on PE-4:
configure
    service
        sdp 46 mpls create
            far-end 192.0.2.6
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        system
            bgp-evpn
                ethernet-segment "ESI-45" create
                    esi 01:00:00:00:00:45:00:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 46
                    no shutdown
                exit
```

The configuration is similar on PE-5. SDP 56 is configured toward MTU-6 and ES "ESI-45" is configured with SDP 56 instead.

On PE-4, R-VPLSs "evi-3" and "evi-4" are configured with EVPN-MPLS, as follows:

```
# on PE-4:
configure
    service
        vpls 3 name "evi-3" customer 1 create
            description "EVPN-MPLS R-VPLS 3"
            allow-ip-int-bind
            exit
            bgp
            exit
            bgp-evpn
                evi 3
                mpls bgp 1
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            spoke-sdp 46:3 create
                no shutdown
            exit
            no shutdown
        exit
        vpls 4 name "evi-4" customer 1 create
            description "EVPN-MPLS R-VPLS 4"
            allow-ip-int-bind
            exit
            bgp
            exit
            bgp-evpn
                evi 4
                mpls bgp 1
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            spoke-sdp 46:4 create
                no shutdown
            exit
            no shutdown
        exit
```

The configuration is similar on PE-5; only the spoke-SDPs are different (spoke-SDP 56:3 and 56:4).

On PE-4, IES 34 is configured with interfaces "int-evi-3" and "int-evi-4", as follows. Passive VRRP is configured on both interfaces. With passive VRRP configured on both PE-4 and PE-5, both PEs behave as primary.

```
# on PE-4:
configure
    service
        ies 34 name "IES-34" customer 1 create
            interface "int-evi-3" create
                address 10.0.3.4/24
                mac 00:00:00:00:03:04
                vrrp 1 passive
                    backup 10.0.3.254
                    ping-reply
                    traceroute-reply
                exit
                vpls "evi-3"
                exit
            exit
            interface "int-evi-4" create
                address 10.0.4.4/24
                mac 00:00:00:00:04:04
                vrrp 1 passive
                    backup 10.0.4.254
                    ping-reply
                    traceroute-reply
                exit
                vpls "evi-4"
                exit
            exit
            no shutdown
        exit
```

The configuration of IES 34 is similar on PE-5, but the interface IP addresses are different: 10.0.3.5/24 and 10.0.4.5/24. The MAC addresses are also different.

To enable routing between CE-1 and CE-64 in a different subnet, BGP sessions are established with CE-1 (neighbor 10.0.3.1 in AS 64501) and CE-64 (neighbor 10.0.4.64 in AS 64500) for the IPv4 address family. The CEs export prefixes, but no export policy needs to be configured on PE-4 and PE-5. The BGP configuration on PE-4 is as follows:

```
# on PE-4:
configure
    router Base
        bgp
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "external"
                family ipv4
                type external
                local-as 64500
                peer-as 64501
                neighbor 10.0.3.1
                exit
            exit
            group "internal-evpn"
                family evpn
                type internal
                neighbor 192.0.2.2
                exit
            exit
            group "internal-ipv4"
                family ipv4
                peer-as 64500
                local-address 10.0.3.4
                neighbor 10.0.4.64
                exit
            exit
            no shutdown
        exit
```

The BGP configuration on PE-5 is almost identical; the local address is 10.0.3.5 instead.

Configuration on CE-1

The configuration on CE-1 includes the following:

- Router interface to VPLS "evi-3" (ESI-23_3) with IP address 10.0.3.1/24 and LAG-1:3 assigned to it
- Loopback interface with IP address 172.16.1.1/24 for test purposes
- Static default route with next-hop 10.0.3.254, which is the VRRP backup address for IES interface "int-evi-3" on PE-4 and PE-5
- Export policy to export prefix 172.16.1.0/24
- BGP sessions for the IPv4 address family toward PE-4 (10.0.3.4), PE-5 (10.0.3.5), and CE-63 (10.0.3.63)

The router configuration on CE-1 is as follows:

```
*A:CE-1>config>router# info
----------------------------------------------
#--------------------------------------------------
echo "Router (Network Side) Configuration"
#--------------------------------------------------
    router Base
        interface "int-CE-1-evi-3_ES-23"
            address 10.0.3.1/24
            port lag-1:3
            no shutdown
        exit
        interface "lo1"
            address 172.16.1.1/24
            loopback
            no shutdown
        exit
        interface "system"
            address 192.0.2.1/32
            no shutdown
        exit
        autonomous-system 64501
#--------------------------------------------------
echo "Static Route Configuration"
#--------------------------------------------------
        static-route-entry 0.0.0.0/0
            next-hop 10.0.3.254
                no shutdown
            exit
        exit
    exit

#--------------------------------------------------
echo "Policy Configuration"
#--------------------------------------------------
        policy-options
            begin
            prefix-list "172.16.1.x"
                prefix 172.16.1.0/24 exact
            exit
            policy-statement "export-bgp-ipv4"
                entry 10
                    from
                        prefix-list "172.16.1.x"
                    exit
                    action accept
                    exit
                exit
            exit
            commit
        exit
#--------------------------------------------------
echo "BGP Configuration"
#--------------------------------------------------
        bgp
            router-id 10.0.3.1
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "external"
                family ipv4
                type external
                export "export-bgp-ipv4"
                local-as 64501
                peer-as 64500
                neighbor 10.0.3.4
                exit
                neighbor 10.0.3.5
                exit
                neighbor 10.0.3.63
                exit
            exit
            no shutdown
        exit
----------------------------------------------
```

Configuration on MTU-6

The configuration on MTU-6 includes the following:

- Router interfaces
- IS-IS
- LDP
- One policy to export prefix 172.16.63.0/24 and another policy to export prefix 172.16.64.0/24
- BGP is not configured in the base router

The following service configuration on MTU-6 includes the SDP configuration and the VPLSs "VPLS-3" and "VPLS-4", which are not routed:

```
# on MTU-6:
configure
    service
        sdp 64 mpls create
            far-end 192.0.2.4
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        sdp 65 mpls create
            far-end 192.0.2.5
            ldp
```

```
                keep-alive
                    shutdown
                exit
                no shutdown
            exit
            vpls 3 name "VPLS-3" customer 1 create
                endpoint "CORE" create
                exit
                stp
                    shutdown
                exit
                sap pxc-1.a:3 create
                    no shutdown
                exit
                spoke-sdp 64:3 endpoint "CORE" create
                    stp
                        shutdown
                    exit
                    no shutdown
                exit
                spoke-sdp 65:3 endpoint "CORE" create
                    stp
                        shutdown
                    exit
                    no shutdown
                exit
                no shutdown
            exit
            vpls 4 name "VPLS-4" customer 1 create
                endpoint "CORE" create
                exit
                stp
                    shutdown
                exit
                sap pxc-1.a:4 create
                    no shutdown
                exit
                sap pxc-1.a:64 create
                    no shutdown
                exit
                spoke-sdp 64:4 endpoint "CORE" create
                    stp
                        shutdown
                    exit
                    no shutdown
                exit
                spoke-sdp 65:4 endpoint "CORE" create
                    stp
                        shutdown
                    exit
                    no shutdown
                exit
                no shutdown
            exit
```

In this example, CE-63 and CE-64 are simulated by VPRNs "CE-63" and "CE-64". The default static route has next-hop 10.0.3.254, which is the VRRP backup address on interface "int-evi-3" in IES 34 on both PE-4 and PE-5. BGP is configured within CE-63 and CE-64. The prefix 172.16.63.0/24 is exported by BGP in CE-63 and prefix 172.16.64.0/24 is exported by BGP in CE-64. The configuration of CE-63 and CE-64 is as follows:

```
# on MTU-6:
```

```
configure
    service
        vprn 63 name "CE-63" customer 1 create
            autonomous-system 64500
            interface "int-1_63" create
                address 10.0.3.63/24
                mac 00:00:00:00:03:63
                sap pxc-1.b:3 create
                exit
            exit
            interface "lo1" create
                address 172.16.63.63/24
                loopback
            exit
            static-route-entry 0.0.0.0/0
                next-hop 10.0.3.254
                    no shutdown
                exit
            exit
            bgp
                router-id 10.0.3.63
                enable-peer-tracking
                rapid-withdrawal
                split-horizon
                group "external"
                    family ipv4
                    type external
                    export "export-bgp-ipv4-63"
                    local-as 64500
                    peer-as 64501
                    neighbor 10.0.3.1
                    exit
                exit
                no shutdown
            exit
            no shutdown
        exit
        vprn 64 name "CE-64" customer 1 create
            autonomous-system 64500
            interface "int-2_64" create
                address 10.0.4.64/24
                mac 00:00:00:00:04:64
                sap pxc-1.b:4 create
                exit
            exit
            interface "int-test" create
                address 172.16.64.64/24
                mac 00:00:00:06:64:64
                sap pxc-1.b:64 create
                exit
            exit
            static-route-entry 0.0.0.0/0
                next-hop 10.0.4.254
                    no shutdown
                exit
            exit
            bgp
                router-id 10.0.4.64
                enable-peer-tracking
                rapid-withdrawal
                split-horizon
                group "internal-ipv4"
                    family ipv4
                    type internal
                    export "export-bgp-ipv4-64"
                    neighbor 10.0.3.4
                    exit
                    neighbor 10.0.3.5
                    exit
                exit
                no shutdown
            exit
            no shutdown
        exit
```

Verification

In the AA MH ES "ESI-23_3", PE-3 is the designated forwarder (DF) for R-VPLS "evi-3" and PE-2 is NDF, as follows:

```
*A:PE-2# show service id "evi-3" ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
lag-1:3               ESI-23_3                         NDF
===============================================================================
No sdp entries
No vxlan instance entries
```

```
*A:PE-3# show service id "evi-3" ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
lag-1:3               ESI-23_3                         DF
===============================================================================
No sdp entries
No vxlan instance entries
```

In the SA MH ES "ESI-45", PE-4 is NDF for R-VPLS "evi-3" and DF for R-VPLS "evi-4", as follows:

```
*A:PE-4# show service id "evi-3" ethernet-segment
No sap entries

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
46:3                  ESI-45                           NDF
===============================================================================
No vxlan instance entries
```

```
*A:PE-4# show service id "evi-4" ethernet-segment
No sap entries

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
46:4                  ESI-45                           DF
===============================================================================
No vxlan instance entries
```

The reverse is true for PE-5, which is DF for R-VPLS "evi-3" and NDF for R-VPLS "evi-4", as follows:

```
*A:PE-5# show service id "evi-3" ethernet-segment
No sap entries

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
56:3                  ESI-45                           DF
===============================================================================
No vxlan instance entries
```

```
*A:PE-5# show service id "evi-4" ethernet-segment
No sap entries

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
56:4                  ESI-45                           NDF
===============================================================================
No vxlan instance entries
```

CE-63 (VPRN 63 on MTU-6) has an external BGP IPv4 session with CE-1, whereas CE-64 (VPRN 64 on MTU-6) has internal BGP IPv4 sessions with IES interface "int-evi-3" on PE-4 and PE-5, as follows:

```
*A:MTU-6# show router service-name "CE-64" bgp summary all

===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
Description
ServiceId          AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------

10.0.3.1
63              64501      11    0 00h02m53s 2/1/1 (IPv4)
                           10    0
10.0.3.4
64              64500      10    0 00h02m57s 1/1/1 (IPv4)
                           10    0
10.0.3.5
64              64500      10    0 00h02m57s 1/0/1 (IPv4)
                           10    0
-------------------------------------------------------------------------------
```

The difference is that CE-63 (with IP address 10.0.3.63) is in the same subnet as CE-1 (10.0.3.1), whereas CE-64 is not (10.0.4.64). Routing between these subnets can be done in IES 34 on PE-4 and PE-5. CE-63 exports prefix 172.16.63.0/24 directly to CE-1, whereas CE-64 exports prefix 172.16.64.0/24 to PE-4 and PE-5 instead, which will advertise prefix 172.16.64.0/24 to their BGP peer CE-1. The following route table on CE-1 shows BGP route 172.16.63.0/63 with next-hop 10.0.3.63 (CE-63) and BGP route 172.16.64.0/64 with next-hop 10.0.3.4 (interface "int-evi-3" on PE-4):

```
*A:CE-1# show router route-table

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
0.0.0.0/0                                     Remote  Static    00h44m17s  5
       10.0.3.254                                                   1
10.0.3.0/24                                   Local   Local     00h44m17s  0
       int-CE-1-evi-3_ES-23                                         0
172.16.1.0/24                                 Local   Local     00h53m46s  0
       lo1                                                          0
172.16.63.0/24                                Remote  BGP       00h02m16s  170
       10.0.3.63                                                    0
172.16.64.0/24                                Remote  BGP       00h02m18s  170
       10.0.3.4                                                     0
192.0.2.1/32                                  Local   Local     00h53m46s  0
       system                                                       0
-------------------------------------------------------------------------------
No. of Routes: 6
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

In IES 34 on PE-4 (and PE-5), routing can be done between subnet 10.0.3.0/24 and 10.0.4.0/24. The following route table on PE-4 shows BGP route 172.16.1.0/24 with next-hop CE-1 (10.0.3.1) and BGP route 172.16.64.0/24 with next-hop CE-64 (10.0.4.64). The same entries occur in the route table on PE-5.

```
*A:PE-4# show router route-table

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.3.0/24                                   Local   Local     00h38m56s  0
       int-evi-3                                                    0
10.0.4.0/24                                   Local   Local     00h38m56s  0
       int-evi-4                                                    0
172.16.1.0/24                                 Remote  BGP       00h02m15s  170
       10.0.3.1                                                     0
172.16.64.0/24                                Remote  BGP       00h32m25s  170
       10.0.4.64                                                    0
---snip---
```

The route table of CE-63 (VPRN 63 on MTU-6) shows a BGP route for prefix 172.16.1.0/24 with next-hop 10.0.3.1 (CE-1), as follows:

```
*A:MTU-6# show router service-name "CE-63" route-table protocol bgp

===============================================================================
Route Table (Service: 63)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
172.16.1.0/24                                 Remote  BGP       00h03m49s  170
       10.0.3.1                                                     0
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The route table of CE-64 (VPRN 64 on MTU-6) shows a BGP route for prefix 172.16.1.0/24 with next-hop 10.0.4.254 (VRRP backup address for IES interface "int-evi-4" on PE-4 and PE-5), as follows:

```
*A:MTU-6# show router service-name "CE-64" route-table

===============================================================================
Route Table (Service: 64)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
0.0.0.0/0                                     Remote  Static    00h39m39s  5
       10.0.4.254                                                   1
10.0.4.0/24                                   Local   Local     00h39m39s  0
       int-2_64                                                     0
172.16.1.0/24                                 Remote  BGP       00h03m30s  170
       10.0.4.254                                                   1
172.16.64.0/24                                Local   Local     00h39m39s  0
       int-test                                                     0
-------------------------------------------------------------------------------
No. of Routes: 4
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The connectivity between CE-1 and CE-63 is verified as follows:

```
*A:CE-1# ping 172.16.63.63 source 172.16.1.1
PING 172.16.63.63 56 data bytes
64 bytes from 172.16.63.63: icmp_seq=1 ttl=64 time=2.96ms.
64 bytes from 172.16.63.63: icmp_seq=2 ttl=64 time=3.10ms.
---snip---
```

The following traceroute command verifies the connectivity between CE-1 and CE-64. The intermediate hop is 10.0.3.4, the IP address of the IES interface "int-evi-3" on PE-4:

```
*A:CE-1# traceroute 172.16.64.64 source 172.16.1.1
```

```
traceroute to 172.16.64.64 from 172.16.1.1, 30 hops max, 40 byte packets
  1  10.0.3.4 (10.0.3.4)    2.16 ms  2.29 ms  2.16 ms
  2  172.16.64.64 (172.16.64.64)    3.17 ms  3.36 ms  3.31 ms
```

When the traceroute is launched from CE-64, the intermediate hop is 10.0.4.4, the IP address of the IES interface "int-evi-4" on PE-4:

```
*A:MTU-6# traceroute router-instance "CE-64" 172.16.1.1
traceroute to 172.16.1.1, 30 hops max, 40 byte packets
  1  10.0.4.4 (10.0.4.4)    1.90 ms  2.08 ms  2.25 ms
  2  172.16.1.1 (172.16.1.1)    2.86 ms  2.92 ms  2.89 ms
```

The following ARP table on CE-1 contains entries for different nodes in the 10.0.3.0/24 subnet:

```
*A:CE-1# show router arp

===============================================================================
ARP Table (Router: Base)
===============================================================================
IP Address      MAC Address       Expiry    Type   Interface
-------------------------------------------------------------------------------
192.0.2.1       00:01:fe:00:00:00 00h00m00s Oth    system
10.0.3.1        00:01:fe:00:01:41 00h00m00s Oth[I] int-CE-1-evi-3_ES-23
10.0.3.4        00:00:00:00:03:04 03h26m35s Dyn[I] int-CE-1-evi-3_ES-23
10.0.3.5        00:00:00:00:03:05 03h26m51s Dyn[I] int-CE-1-evi-3_ES-23
10.0.3.63       00:00:00:00:03:63 03h40m39s Dyn[I] int-CE-1-evi-3_ES-23
10.0.3.254      00:00:5e:00:01:01 03h30m03s Dyn[I] int-CE-1-evi-3_ES-23
172.16.1.1      00:01:fe:00:00:00 00h00m00s Oth    lo1
-------------------------------------------------------------------------------
No. of ARP Entries: 7
===============================================================================
```

The ARP table on PE-4 contains entries for different nodes in subnets 10.0.3.0/24 and 10.0.4.0/24:

```
*A:PE-4# show router arp

===============================================================================
ARP Table (Router: Base)
===============================================================================
IP Address      MAC Address       Expiry    Type   Interface
-------------------------------------------------------------------------------
---snip---
10.0.3.1        00:01:fe:00:01:41 03h30m06s Dyn[I] int-evi-3
10.0.3.2        00:00:00:00:03:02 00h00m00s Evp[I] int-evi-3
10.0.3.3        00:00:00:00:03:03 00h00m00s Evp[I] int-evi-3
10.0.3.4        00:00:00:00:03:04 00h00m00s Oth[I] int-evi-3
10.0.3.5        00:00:00:00:03:05 00h00m00s Evp[I] int-evi-3
10.0.3.63       00:00:00:00:03:63 03h40m44s Dyn[I] int-evi-3
10.0.3.254      00:00:5e:00:01:01 00h00m00s Oth[I] int-evi-3
10.0.4.4        00:00:00:00:04:04 00h00m00s Oth[I] int-evi-4
10.0.4.5        00:00:00:00:04:05 00h00m00s Evp[I] int-evi-4
10.0.4.64       00:00:00:00:04:64 03h40m44s Dyn[I] int-evi-4
10.0.4.254      00:00:5e:00:01:01 00h00m00s Oth[I] int-evi-4
---snip---
-------------------------------------------------------------------------------
```

The FDB on PE-4 shows that MAC address 00:00:00:00:04:64-corresponding to 10.0.4.64 on CE-64-is learned on SDP 46:6, as follows.

```
*A:PE-4# show service id "evi-4" fdb detail
```

```
===============================================================================
Forwarding Database, Service 4
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
4          00:00:00:00:04:04 cpm                     Intf     11/08/23 07:33:19
4          00:00:00:00:04:05 mpls-1:                 EvpnS:P  11/08/23 07:33:25
                             192.0.2.5:524280
           ldp:65539
4          00:00:00:00:04:64 sdp:46:4                LT/0     11/08/23 07:47:29
4          00:00:5e:00:01:01 cpm                     Intf     11/08/23 07:33:19
-------------------------------------------------------------------------------
No. of MAC Entries: 4
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

The FDB on PE-5 shows that MAC address 00:00:00:00:04:64 -corresponding to 10.0.4.64 on CE-64-is advertised as an EVPN MAC route with ESI "ESI-45", as follows:

```
*A:PE-5# show service id "evi-4" fdb detail

===============================================================================
Forwarding Database, Service 4
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
4          00:00:00:00:04:04 mpls-1:                 EvpnS:P  11/08/23 07:33:27
                             192.0.2.4:524280
           ldp:65538
4          00:00:00:00:04:05 cpm                     Intf     11/08/23 07:33:25
4          00:00:00:00:04:64 eES:                    Evpn     11/08/23 07:47:29
                             01:00:00:00:00:45:00:00:00:01
4          00:00:5e:00:01:01 cpm                     Intf     11/08/23 07:33:25
-------------------------------------------------------------------------------
No. of MAC Entries: 4
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

Conclusion

With EVPN R-VPLS attached to IES services, EVPN services are connected to the base router, so the traffic can be routed in the global routing table (GRT).

EVPN VPLS Services Using SRv6 Transport

This chapter provides information about EVPN VPLS using SRv6 transport.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

The information and configuration in this chapter are based on SR OS Release 24.3.R1.

EVPN VPLS services over SRv6 are supported on FP4-based platforms in SR OS Release 22.7.R1 and later. For FP platforms, both all-active and single-active multihoming modes in EVPN VPLS services over SRv6 are supported in SR OS Release 23.10.R1 and later.

For migration scenarios from EVPN MPLS to EVPN SRv6 or from EVPN VXLAN to EVPN SRv6, see the EVPN VPLS with MPLS to SRv6 or VXLAN to SRv6 Stitching chapter.

Overview

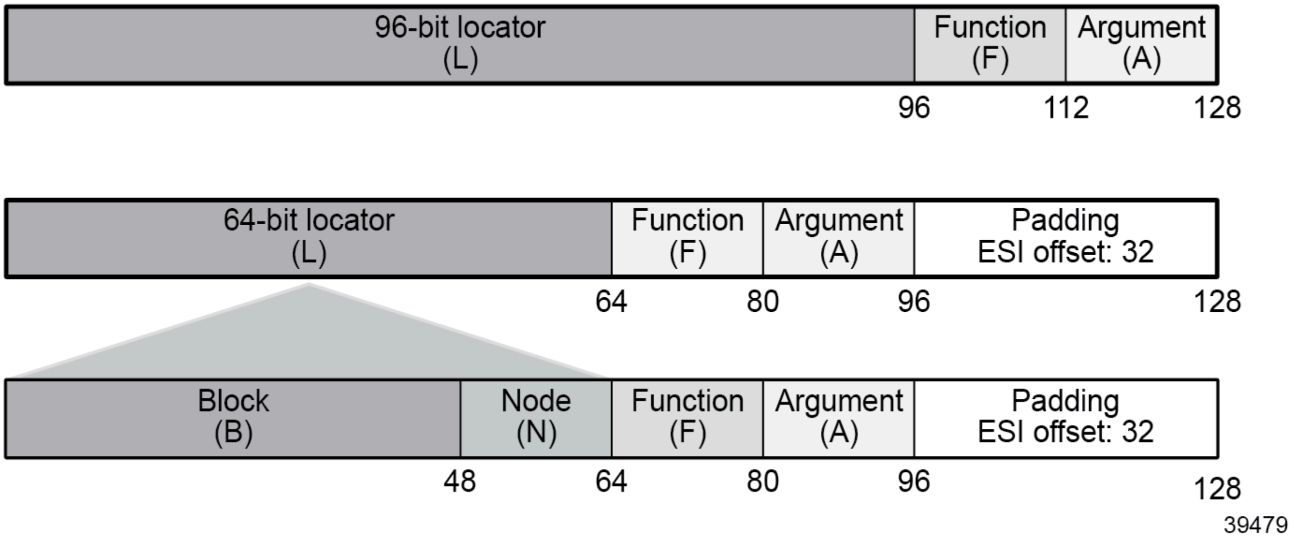
SRv6 locator and micro-segment locator

An EVPN VPLS service using SRv6 transport can be configured with a locator for segment identifiers (SIDs) or with a micro-segment locator for micro-SIDs (uSIDs). The micro-SID is an extension of the SRv6 architecture that allows for better efficiency. For micro-SIDs, multiple uSID instructions can be encoded within a single 128-bit SID address, with a maximum of six uSID instructions. Any SID in the destination address or segment routing header can be an SRv6 uSID carrier containing one or more uSIDs.

Locator for SID

Figure 116: SRv6 SID encoding shows two examples of SRv6 SID encoding: one with a 96-bit locator and one with a 64-bit locator. The 64-bit locator contains a 48-bit block address B and a 16-bit node-specific address N.

Figure 116: SRv6 SID encoding



The block length, function length, and argument length can be configured with the following command:

```
*A:PE-2>config>router>segment-routing>srv6>locator# ?
 [no] algorithm       - Configure a Flex Algorithm for segment routing v6
      argument-length - Configure segment argument length
 [no] block-length    - Configure block length for segment routing v6 locator
 [no] function-length - Configure function length for segment routing v6 locator
 [no] label-block     - Configure reserved label block for allocation of static and
                        dynamic labels for the segment routing v6 locator
      prefix          + Configure ipv6 prefix parameters for locator
 [no] shutdown        - Administratively disable or enable the locator
      static-function + Configure static function parameters for segment routing v6 locator
 [no] termination-fpe - Configure fpe for segment routing v6 locator
```

The only two possible values for the argument length are 0 (default) and 16, as follows:

```
*A:PE-2>config>router>segment-routing>srv6>locator# argument-length ?
  - argument-length {0|16}
```

The block length can have any value from 0 to 96, as follows:

```
*A:PE-2>config>router>segment-routing>srv6>locator# block-length ?
  - block-length <block-length>
  - no block-length

 <block-length>       : [0..96]
```

The function length can be 16 or any value from 20 to 96, as follows:

```
*A:PE-2>config>router>segment-routing>srv6>locator# function-length ?
  - function-length <function-length>
  - no function-length

 <function-length>    : 16 | [20..96]
```

For an EVPN VPLS service using SRv6 transport and configured with a locator, the egress PE signals the following functions to the ingress PE:

- End.DT2U for known unicast traffic, encoded in EVPN MAC/IP advertisement routes
- End.DT2M for BUM traffic, encoded in Inclusive Multicast Ethernet Tag (IMET) routes

In EVPN VPLS services, End.DT2U must be configured (with or without a static value) so that MAC/IP routes can be advertised. Similarly, End.DT2M must be configured (with or without a—different—static value) so that IMET routes can be advertised.

Micro-segment locator for micro-SID

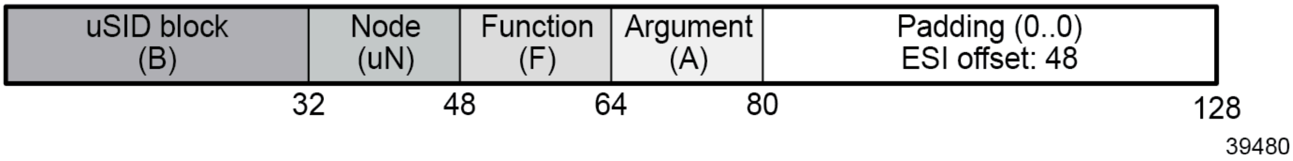
A maximum of six 16-bit micro-SID instructions can be encoded within a single 128-bit SID address, as follows:

<32-bit prefix>:<uSID1>:<uSID2>:<uSID3>:<uSID4>:<uSID5>:<uSID6>

The prefix length or block length is 32 in the preceding example, but can be any multiple of 8, with a maximum of 64. Any SRv6 instruction starting with this 32-bit prefix can contain up to six micro-instructions <uSIDx>. In the case that less than six uSIDs are used with this 32-bit prefix, the unused micro-instructions are set to 0x0000.

Figure 117: SRv6 micro-SID encoding shows an example of SRv6 micro-SID encoding with a 32-bit uSID prefix (block B), a 16-bit uSID identifying the node (uN), a 16-bit function, and optionally a 16-bit argument:

Figure 117: SRv6 micro-SID encoding



Micro-segments can be configured with the following command:

```
*A:PE-2>config>router>segment-routing>srv6# micro-segment ?
  - micro-segment
  - no micro-segment

      argument-length - Configure micro-segment argument length
 [no] block           + Configure micro-segment block
      block-length    - Configure micro-segment block length
      global-sid-ent* - Configure micro-segment global SID entries
      sid-length      - Configure micro-segment SID length
```

The argument length is either 0 or 16, as follows:

```
*A:PE-2>conf>router>sr>srv6>micro-segment# argument-length ?
  - argument-length {0|16}
```

For micro-segments, the block length must be a multiple of 8, with a maximum of 64, as follows:

```
*A:PE-2>conf>router>sr>srv6>micro-segment# block-length ?
```

```
  - block-length {8|16|24|32|40|48|56|64}
```

The micro-segment SID length can only be 16, as follows:

```
*A:PE-2>conf>router>sr>srv6>micro-segment# sid-length ?
  - sid-length {16}
```

In the case of a micro-segment locator, the egress PE signals the following functions to the ingress PE:

- End.uDT2U for known unicast traffic
- End.uDT2M for BUM traffic

EVPN VPLS services using SRv6 transport

An EVPN VPLS service using SRv6 can either be configured with a locator or a micro-segment locator, as follows:

```
*A:PE-2>config>service>vpls>srv6$ ?
 [no] locator         + Configure locator information
 [no] micro-segment-* + Configure micro segment locator information
```

The EVPN VPLS can be configured with an SRv6 locator, as follows:

```
*A:PE-2>config>service>vpls>srv6# locator "PE2-loc" ?
  - locator <name>
  - no locator <name>

 <name>               : [64 chars max]


      function        + Configure function value information
```

When a locator is configured, the possible functions are End.DT2U for known unicast traffic and End.DT2M for BUM traffic, as follows:

```
*A:PE-2>config>service>vpls>srv6>locator# function ?
  - function

 [no] end-dt2m        - Configure endpoint DT2M function value
 [no] end-dt2u        - Configure endpoint DT2U function value
```

When a micro-segment locator is configured, the functions are End.uDT2U for known unicast traffic and End.uDT2M for BUM traffic, as follows:

```
*A:PE-2>config>service>vpls>srv6# micro-segment-locator "PE2-mloc" function ?
  - function

 [no] udt2m           - Configure micro-segment DT2M function value
 [no] udt2u           - Configure micro-segment DT2U function value
```

In the **bgp-evpn segment-routing-v6** context, the following can be configured:

```
*A:PE-2>config>service>vpls>bgp-evpn# segment-routing-v6 ?
  - segment-routing-v6 [bgp <bgp-instance>] [srv6-instance <[1..1]>] [default-locator <name>]
    [create]
  - no segment-routing-v6 [bgp <bgp-instance>]
```

```
<bgp-instance>       : [1..2]
<name>               : [64 chars max]
<create>             : keyword


[no] default-route-* - Configure default-route-tag to match against export policies
     ecmp            - Configure maximum ECMP routes information
[no] evi-three-byte* - Enable/Disable evi-three-byte-auto-rt
[no] force-qinq-vc-* - Forces qinq-vc-type forwarding in the data-path
[no] force-vlan-vc-* - Forces vlan-vc-type forwarding in the data-path
     mh-mode         - Configure multi-homing mode
[no] oper-group      - Configure oper-group
     resolution      - Configure route resolution options
[no] restrict-prote* - Enable/disable protected src MAC restriction
     route-next-hop  - Configure route next-hop
[no] shutdown        - Enable/disable SRV6
[no] source-address  - Configure source IPv6 address
[no] split-horizon-* - Configure a split-horizon-group
```

The following parameters are specific to BGP-EVPN SRv6:

- The **default-route-tag** command is used to match BGP-EVPN routes for the service on export policies, for example, to add or modify BGP attributes.
- The **ecmp** command is used for aliasing on remote SRv6 EVPN Ethernet segment (ES) destinations.
- The **evi-three-byte-auto-rt** command is used to enable or disable 3-byte EVI auto-RT.
- The **force-qinq-vc-forwarding** command is used to preserve two VLAN tags in the SRv6 tunnel.
- The **force-vlan-vc-forwarding** command is used to preserve one VLAN tag in the SRv6 tunnel.
- The **mh-mode** option can be configured with the values access or network (needed in the case of multi-instance VPLS services).
- The **oper-group** command is required for fault propagation purposes.
- The **resolution {route-table|tunnel-table|fallback-tunnel-to-route-table}** command allows for setting the resolution of SRv6 routes in the route table or tunnel table (needed for SRv6 policies), and even a fallback from tunnel to route table resolution.
- The **restrict-protected-src discard-frame** command is required for loop avoidance along with **mac-duplication blackhole**.
- The **route-next-hop** command controls the BGP next hop used for service routes. The default is system IPv4 address.
- The **source-address** command does not need to be reachable or even exist on a local interface. This is possible because the source address is not looked up in the data path at the remote PE. If not configured, the source address is inherited from the locator's source address.
- The **split-horizon-group** command is used for a seamless integration with spoke SDPs and migration from the EVPN-MPLS services (multi-instance services).

AD per-ES routes in EVPN SRv6 multihoming

An EVPN VPLS service over SRv6 cannot be configured with all-active or single-active multihoming with the ESI label when the SRv6 locator has—the default—argument length 0. The following error is raised when attempting to configure a SAP with a LAG that is associated to an all-active ES in an EVPN VPLS service with SRv6 locator with argument length 0:

```
*A:PE-2>config>service>vpls# sap lag-1:3 create
MINOR: SVCMGR #6044 Cannot create sap - Associated Srv6 Svc 3 has locators with no argument
 length configured
```

The arg.fe2 argument is advertised along with the AD per-ES routes for all-active or single-active multihoming with the ESI label. The arg.fe2 argument is used along with the End.DT2M or End.uDT2M functions and supports the transposition into the ESI label extended community label field.

The arg.fe2 argument is dynamically allocated and encoded as follows:

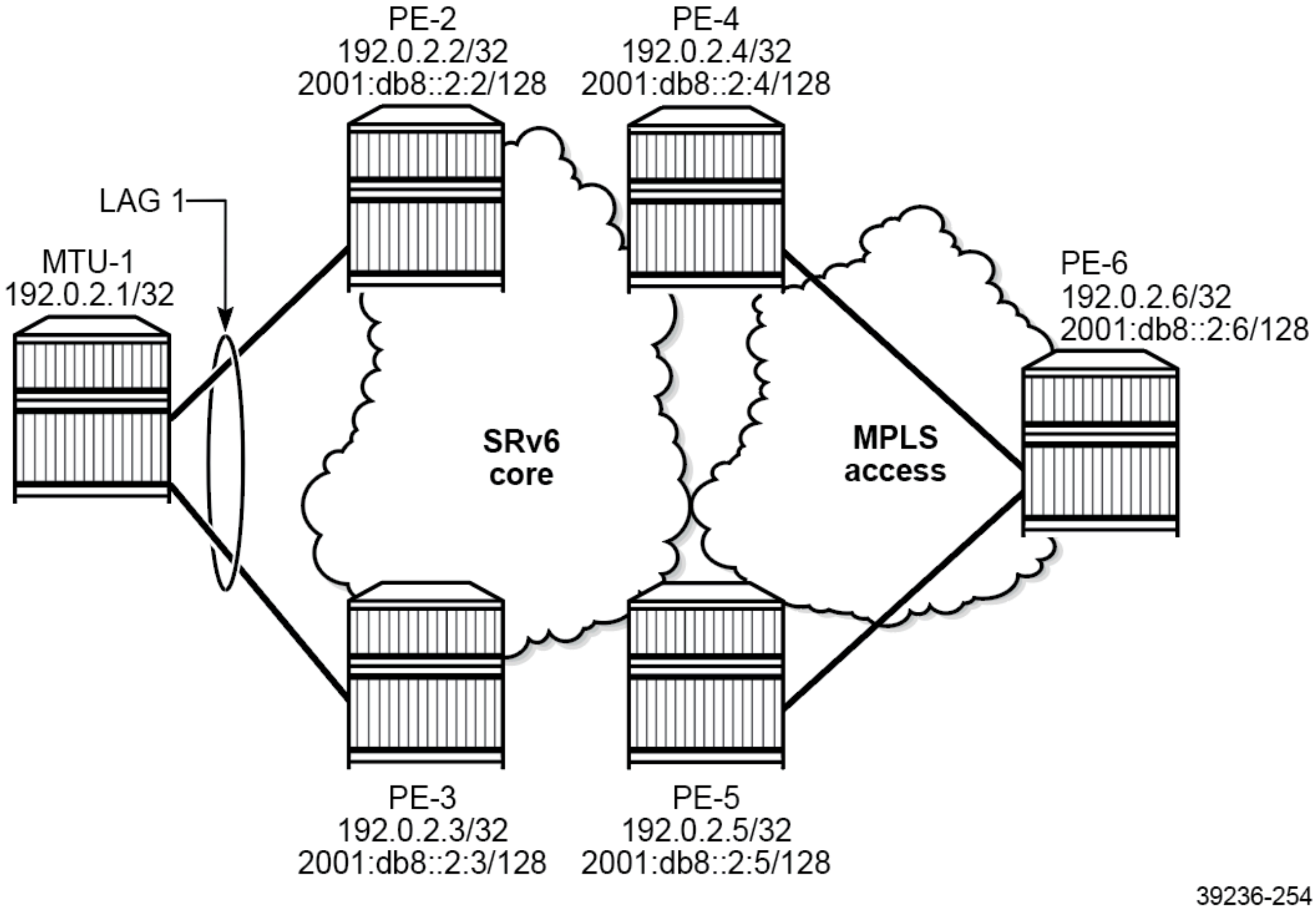
- The allocated arg.fe2 value is encoded in the high-order 16 bits of the ESI label field.
- The SRv6 SID value is always 0.
- The SID structure advertised in the SRv6 SID sub-sub-TLV follows the length of the locator associated with the VPLS (block length, node length, function length).

Note: If the ES is configured as **multi-homing single-active no-esi-label**, the AD per-ES routes are advertised with ESI label 3 (for an implicit-null label) and without the arg.fe2 argument.

Configuration

The Figure 118: Example topology shows the topology with six SR OS nodes:

Figure 118: Example topology



The initial configuration includes:

- cards, MDAs, ports
- LAG 1 between MTU-1, PE-2, and PE-3
- router interfaces
- IS-IS on the router interfaces
- BGP for the EVPN address family on PE-2, PE-3, PE-4, and PE-5

As an example, the BGP configuration on PE-2 is as follows:

```
# on PE-2:
configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
```

```
                group "internal"
                    family evpn
                    peer-as 64500
                    neighbor 2001:db8::2:3
                    exit
                    neighbor 2001:db8::2:4
                    exit
                    neighbor 2001:db8::2:5
                    exit
                exit
            exit
```

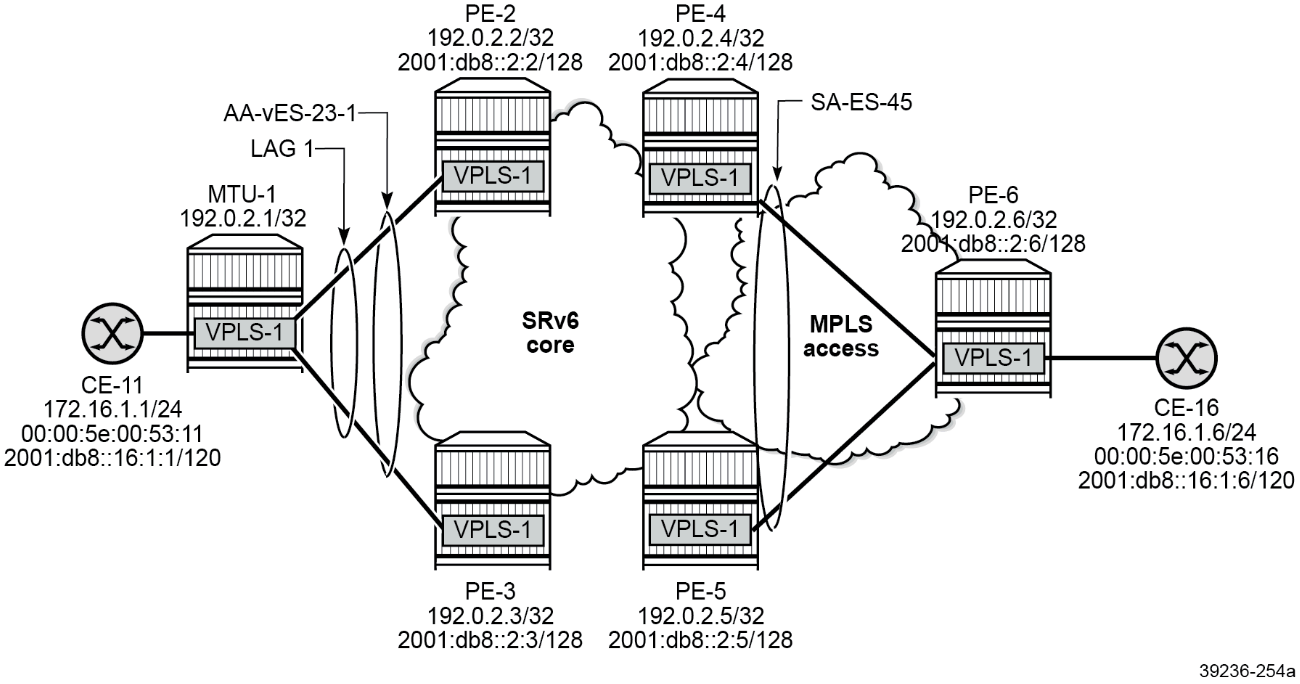
Two use cases are described in the following sections:

- EVPN VPLS using SRv6 with locator
- EVPN VPLS using SRv6 with micro-segment locator

EVPN VPLS using SRv6 with locator

The core PEs in Figure 119: Example topology with VPLS-1 are connected through an SRv6 network. All-active multihoming applies between PE-2 and PE-3, while single-active multihoming applies between PE-4 and PE-5.

Figure 119: Example topology with VPLS-1



SRv6 configuration with locator

SRv6 is configured on the core PEs. The configuration on PE-2 is as follows:

```
# on PE-2:
configure
    card 1
        mda 1
            xconnect
                mac 1 create
                    loopback 1 create
                    exit
                    loopback 2 create
                    exit
                exit
            exit
            no shutdown
        exit
    exit
    port-xc
        pxc 1 create
            port 1/1/m1/1
            no shutdown
        exit
        pxc 2 create
            port 1/1/m1/2
            no shutdown
        exit
    exit
    port pxc-1.a
        ethernet
        exit
        no shutdown
    exit
    port pxc-1.b
        ethernet
        exit
        no shutdown
    exit
    port pxc-2.a
        ethernet
        exit
        no shutdown
    exit
    port pxc-2.b
        ethernet
        exit
        no shutdown
    exit
    port 1/1/m1/1
        no shutdown
    exit
    port 1/1/m1/2
        no shutdown
    exit
    fwd-path-ext
        fpe 1 create
            path pxc 1
            srv6 origination
                interface-a
                exit
                interface-b
                exit
            exit
        exit
        fpe 2 create
            path pxc 2
            srv6 termination
                interface-a
                exit
                interface-b
                exit
            exit
        exit
    exit
    router Base
---snip---
        mpls-labels
            reserved-label-block "srv6-labels"          # optional
                start-label 20000 end-label 20999
            exit
        exit
        segment-routing
            segment-routing-v6
                origination-fpe 1
                source-address 2001:db8::2:2
                locator "PE2-loc"
                    block-length 48
                    label-block "srv6-labels"
                    function-length 16
                    termination-fpe 2
                    prefix
                        ip-prefix 2001:db8:aaaa:102::/64
                    exit
                    static-function
                    exit
                    argument-length 16
                    no shutdown
                exit
                base-routing-instance
                    locator "PE2-loc"
                        function
                            end-x-auto-allocate srh-mode psp protection unprotected
                            end 1
                                srh-mode usp
                            exit
                        exit
                    exit
                exit
            exit
        exit
        isis 0
            area-id 49.0001
            traffic-engineering
            traffic-engineering-options
                ipv6
                application-link-attributes
                exit
            exit
            advertise-passive-only
            advertise-router-capability as
            ipv6-routing native
            level 1
                wide-metrics-only
            exit
            level 2
```

```
                wide-metrics-only
            exit
            segment-routing-v6
                locator "PE2-loc"
                    level 1
                    exit
                    level 2
                    exit
                exit
                no shutdown
            exit
            interface "system"
                passive
            exit
            interface "int-PE-2-PE-3"
                level-capability level-2
                interface-type point-to-point
            exit
            interface "int-PE-2-PE-4"
                level-capability level-2
                interface-type point-to-point
            exit
            interface "int-PE-2-PE-5"
                level-capability level-2
                interface-type point-to-point
            exit
            no shutdown
```

The configuration on the other PEs is similar. On PE-5, no reserved MPLS label block is configured and the labels will be allocated by the system from the dynamic label range. The function length of the locator "PE5-loc" is 24 and the argument length is 16.

EVPN VPLS service configuration

VPLS-1 is configured on all nodes. The configuration on MTU-1 is as follows:

```
# on MTU-1:
configure
    service
        vpls 1 name "VPLS-1" customer 1 create
            stp
                shutdown
            exit
            sap 1/1/c10/1:1 create
                description "SAP to CE-11"
                no shutdown
            exit
            sap lag-1:1 create
                no shutdown
            exit
            no shutdown
```

On PE-2 and PE-3, all-active multihoming is configured, as follows:

```
# on PE-2:
configure
    service
        system
            bgp-evpn
                ethernet-segment "AA-vES-23-1" virtual create
```

```
                    esi 01:00:00:00:00:23:00:00:01:01
                    es-orig-ip 2001:db8::2:2          # on PE-3: 2001:db8::2:3
                    route-next-hop 2001:db8::2:2      # on PE-3: 2001:db8::2:3
                    es-activation-timer 3
                    service-carving
                        mode manual
                        manual
                            preference non-revertive create
                                value 100            # on PE-3: value 150
                            exit
                        exit
                    exit
                    multi-homing all-active
                    lag 1
                    dot1q
                        q-tag-range 1 to 3
                    exit
                    no shutdown
                exit
            exit
        exit
        vpls 1 name "VPLS-1" customer 1 create
            segment-routing-v6 1 create
                locator "PE2-loc"                    # on PE-3: "PE3-loc"
                    function
                        end-dt2u    # label dynamically allocated; label block
                        end-dt2m    # label dynamically allocated; label block
                    exit
                exit
            exit
            bgp
            exit
            bgp-evpn
                evi 1
                segment-routing-v6 bgp 1 srv6-instance 1 default-locator "PE2-loc" create
                    source-address 2001:db8::2:2     # on PE-3: "PE3-loc" - 2001:db8::2:3
                    ecmp 2
                    route-next-hop 2001:db8::2:2     # on PE-3: "PE3-loc" - 2001:db8::2:3
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap lag-1:1 create
                no shutdown
            exit
            no shutdown
```

On PE-4, the function values are manually configured with a value of 10 for End.DT2U and a value of 11 for End.DT2M. The locator "PE4-loc" refers to the MPLS label block and the number of static functions is 15. Single-active multihoming applies between PE-4 and PE-5. The corresponding SDPs use LDP IPv6 tunnels, as follows:

```
# on PE-4:
configure
    router Base
        segment-routing
             segment-routing-v6
                locator "PE4-loc"
                    block-length 48
                    label-block "srv6-labels"
                    function-length 16
                    termination-fpe 2
                    prefix
                        ip-prefix 2001:db8:aaaa:104::/64
                    exit
                    static-function
                        max-entries 15
                    exit
                    argument-length 16
                    no shutdown
                exit
            exit
        exit
    exit
    service
        sdp 46 mpls create
            far-end 2001:db8::2:6
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        system
            bgp-evpn
                ethernet-segment "SA-ES-45" create
                    esi 01:00:00:00:00:45:00:00:00:01
                    es-orig-ip 2001:db8::2:4
                    route-next-hop 2001:db8::2:4
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 46
                    no shutdown
                exit
            exit
        exit
        vpls 1 name "VPLS-1" customer 1 create
            segment-routing-v6 1 create
                locator "PE4-loc"
                    function
                        end-dt2u 10     # configured function value
                        end-dt2m 11     # configured function value
                    exit
                exit
            exit
            bgp
            exit
            bgp-evpn
                evi 1
                segment-routing-v6 bgp 1 srv6-instance 1 default-locator "PE4-loc" create
                    source-address 2001:db8::2:4
                    ecmp 2
                    route-next-hop 2001:db8::2:4
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            spoke-sdp 46:1 create
                no shutdown
            exit
            no shutdown
```

The configuration on PE-5 does not include an MPLS label block and the function values are dynamically allocated from the dynamic label range, as follows:

```
# on PE-5:
configure
    router Base
        segment-routing
            segment-routing-v6
                origination-fpe 1
                source-address 2001:db8::2:5
                locator "PE5-loc"
                    block-length 48
                    function-length 24
                    termination-fpe 2
                    prefix
                        ip-prefix 2001:db8:aaaa:105::/64
                    exit
                    static-function
                    exit
                    argument-length 16
                    no shutdown
                exit
                base-routing-instance
                    locator "PE5-loc"
                        function
                            end-x-auto-allocate srh-mode psp protection unprotected
                            end 1
                                srh-mode usp
                            exit
                        exit
                    exit
                exit
            exit
        exit
    exit
    service
        sdp 56 mpls create
            far-end 2001:db8::2:6
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        system
            bgp-evpn
                ethernet-segment "SA-ES-45" create
                    esi 01:00:00:00:00:45:00:00:00:01
                    es-orig-ip 2001:db8::2:5
                    route-next-hop 2001:db8::2:5
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 56
                    no shutdown
                exit
            exit
        exit
        vpls 1 name "VPLS-1" customer 1 create
```

```
            segment-routing-v6 1 create
                locator "PE5-loc"
                    function
                        end-dt2u        # dynamic label; no label block defined
                        end-dt2m        # dynamic label; no label block defined
                    exit
                exit
            exit
            bgp
            exit
            bgp-evpn
                evi 1
                segment-routing-v6 bgp 1 srv6-instance 1 default-locator "PE5-loc" create
                    source-address 2001:db8::2:5
                    ecmp 2
                    route-next-hop 2001:db8::2:5
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            spoke-sdp 56:1 create
                no shutdown
            exit
            no shutdown
```

The service configuration on MTU-6 is as follows:

```
# on MTU-6:
configure
    service
        sdp 64 mpls create
            far-end 2001:db8::2:4
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        sdp 65 mpls create
            far-end 2001:db8::2:5
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        vpls 1 name "VPLS-1" customer 1 create
            endpoint "CORE" create
            exit
            stp
                shutdown
            exit
            sap 1/1/c10/1:1 create
                no shutdown
            exit
            spoke-sdp 64:1 endpoint "CORE" create
                stp
                    shutdown
                exit
                no shutdown
            exit
            spoke-sdp 65:1 endpoint "CORE" create
                stp
                    shutdown
                exit
                no shutdown
            exit
            no shutdown
```

Verification

The following command shows the configured BGP-EVPN SRv6 parameters:

```
*A:PE-2# show service id "VPLS-1" bgp-evpn segment-routing-v6

===============================================================================
BGP EVPN Segment Routing v6 Information
===============================================================================
Admin State             : Enabled                 Bgp Instance  : 1
Srv6 Instance           : 1
Default Locator         : PE2-loc

Oper Group              : (none)
Default Route Tag       : 0x0
Source Address          : 2001:db8::2:2
ECMP                    : 2
Force Vlan VC Fwd       : Disabled
Next Hop Type           : explicit
Next Hop Address        : 2001:db8::2:2
Evi 3-byte Auto-RT      : disabled
Route Resolution        : route-table
Force QinQ VC Fwd       : none
MH Mode                 : network
Rest Prot Src Mac       : disabled
Split Horizon Group     : n/a
===============================================================================
```

The following command shows the SID values and the status of the End.DT2U and End.DT2M functions in SRv6 instance 1 in VPLS-1 on PE-3:

```
*A:PE-3# show service id "VPLS-1" segment-routing-v6 instance 1

===============================================================================
Segment Routing v6 Instance 1 Service 1
===============================================================================
Locator
 Type          Function  SID                                     Status
-------------------------------------------------------------------------------
PE3-loc
  End.DT2U     *3        2001:db8:aaaa:103:3::                   ok
  End.DT2M     *4        2001:db8:aaaa:103:4::                   ok
===============================================================================
Legend: * - System allocated
```

The SID 2001:db8:aaaa:103:4:: corresponding to the End.DT2M function is one of the SRv6 destinations for VPLS-1 on PE-2, as follows:

```
*A:PE-2# show service id "VPLS-1" segment-routing-v6 destinations

===============================================================================
TEP, SID (Instance 1)
===============================================================================
TEP Address                Segment Id                          Oper  Mcast Num
                                                               State       MACs
-------------------------------------------------------------------------------
2001:db8::2:3              2001:db8:aaaa:103:4::               Up    BUM   0
2001:db8::2:4              2001:db8:aaaa:104:b::               Up    BUM   0
2001:db8::2:5              2001:db8:aaaa:105:7ff:f800::        Up    BUM   0
-------------------------------------------------------------------------------
Number of TEP, SID: 3
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
Segment Routing v6 Ethernet Segment Dest (Instance 1)
===============================================================================
Eth SegId                                 Num. Macs     Last Update
-------------------------------------------------------------------------------
01:00:00:00:00:45:00:00:00:01             1             03/18/2024 15:42:31
-------------------------------------------------------------------------------
Number of entries: 1
-------------------------------------------------------------------------------
===============================================================================
```

The two other SRv6 destinations for VPLS-1 on PE-2 correspond to the End.DT2M function in VPLS-1 on PE-4 and PE-5.

Multihoming, route tables, FDBs

PE-3 acts as the designated forwarder (DF) for VPLS-1 in the all-active ES "AA-vES-23-1", as follows:

```
*A:PE-3# show service id "VPLS-1" ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
lag-1:1               AA-vES-23-1                      DF
===============================================================================
No sdp entries
No vxlan instance entries
```

PE-2 is an NDF in the all-active ES "AA-vES-23-1", as follows:

```
*A:PE-2# show service id "VPLS-1" ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
lag-1:1               AA-vES-23-1                      NDF
===============================================================================
No sdp entries
No vxlan instance entries
```

PE-5 acts as the DF for VPLS-1 in the single-active ES "SA-ES-45", as follows:

```
*A:PE-5# show service id "VPLS-1" ethernet-segment
No sap entries

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
56:1                  SA-ES-45                         DF
===============================================================================
No vxlan instance entries
```

PE-4 is an NDF in the single-active ES "SA-ES-45", as follows:

```
*A:PE-4# show service id "VPLS-1" ethernet-segment
No sap entries

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
46:1                  SA-ES-45                         NDF
===============================================================================
No vxlan instance entries
```

The following route table for IPv6 shows that PE-2 has SRv6-ISIS-tunneled routes to the locator prefixes on PE-3, PE-4, and PE-5:

```
*A:PE-2# show router route-table ipv6 next-hop-type tunneled

===============================================================================
IPv6 Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
2001:db8:aaaa:103::/64                        Remote  ISIS      00h07m51s  18
       2001:db8:aaaa:103::/64 (tunneled:SRV6-ISIS)                  10
2001:db8:aaaa:104::/64                        Remote  ISIS      00h07m43s  18
       2001:db8:aaaa:104::/64 (tunneled:SRV6-ISIS)                  10
2001:db8:aaaa:105::/64                        Remote  ISIS      00h07m21s  18
       2001:db8:aaaa:105::/64 (tunneled:SRV6-ISIS)                  10
-------------------------------------------------------------------------------
No. of Routes: 3
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

After traffic has been sent between CE-11 and CE-16, the VPLS-1 FDBs are populated. The FDB on MTU-1 shows that CE-16 can be reached through the LAG, while CE-11 can be reached via SAP 1/1/c10/1:1, as follows:

```
*A:MTU-1# show service id "VPLS-1" fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:5e:00:53:11 sap:1/1/c10/1:1         L/120    03/18/24 15:27:04
1          00:00:5e:00:53:16 sap:lag-1:1             L/120    03/18/24 15:28:01
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

The VPLS-1 FDBs on PE-2 and PE-3 are similar. CE-11 can be reached through the LAG, while CE-16 can be reached via the single-active ES between PE-4 and PE-5, as follows:

```
*A:PE-2# show service id "VPLS-1" fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:5e:00:53:11 sap:lag-1:1             LT/90    03/18/24 15:28:01
1          00:00:5e:00:53:16 eES:                    Evpn     03/18/24 15:28:01
                             01:00:00:00:00:45:00:00:00:01
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

```
*A:PE-3# show service id "VPLS-1" fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:5e:00:53:11 sap:lag-1:1             Evpn     03/18/24 15:28:01
1          00:00:5e:00:53:16 eES:                    Evpn     03/18/24 15:28:01
                             01:00:00:00:00:45:00:00:00:01
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

The FDB on NDF PE-4 shows that CE-11 can be reached via the all-active ES between PE-2 and PE-3, while CE-16 can be reached via the single-active ES between PE-4 and PE-5, but not through the SDP toward MTU-6, as follows:

```
*A:PE-4# show service id "VPLS-1" fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
```

```
-------------------------------------------------------------------------------
1          00:00:5e:00:53:11 eES:                    Evpn     03/18/24 15:28:01
                             01:00:00:00:00:23:00:00:01:01
1          00:00:5e:00:53:16 eES:                    Evpn     03/18/24 15:28:01
                             01:00:00:00:00:45:00:00:00:01
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

The FDB on DF PE-5 shows that CE-11 can be reached via the all-active ES between PE-2 and PE-3, while CE-16 can be reached via the SDP toward MTU-6, as follows:

```
*A:PE-5# show service id "VPLS-1" fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:5e:00:53:11 eES:                    Evpn     03/18/24 15:28:01
                             01:00:00:00:00:23:00:00:01:01
1          00:00:5e:00:53:16 sdp:56:1                LT/90    03/18/24 15:28:01
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

The FDB on MTU-6 shows that CE-11 can be reached via the SDP toward DF PE-5, while CE-16 can be reached via SAP 1/1/c10/1:1, as follows:

```
*A:MTU-6# show service id "VPLS-1" fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:5e:00:53:11 sdp:65:1                L/0      03/18/24 15:28:01
1          00:00:5e:00:53:16 sap:1/1/c10/1:1         L/0      03/18/24 15:27:45
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

BGP-EVPN routes

PE-2 received the following EVPN-MAC route from PE-5 with the endpoint behavior 0x17 = 23, which corresponds to the End.DT2U function for known unicast traffic:

```
# on PE-2:
40 2024/03/18 15:28:00.714 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:5
"Peer 1: 2001:db8::2:5: UPDATE
Peer 1: 2001:db8::2:5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 125
    Flag: 0x90 Type: 14 Len: 56 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 16 Global NextHop 2001:db8::2:5
        Type: EVPN-MAC Len: 33 RD: 192.0.2.5:1 ESI: 01:00:00:00:00:45:00:00:00:01,
                       tag: 0, mac len: 48 mac: 00:00:5e:00:53:16, IP len: 0,
                       IP: NULL, label1: 8388496 (Raw Label: 0x7fff90)
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64500:1
    Flag: 0xc0 Type: 40 Len: 37 Prefix-SID-attr:
       SRv6 Services TLV (37 bytes):-
           Type: SRV6 L2 Service TLV (6)
           Length: 34 bytes, Reserved: 0x0
         SRv6 Service Information Sub-TLV (33 bytes)
             Type: 1 Len: 30 Rsvd1: 0x0
             SRv6 SID: 2001:db8:aaaa:105::
             SID Flags: 0x0 Endpoint Behavior: 0x17 Rsvd2: 0x0
             SRv6 SID Sub-Sub-TLV
                Type: 1 Len: 6
                BL:48 NL:16 FL:24 AL:0 TL:20 TO:68
"
```

The following **show** command for the EVPN-MAC route containing the CE-16 MAC address shows the behavior End.DT2U (23):

```
*A:PE-2# show router bgp routes evpn mac mac-address 00:00:5e:00:53:16 detail
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Original Attributes

Network        : n/a
Nexthop        : 2001:db8::2:5
Path Id        : None
From           : 2001:db8::2:5
Res. Nexthop   : fe80::1e:1ff:fe01:15
Local Pref.    : 100                    Interface Name : int-PE-2-PE-5
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64500:1
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.5
Origin         : IGP
Flags          : Used Valid Best
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : MAC
ESI            : 01:00:00:00:00:45:00:00:00:01
Tag            : 0
```

```
IP Address     : n/a
Route Dist.    : 192.0.2.5:1
Mac Address    : 00:00:5e:00:53:16
MPLS Label1    : 524281                MPLS Label2    : n/a
Route Tag      : 0
Neighbor-AS    : n/a
DB Orig Val    : N/A                   Final Orig Val : N/A
Source Class   : 0                     Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h06m13s
SRv6 TLV Type  : SRv6 L2 Service TLV (6)
SRv6 SubTLV    : SRv6 SID Information (1)
Sid            : 2001:db8:aaaa:105::
Full Sid       : 2001:db8:aaaa:105:7ff:f900::
Behavior       : End.DT2U (23)
SRv6 SubSubTLV : SRv6 SID Structure (1)
Loc-Block-Len  : 48                    Loc-Node-Len   : 16
Func-Len       : 24                    Arg-Len        : 0
Tpose-Len      : 20                    Tpose-offset   : 68
---snip---
```

PE-2 received the following EVPN IMET route from PE-5 with the endpoint behavior 0x18 = 24, which corresponds to the End.DT2M function for BUM traffic:

```
25 2024/03/18 15:27:44.001 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:5
"Peer 1: 2001:db8::2:5: UPDATE
Peer 1: 2001:db8::2:5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 145
    Flag: 0x90 Type: 14 Len: 52 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 16 Global NextHop 2001:db8::2:5
        Type: EVPN-INCL-MCAST Len: 29 RD: 192.0.2.5:1, tag: 0,
                              orig_addr len: 128, orig_addr: 2001:db8::2:5
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64500:1
    Flag: 0xc0 Type: 22 Len: 21 PMSI:
        Tunnel-type Ingress Replication (6)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf: not required]
        MPLS Label 8388480
        Tunnel-Endpoint 2001:db8::2:5
    Flag: 0xc0 Type: 40 Len: 37 Prefix-SID-attr:
       SRv6 Services TLV (37 bytes):-
           Type: SRV6 L2 Service TLV (6)
           Length: 34 bytes, Reserved: 0x0
         SRv6 Service Information Sub-TLV (33 bytes)
             Type: 1 Len: 30 Rsvd1: 0x0
             SRv6 SID: 2001:db8:aaaa:105::
             SID Flags: 0x0 Endpoint Behavior: 0x18 Rsvd2: 0x0
             SRv6 SID Sub-Sub-TLV
                Type: 1 Len: 6
                BL:48 NL:16 FL:24 AL:16 TL:20 TO:68
"
```

The following **show** command for the EVPN IMET route that is received from PE-5 shows the behavior End.DT2M (24):

```
*A:PE-2# show router bgp routes evpn incl-mcast originator-ip 2001:db8::2:5 hunt
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : n/a
Nexthop        : 2001:db8::2:5
Path Id        : None
From           : 2001:db8::2:5
Res. Nexthop   : fe80::1e:1ff:fe01:15
Local Pref.    : 100                    Interface Name : int-PE-2-PE-5
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64500:1
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.5
Origin         : IGP
Flags          : Used Valid Best
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : INCL-MCAST
Tag            : 0
Originator IP  : 2001:db8::2:5
Route Dist.    : 192.0.2.5:1
Route Tag      : 0
Neighbor-AS    : n/a
DB Orig Val    : N/A                    Final Orig Val : N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h07m17s
SRv6 TLV Type  : SRv6 L2 Service TLV (6)
SRv6 SubTLV    : SRv6 SID Information (1)
Sid            : 2001:db8:aaaa:105::
Full Sid       : 2001:db8:aaaa:105:7ff:f800::
Behavior       : End.DT2M (24)
SRv6 SubSubTLV : SRv6 SID Structure (1)
Loc-Block-Len  : 48                     Loc-Node-Len   : 16
Func-Len       : 24                     Arg-Len        : 16
Tpose-Len      : 20                     Tpose-offset   : 68
-------------------------------------------------------------------------------
PMSI Tunnel Attributes :
Tunnel-type    : Ingress Replication
Flags          : Type: RNVE(0) BM: 0 U: 0 Leaf: not required
MPLS Label     : 8388480
Tunnel-Endpoint: 2001:db8::2:5
-------------------------------------------------------------------------------


-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

PE-2 receives the following AD per-EVI route from PE-5:

```
28 2024/03/18 15:27:44.001 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:5
"Peer 1: 2001:db8::2:5: UPDATE
Peer 1: 2001:db8::2:5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 117
    Flag: 0x90 Type: 14 Len: 48 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 16 Global NextHop 2001:db8::2:5
        Type: EVPN-AD Len: 25 RD: 192.0.2.5:1 ESI: 01:00:00:00:00:45:00:00:00:01,
                      tag: 0 Label: 8388560 (Raw Label: 0x7fffd0) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64500:1
    Flag: 0xc0 Type: 40 Len: 37 Prefix-SID-attr:
       SRv6 Services TLV (37 bytes):-
           Type: SRV6 L2 Service TLV (6)
           Length: 34 bytes, Reserved: 0x0
         SRv6 Service Information Sub-TLV (33 bytes)
             Type: 1 Len: 30 Rsvd1: 0x0
             SRv6 SID: 2001:db8:aaaa:105::
             SID Flags: 0x0 Endpoint Behavior: 0x17 Rsvd2: 0x0
             SRv6 SID Sub-Sub-TLV
                Type: 1 Len: 6
                BL:48 NL:16 FL:24 AL:0 TL:20 TO:68
"
```

PE-2 receives the following AD per-ES route from PE-5:

```
31 2024/03/18 15:27:44.001 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:5
"Peer 1: 2001:db8::2:5: UPDATE
Peer 1: 2001:db8::2:5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 125
    Flag: 0x90 Type: 14 Len: 48 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 16 Global NextHop 2001:db8::2:5
        Type: EVPN-AD Len: 25 RD: 192.0.2.5:1 ESI: 01:00:00:00:00:45:00:00:00:01,
                      tag: MAX-ET Label: 0 (Raw Label: 0x0) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:1
        esi-label:16/Single-Active
    Flag: 0xc0 Type: 40 Len: 37 Prefix-SID-attr:
       SRv6 Services TLV (37 bytes):-
           Type: SRV6 L2 Service TLV (6)
           Length: 34 bytes, Reserved: 0x0
         SRv6 Service Information Sub-TLV (33 bytes)
             Type: 1 Len: 30 Rsvd1: 0x0
             SRv6 SID: ::
             SID Flags: 0x0 Endpoint Behavior: 0x18 Rsvd2: 0x0
             SRv6 SID Sub-Sub-TLV
                Type: 1 Len: 6
                BL:48 NL:16 FL:24 AL:16 TL:16 TO:88
"
```

ES "SA-ES-45" is configured in a single-active multihoming mode with the ESI label; therefore, PE-5 sends the AD per-ES routes that carry the arg.fe2 value. The SID structure in the SRv6 sub-sub-TLV indicates that the argument length is 16 bits. The 16-bit length of the arg.fe2 value is transposed into the ESI label extended community label field. The transposition length is 16 (argument length 16) and the transposition offset is 88 (block length 48 + node length 16 + function length 24). The SRv6 SID is :: (0).

End.DT2U and End.DT2M functions

On PE-4, function value 10 is configured for End.DT2U and function value 11 for End.DT2M, as follows:

```
*A:PE-4# show router segment-routing-v6 local-sid end-dt2m end-dt2u

===============================================================================
Segment Routing v6 Local SIDs
===============================================================================
SID                                               Type           Function
  Locator
  Context
-------------------------------------------------------------------------------
2001:db8:aaaa:104:a::                             End.DT2U       10
  PE4-loc
  SvcId: 1 Name: VPLS-1
2001:db8:aaaa:104:b::                             End.DT2M       11
  PE4-loc
  SvcId: 1 Name: VPLS-1
-------------------------------------------------------------------------------
SIDs : 2
-------------------------------------------------------------------------------
===============================================================================
```

On PE-5, the function values 524280 and 524281 are dynamically allocated from the dynamic label range, as follows:

```
*A:PE-5# show router segment-routing-v6 local-sid end-dt2m end-dt2u

===============================================================================
Segment Routing v6 Local SIDs
===============================================================================
SID                                               Type           Function
  Locator
  Context
-------------------------------------------------------------------------------
2001:db8:aaaa:105:7ff:f800::                      End.DT2M       524280
  PE5-loc
  SvcId: 1 Name: VPLS-1
2001:db8:aaaa:105:7ff:f900::                      End.DT2U       524281
  PE5-loc
  SvcId: 1 Name: VPLS-1
-------------------------------------------------------------------------------
SIDs : 2
-------------------------------------------------------------------------------
===============================================================================
```

On PE-2, a reserved label block is configured and associated to the SRv6 locator. Function value 3 is dynamically allocated for End.DT2U and function value 4 for End.DT2M. SID 2001:db8:aaaa:102:3:: is used for the End.DT2U function and SID 2001:db8:aaaa:102:4:: for the End.DT2M function, as follows:

```
*A:PE-2# show service id "VPLS-1" segment-routing-v6 detail
```

```
===============================================================================
Segment Routing v6 Instance 1 Service 1
===============================================================================
Locator
 Type          Function  SID                                     Status
-------------------------------------------------------------------------------
PE2-loc
  End.DT2U     *3        2001:db8:aaaa:102:3::                   ok
  End.DT2M     *4        2001:db8:aaaa:102:4::                   ok
===============================================================================
Legend: * - System allocated
```

On PE-2, the following MPLS label blocks are defined, where the reserved label block "srv6-labels" is manually configured:

```
*A:PE-4# show router mpls-labels label-range

===============================================================================
Label Ranges
===============================================================================
Label Type      Start Label End Label   Aging       Available   Total
-------------------------------------------------------------------------------
Static          32          18431       -           18400       18400
Dynamic         18432       524287      0           504848      505856
    Seg-Route   0           0           -           0           0

-------------------------------------------------------------------------------
Reserved Label Blocks
-------------------------------------------------------------------------------
Reserved Label                               Start       End         Total
Block Name                                   Label       Label
-------------------------------------------------------------------------------
srv6-labels                                  20000       20999       1000
-------------------------------------------------------------------------------
No. of Reserved Label Blocks: 1
-------------------------------------------------------------------------------
===============================================================================
```

On PE-5, no reserved label block is configured, as follows:

```
*A:PE-5# show router mpls-labels label-range

===============================================================================
Label Ranges
===============================================================================
Label Type      Start Label End Label   Aging       Available   Total
-------------------------------------------------------------------------------
Static          32          18431       -           18400       18400
Dynamic         18432       524287      0           505846      505856
    Seg-Route   0           0           -           0           0
===============================================================================
```

The dynamically allocated function labels 524280 and 524281 are taken from the dynamic MPLS label range:

```
*A:PE-5# show service id "VPLS-1" segment-routing-v6 detail

===============================================================================
Segment Routing v6 Instance 1 Service 1
===============================================================================
Locator
 Type          Function  SID                                     Status
-------------------------------------------------------------------------------
PE5-loc
  End.DT2U     *524281   2001:db8:aaaa:105:7ff:f900::            ok
  End.DT2M     *524280   2001:db8:aaaa:105:7ff:f800::            ok
===============================================================================
Legend: * - System allocated
```

In the preceding output, SID 2001:db8:aaaa:105:7ff:f800:: is used for BUM traffic, as indicated by the End.DT2M function. This SID is advertised by PE-5 in an IMET route to PE-2 and appears in the list of SRv6 destinations in VPLS-1 on PE-2, as follows:

```
*A:PE-2# show service id "VPLS-1" segment-routing-v6 destinations

===============================================================================
TEP, SID (Instance 1)
===============================================================================
TEP Address                Segment Id                          Oper  Mcast Num
                                                               State       MACs
-------------------------------------------------------------------------------
2001:db8::2:3              2001:db8:aaaa:103:4::               Up    BUM   0
2001:db8::2:4              2001:db8:aaaa:104:b::               Up    BUM   0
2001:db8::2:5              2001:db8:aaaa:105:7ff:f800::        Up    BUM   0
-------------------------------------------------------------------------------
Number of TEP, SID: 3
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
Segment Routing v6 Ethernet Segment Dest (Instance 1)
===============================================================================
Eth SegId                                 Num. Macs     Last Update
-------------------------------------------------------------------------------
01:00:00:00:00:45:00:00:00:01             1             03/18/2024 15:28:01
-------------------------------------------------------------------------------
Number of entries: 1
-------------------------------------------------------------------------------
===============================================================================
```

Ping or traceroute can be sent to the entire SIDs, including the arguments. The ping or traceroute messages are replied to, ignoring these arguments. As an example, PE-2 launches the following commands to some of the SRv6 SIDs (without arguments) in the preceding output:

```
*A:PE-2# ping 2001:db8:aaaa:104:b:: rapid
PING 2001:db8:aaaa:104:b:: 56 data bytes
!!!!!
---- 2001:db8:aaaa:104:b:: PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 2.22ms, avg = 2.30ms, max = 2.48ms, stddev = 0.094ms

*A:PE-2# traceroute 2001:db8:aaaa:105:7ff:f800::
traceroute to 2001:db8:aaaa:105:7ff:f800::, 30 hops max, 60 byte packets
  1  2001:db8::2:5 (2001:db8::2:5)    3.14 ms  3.02 ms  2.66 ms
```

The following command shows that the SRv6 locator of PE-5 (2001:db8:aaaa:105::/64) is resolved:

```
*A:PE-2# show router bgp next-hop evpn 2001:db8::2:5 service-id 1 detail
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
===============================================================================
BGP VPN Next Hop
===============================================================================
---snip---
-------------------------------------------------------------------------------
VPN Next Hop          : 2001:db8::2:5
Autobind              : gre/rtm
Labels                : --
User-labels           : 1
Admin-tag-policy      : --
Strict-tunnel-tagging : N
Color                 : --
UPA Trigger Next Hop  : --
Locator               : 2001:db8:aaaa:105::/64
Created               : 00h11m35s
Last-modified         : 00h11m35s
-------------------------------------------------------------------------------
Resolving Prefix : 2001:db8::2:5/128
Preference       : 18                   Metric           : 10
Reference Count  : 3                    Owner            : GRE
Fib Programmed   : Y
Resolved Next Hop: fe80::1e:1ff:fe01:15
Egress Label     : n/a                  TunnelId         : 4294967293
Locator State    : Resolved
-------------------------------------------------------------------------------
---snip---
```

The following tunnel table on PE-2 shows the SRv6-ISIS tunnels to the SRv6 locators 2001:db8:aaaa:103::/64 on PE-3, 2001:db8:aaaa:104::/64 on PE-4, and 2001:db8:aaaa:105::/64 on PE-5:

```
*A:PE-2# show router tunnel-table ipv6 protocol srv6-isis

===============================================================================
IPv6 Tunnel Table (Router: Base)
===============================================================================
Destination                                     Owner     Encap TunnelId  Pref
Nexthop                                         Color           Metric
-------------------------------------------------------------------------------
2001:db8:aaaa:103::/64                          srv6-isis SRV6  524289    0
  fe80::14:1ff:fe01:1f-"int-PE-2-PE-3"                            10
2001:db8:aaaa:104::/64                          srv6-isis SRV6  524290    0
  fe80::1a:1ff:fe01:b-"int-PE-2-PE-4"                             10
2001:db8:aaaa:105::/64                          srv6-isis SRV6  524291    0
  fe80::1e:1ff:fe01:15-"int-PE-2-PE-5"                            10
-------------------------------------------------------------------------------
Flags: B = BGP or MPLS backup hop available
       L = Loop-Free Alternate (LFA) hop available
       E = Inactive best-external BGP route
       k = RIB-API or Forwarding Policy backup hop
===============================================================================
```

EVPN and VPLS integration

EVPN and VPLS integration (RFC 8560) is supported when SRv6 transport is used. SDP bindings signaled by TLDP (manually configured or BGP-AD) or BGP can coexist in VPLS services where EVPN SRv6 is enabled. The SR OS node allows for the creation of an EVPN destination and an SDP binding to the same far end, but the SDP binding is kept operationally down with a flag indicating an EVPN route conflict.

On PE-2 and PE-5, LDP SDPs are created, as follows:

```
# on PE-2:
configure
    router Base
        ldp
            interface-parameters
                interface "int-PE-2-PE-5" dual-stack    # on PE-5: "int-PE-5-PE-2"
                    ipv6
                        no shutdown
                    exit
                exit
            exit
        exit
    exit
    service
        sdp 25 mpls create                              # on PE-5: sdp 52
            far-end 2001:d8::2:5                        # on PE-5: 2001:db8::2:2
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        vpls "VPLS-1"
            spoke-sdp 25:1 create                      # on PE-5: spoke-sdp 52:1
            exit
        exit
```

PE-2 still has an SRv6-ISIS tunnel toward 2001:db8:aaaa:105::/64 on PE-5, as follows:

```
*A:PE-2# show router tunnel-table ipv6 protocol srv6-isis

===============================================================================
IPv6 Tunnel Table (Router: Base)
===============================================================================
Destination                                     Owner     Encap TunnelId  Pref
Nexthop                                         Color           Metric
-------------------------------------------------------------------------------
2001:db8:aaaa:103::/64                          srv6-isis SRV6  524289    0
  fe80::14:1ff:fe01:1f-"int-PE-2-PE-3"                            10
2001:db8:aaaa:104::/64                          srv6-isis SRV6  524290    0
  fe80::1a:1ff:fe01:b-"int-PE-2-PE-4"                             10
2001:db8:aaaa:105::/64                          srv6-isis SRV6  524291    0
  fe80::1e:1ff:fe01:15-"int-PE-2-PE-5"                            10
-------------------------------------------------------------------------------
Flags: B = BGP or MPLS backup hop available
       L = Loop-Free Alternate (LFA) hop available
       E = Inactive best-external BGP route
       k = RIB-API or Forwarding Policy backup hop
===============================================================================
```

The spoke SDP from PE-2 to PE-5 is operationally down and no egress label is allocated, as follows:

```
*A:PE-2# show service id "VPLS-1" sdp 25:1

===============================================================================
Service Destination Point (Sdp Id : 25:1)
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
25:1             Spok                     Up      Down      524284    None
```

```
                          2001:db8::2:5
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

The reason why the spoke SDP from PE-2 to PE-5 is operationally down is because of an EVPN route conflict, as follows:

```
*A:PE-2# show service id "VPLS-1" sdp 25:1 detail | match Flag context all
Flags              : NoEgrVCLabel
                     EvpnRouteConflict
```

Tools commands

The following command shows the EVPN usage statistics:

```
*A:PE-2# tools dump service evpn usage

vxlan-srv6-evpn-mpls usage statistics at 03/18/2024 15:42:08:

MPLS-TEP                                        :             0
VXLAN-TEP                                       :             0
SRV6-TEP                                        :             3
Total-TEP                                       :      3/ 16383

Mpls Dests (TEP, Egress Label + ES + ES-BMAC)   :             0
Mpls Etree Leaf Dests                           :             0
Vxlan Dests (TEP, Egress VNI + ES)              :             0
Srv6 Dests (TEP, SID  + ES)                     :             4
Total-Dest                                      :      4/196607

Sdp Bind +  Evpn Dests                          :      4/245759
ES L2/L3 PBR                                    :      0/ 32767
Evpn Etree Remote BUM Leaf Labels               :             0
```

```
*A:PE-2# tools dump router segment-routing-v6 usage
Segment Routing v6 Usage
Service SID index: 6/262128
```

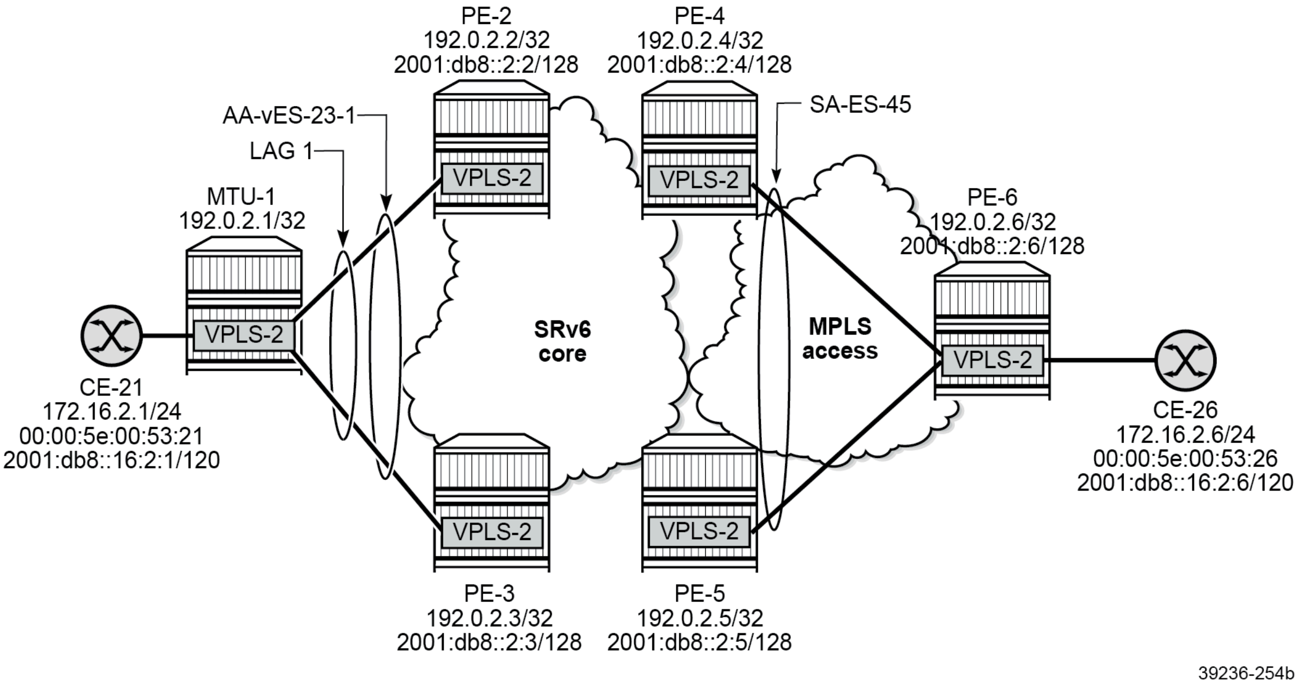
In the case of failure to instantiate an EVPN destination, the **tools dump service id "VPLS-1" srv6** command provides the following statistics:

```
*A:PE-2# tools dump service id "VPLS-1" srv6

TEP, Egress Bind Failure statistics at 03/18/2024 15:42:08:

statistics last cleared at 03/18/2024 15:20:23:

Failures: None
```

EVPN VPLS using SRv6 with micro-segment locator

Figure 120: Example topology with VPLS-2 shows the topology with VPLS-2. The same nodes and ESs are used, but the VPLS-2 service uses micro-SIDs, whereas VPLS-1 uses regular SIDs. The CEs are different as well.

Figure 120: Example topology with VPLS-2



SRv6 micro-segment configuration on core PEs

On all core PEs, a micro-segment and a micro-segment locator are configured on PE-2, as follows:

```
# on PE-2:
configure
    router Base
        mpls-labels
             reserved-label-block "res-block1"
                start-label 19000 end-label 19999
            exit
        exit
        segment-routing
            segment-routing-v6
                micro-segment
                    argument-length 16
                    block "PE2-ms-block1"  # PE3-ms-block1, PE4-ms-block1, PE5-ms-block1
                        label-block "res-block1"
                        termination-fpe 2
                        prefix
                            ip-prefix 2001:bbbb::/32    # on all PEs; block length 32
                        exit
                        static-function
                            max-entries 5
                        exit
                        no shutdown
                    exit
                exit
                micro-segment-locator "PE2-mloc"       # or PE3-mloc, PE4-mloc, PE5-mloc
                    block "PE2-ms-block1"
                    un
                        srh-mode usp
                        value 2         # on PE-3: value 3; PE-4: value 4; PE-5: value 5
                    exit
                    no shutdown
                exit
                base-routing-instance
                    micro-segment-locator "PE2-mloc"     # PE3-mloc, PE4-mloc, PE5-mloc
                        function
                            ua-auto-allocate srh-mode psp protection unprotected
                            ua 1
                                srh-mode usp
                            exit
                        exit
                    exit
                exit
            exit
        exit
        isis 0
            segment-routing-v6
                micro-segment-locator "PE2-mloc"       # or PE3-mloc, PE4-mloc, PE5-mloc
                    level 1
                    exit
                    level 2
                    exit
                exit
```

The micro-segment locator contains a micro-segment node value (**un value**) that must be unique network-wide. The **un** function is equivalent to the regular SRv6 End function, but it is configured in the **micro-segment-locator** context instead of the Base instance. The **un value** command creates a node identifier as an IPv6 address that is composed of the block part and followed by a 16-bit SID.

The **ua** micro-SID function encodes the behavior of an adjacency SID.

The following error message is raised when an IP prefix is configured with a length of /48 that is different from the block length of /32:

```
*A:PE-2>conf>router>sr>srv6>ms>block>prefix# ip-prefix 2001:db8:bbbb::/48
MINOR: SRV6 #1006 Not permitted - Prefix length must be equal to block-length 32
```

Service configuration

On MTU-1, VPLS-2 is configured as follows:

```
# on MTU-1:
configure
    service
        vpls 2 name "VPLS-2" customer 1 create
            stp
                shutdown
            exit
            sap 1/1/c10/1:2 create
```

```
                no shutdown
            exit
            sap lag-1:2 create
                no shutdown
            exit
            no shutdown
```

On PE-2 and PE-3, VPLS-2 is configured with all-active multihoming using ES "AA-vES-23-1". The micro-segment locator is configured with the End.uDT2U and End.uDT2M functions, as follows:

```
# on PE-2:
configure
    service
        vpls 2 name "VPLS-2" customer 1 create
            segment-routing-v6 1 create
                micro-segment-locator "PE2-mloc"            # on PE-3: PE3-mloc
                    function
                        udt2u
                        udt2m
                    exit
                exit
            exit
            bgp
            exit
            bgp-evpn
                evi 2
                segment-routing-v6 bgp 1 srv6-instance 1 default-locator "PE2-mloc" create
                    source-address 2001:db8::2:2      # on PE-3: PE3-mloc; 2001:db8::2:3
                    ecmp 2
                    route-next-hop 2001:db8::2:2      # on PE-3: PE3-mloc; 2001:db8::2:3
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap lag-1:2 create
                no shutdown
            exit
            no shutdown
```

The VPLS-2 configuration on PE-4 uses spoke SDP 46:2, while on PE-5, spoke SDP 56:2 is used, as follows:

```
# on PE-4:
configure
    service
        vpls 2 name "VPLS-2" customer 1 create
            segment-routing-v6 1 create
                micro-segment-locator "PE4-mloc"        # on PE-5: PE5-mloc
                    function
                        udt2u
                        udt2m
                    exit
                exit
            exit
            bgp
            exit
            bgp-evpn
                evi 2
                segment-routing-v6 bgp 1 srv6-instance 1 default-locator "PE4-mloc" create
                    source-address 2001:db8::2:4       # on PE-5: PE5-mloc; 2001:db8::2:5
                    ecmp 2
                    route-next-hop 2001:db8::2:4
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            spoke-sdp 46:2 create                      # on PE-5: spoke-sdp 56:2
                no shutdown
            exit
            no shutdown
```

The configuration for VPLS-2 on MTU-6 is as follows:

```
# on MTU-6:
configure
    service
        vpls 2 name "VPLS-2" customer 1 create
            endpoint "CORE" create
            exit
            stp
                shutdown
            exit
            sap 1/1/c10/1:2 create
                no shutdown
            exit
            spoke-sdp 64:2 endpoint "CORE" create
                stp
                    shutdown
                exit
                no shutdown
            exit
            spoke-sdp 65:2 endpoint "CORE" create
                stp
                    shutdown
                exit
                no shutdown
            exit
            no shutdown
```

Verification

PE-2 already had three SRv6-ISIS tunnels to the locators 2001:db8:aaaa:103::/64 on PE-3, 2001:db8:aaaa:104::/64 on PE-4, and 2001:db8:aaaa:105::/64 on PE-5. Now, PE-2 also has SRv6-ISIS tunnels to the micro-segment locators 2001:bbbb:3::/48 on PE-3, 2001:bbbb:4::/48 on PE-4, and 2001:bbbb:5::/48 on PE-5, as follows:

```
*A:PE-2# show router tunnel-table ipv6 protocol srv6-isis

===============================================================================
IPv6 Tunnel Table (Router: Base)
===============================================================================
Destination                                 Owner     Encap TunnelId  Pref
Nexthop                                     Color           Metric
-------------------------------------------------------------------------------
2001:db8:aaaa:103::/64                      srv6-isis SRV6  524289    0
  fe80::14:1ff:fe01:1f-"int-PE-2-PE-3"                        10
2001:db8:aaaa:104::/64                      srv6-isis SRV6  524290    0
  fe80::1a:1ff:fe01:b-"int-PE-2-PE-4"                         10
```

```
2001:db8:aaaa:105::/64                          srv6-isis SRV6  524291    0
  fe80::1e:1ff:fe01:15-"int-PE-2-PE-5"                            10
2001:bbbb:3::/48                                srv6-isis SRV6  524292    0
  fe80::14:1ff:fe01:1f-"int-PE-2-PE-3"                            10
2001:bbbb:4::/48                                srv6-isis SRV6  524293    0
  fe80::1a:1ff:fe01:b-"int-PE-2-PE-4"                             10
2001:bbbb:5::/48                                srv6-isis SRV6  524294    0
  fe80::1e:1ff:fe01:15-"int-PE-2-PE-5"                            10
-------------------------------------------------------------------------------
Flags: B = BGP or MPLS backup hop available
       L = Loop-Free Alternate (LFA) hop available
       E = Inactive best-external BGP route
       k = RIB-API or Forwarding Policy backup hop
===============================================================================
```

The IPv6 route table on PE-2 shows the following tunneled routes to the locators and the micro-segment locators on PE-3, PE-4, and PE-5:

```
*A:PE-2# show router route-table ipv6 next-hop-type tunneled

===============================================================================
IPv6 Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
2001:db8:aaaa:103::/64                        Remote  ISIS      00h22m21s  18
       2001:db8:aaaa:103::/64 (tunneled:SRV6-ISIS)                  10
2001:db8:aaaa:104::/64                        Remote  ISIS      00h22m13s  18
       2001:db8:aaaa:104::/64 (tunneled:SRV6-ISIS)                  10
2001:db8:aaaa:105::/64                        Remote  ISIS      00h21m51s  18
       2001:db8:aaaa:105::/64 (tunneled:SRV6-ISIS)                  10
2001:bbbb:3::/48                              Remote  ISIS      00h03m38s  18
       2001:bbbb:3::/48 (tunneled:SRV6-ISIS)                        10
2001:bbbb:4::/48                              Remote  ISIS      00h03m31s  18
       2001:bbbb:4::/48 (tunneled:SRV6-ISIS)                        10
2001:bbbb:5::/48                              Remote  ISIS      00h03m21s  18
       2001:bbbb:5::/48 (tunneled:SRV6-ISIS)                        10
-------------------------------------------------------------------------------
No. of Routes: 6
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

Multihoming and FDBs

PE-3 is DF in the all-active ES for VPLS-2, as follows:

```
*A:PE-3# show service id "VPLS-2" ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
lag-1:2               AA-vES-23-1                      DF
===============================================================================
No sdp entries
No vxlan instance entries
```

PE-4 is DF in the single-active ES for VPLS-2, as follows:

```
*A:PE-4# show service id "VPLS-2" ethernet-segment
No sap entries

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
46:2                  SA-ES-45                         DF
===============================================================================
No vxlan instance entries
```

When traffic is sent between CE-21 and CE-26, the FDBs are populated; on MTU-1 as follows:

```
*A:MTU-1# show service id "VPLS-2" fdb detail

===============================================================================
Forwarding Database, Service 2
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
2          00:00:5e:00:53:21 sap:1/1/c10/1:2         L/180    03/18/24 15:42:55
2          00:00:5e:00:53:26 sap:lag-1:2             L/180    03/18/24 15:44:04
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

The VPLS-2 FDBs on PE-2 and PE-3 are similar. CE-21 can be reached through the LAG, while CE-26 can be reached via the single-active ES between PE-4 and PE-5, as follows:

```
*A:PE-2# show service id "VPLS-2" fdb detail

===============================================================================
Forwarding Database, Service 2
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
2          00:00:5e:00:53:21 sap:lag-1:2             L/180    03/18/24 15:44:04
2          00:00:5e:00:53:26 eES:                    Evpn     03/18/24 15:44:04
                             01:00:00:00:00:45:00:00:00:01
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

```
*A:PE-3# show service id "VPLS-2" fdb detail

===============================================================================
Forwarding Database, Service 2
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
```

```
-------------------------------------------------------------------------------
2          00:00:5e:00:53:21 sap:lag-1:2             Evpn     03/18/24 15:44:04
2          00:00:5e:00:53:26 eES:                    Evpn     03/18/24 15:44:04
                             01:00:00:00:00:45:00:00:00:01
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

On DF PE-4, CE-21 can be reached via the all-active ES with ESI 01:00:00:00:00:23:00:00:01:01, while CE-26 can be reached via spoke SDP 46:2, as follows:

```
*A:PE-4# show service id "VPLS-2" fdb detail

===============================================================================
Forwarding Database, Service 2
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
2          00:00:5e:00:53:21 eES:                    Evpn     03/18/24 15:44:04
                             01:00:00:00:00:23:00:00:01:01
2          00:00:5e:00:53:26 sdp:46:2                L/180    03/18/24 15:44:04
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

On NDF PE-5, CE-21 can be reached via the all-active ES with ESI 01:00:00:00:00:23:00:00:01:01, while CE-26 can be reached via DF PE-4, as follows:

```
*A:PE-5# show service id "VPLS-2" fdb detail

===============================================================================
Forwarding Database, Service 2
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
2          00:00:5e:00:53:21 eES:                    Evpn     03/18/24 15:44:04
                             01:00:00:00:00:23:00:00:01:01
2          00:00:5e:00:53:26 eES:                    Evpn     03/18/24 15:44:04
                             01:00:00:00:00:45:00:00:00:01
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

On MTU-6, CE-21 can be reached via spoke SDP 64:2 to PE-4, while CE-26 can be reached via SAP 1/1/c10/1:2, as follows:

```
*A:MTU-6# show service id "VPLS-2" fdb detail

===============================================================================
Forwarding Database, Service 2
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
```

```
-------------------------------------------------------------------------------
2         00:00:5e:00:53:21 sdp:64:2               L/180    03/18/24 15:44:04
2         00:00:5e:00:53:26 sap:1/1/c10/1:2        L/180    03/18/24 15:43:59
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

BGP-EVPN routes

PE-2 received the following EVPN-MAC route containing the CE-26 MAC address from DF PE-4 with SRv6 SID 2001:bbbb:4:: and endpoint behavior 0x43 = 67 for End.uDT2U:

```
# on PE-2:
116 2024/03/18 15:44:03.874 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:4
"Peer 1: 2001:db8::2:4: UPDATE
Peer 1: 2001:db8::2:4 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 125
    Flag: 0x90 Type: 14 Len: 56 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 16 Global NextHop 2001:db8::2:4
        Type: EVPN-MAC Len: 33 RD: 192.0.2.4:2 ESI: 01:00:00:00:00:45:00:00:00:01,
                       tag: 0, mac len: 48 mac: 00:00:5e:00:53:26, IP len: 0,
                       IP: NULL, label1: 4195840 (Raw Label: 0x400600)
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64500:2
    Flag: 0xc0 Type: 40 Len: 37 Prefix-SID-attr:
       SRv6 Services TLV (37 bytes):-
           Type: SRV6 L2 Service TLV (6)
           Length: 34 bytes, Reserved: 0x0
         SRv6 Service Information Sub-TLV (33 bytes)
             Type: 1 Len: 30 Rsvd1: 0x0
             SRv6 SID: 2001:bbbb:4::
             SID Flags: 0x0 Endpoint Behavior: 0x43 Rsvd2: 0x0
             SRv6 SID Sub-Sub-TLV
                Type: 1 Len: 6
                BL:32 NL:16 FL:16 AL:0 TL:16 TO:48
"
```

The following command shows the same EVPN-MAC route containing the CE-26 MAC address:

```
*A:PE-2# show router bgp routes evpn mac mac-address 00:00:5e:00:53:26 detail
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Original Attributes
```

```
Network        : n/a
Nexthop        : 2001:db8::2:4
Path Id        : None
From           : 2001:db8::2:4
Res. Nexthop   : fe80::1a:1ff:fe01:b
Local Pref.    : 100                    Interface Name : int-PE-2-PE-4
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64500:2
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.4
Origin         : IGP
Flags          : Used Valid Best
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : MAC
ESI            : 01:00:00:00:00:45:00:00:00:01
Tag            : 0
IP Address     : n/a
Route Dist.    : 192.0.2.4:2
Mac Address    : 00:00:5e:00:53:26
MPLS Label1    : 16390                  MPLS Label2    : n/a
Route Tag      : 0
Neighbor-AS    : n/a
DB Orig Val    : N/A                    Final Orig Val : N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h05m05s
SRv6 TLV Type  : SRv6 L2 Service TLV (6)
SRv6 SubTLV    : SRv6 SID Information (1)
Sid            : 2001:bbbb:4::
Full Sid       : 2001:bbbb:4:4006::
Behavior       : End.uDT2U (67)
SRv6 SubSubTLV : SRv6 SID Structure (1)
Loc-Block-Len  : 32                     Loc-Node-Len   : 16
Func-Len       : 16                     Arg-Len        : 0
Tpose-Len      : 16                     Tpose-offset   : 48
---snip---
```

The full SID 2001:bbbb:4:4006:: contains the 32-bit prefix 2001:bbbb, the uN value 4 for PE-4, and the function 0x4006 = 16390 for the End.uDT2U SID in the following list:

```
*A:PE-2# show router segment-routing-v6 micro-segment-local-sid

===============================================================================
Micro Segment Routing v6 Local SIDs
===============================================================================
SID                                                 Type           Function
  Micro Segment Locator
  Context
-------------------------------------------------------------------------------
2001:bbbb:2::                                       uN             2
  PE2-mloc
  None
2001:bbbb:2:4005::                                  uA             16389
  PE2-mloc
  None
2001:bbbb:2:4006::                                  uDT2U          16390
  PE2-mloc
  SvcId: 2 Name: VPLS-2
2001:bbbb:2:4007::                                  uDT2M          16391
```

```
  PE2-mloc
  SvcId: 2 Name: VPLS-2
2001:bbbb:2:43e8::                                 uA             17384
  PE2-mloc
  None
2001:bbbb:2:43e9::                                 uA             17385
  PE2-mloc
  None
--------------------------------------------------------------------------------
SIDs : 6
--------------------------------------------------------------------------------
================================================================================
```

PE-2 received the following IMET route with endpoint behavior 0x44 = 68 for End.uDT2M from PE-4:

```
94 2024/03/18 15:43:43.589 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:4
"Peer 1: 2001:db8::2:4: UPDATE
Peer 1: 2001:db8::2:4 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 145
    Flag: 0x90 Type: 14 Len: 52 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 16 Global NextHop 2001:db8::2:4
        Type: EVPN-INCL-MCAST Len: 29 RD: 192.0.2.4:2, tag: 0,
                              orig_addr len: 128, orig_addr: 2001:db8::2:4
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:64500:2
    Flag: 0xc0 Type: 22 Len: 21 PMSI:
        Tunnel-type Ingress Replication (6)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf: not required]
        MPLS Label 4196096
        Tunnel-Endpoint 2001:db8::2:4
    Flag: 0xc0 Type: 40 Len: 37 Prefix-SID-attr:
       SRv6 Services TLV (37 bytes):-
           Type: SRV6 L2 Service TLV (6)
           Length: 34 bytes, Reserved: 0x0
         SRv6 Service Information Sub-TLV (33 bytes)
             Type: 1 Len: 30 Rsvd1: 0x0
             SRv6 SID: 2001:bbbb:4::
             SID Flags: 0x0 Endpoint Behavior: 0x44 Rsvd2: 0x0
             SRv6 SID Sub-Sub-TLV
                Type: 1 Len: 6
                BL:32 NL:16 FL:16 AL:16 TL:16 TO:48
"
```

The following shows the same IMET route:

```
*A:PE-2# show router bgp routes evpn incl-mcast community target:64500:2 hunt
================================================================================
 BGP Router ID:192.0.2.2        AS:64500        Local AS:64500
================================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

================================================================================
BGP EVPN Inclusive-Mcast Routes
================================================================================
--------------------------------------------------------------------------------
```

```
RIB In Entries
-------------------------------------------------------------------------------
---snip---
Network        : n/a
Nexthop        : 2001:db8::2:4
Path Id        : None
From           : 2001:db8::2:4
Res. Nexthop   : fe80::1a:1ff:fe01:b
Local Pref.    : 100                    Interface Name : int-PE-2-PE-4
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64500:2
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.4
Origin         : IGP
Flags          : Used Valid Best
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : INCL-MCAST
Tag            : 0
Originator IP  : 2001:db8::2:4
Route Dist.    : 192.0.2.4:2
Route Tag      : 0
Neighbor-AS    : n/a
DB Orig Val    : N/A                    Final Orig Val : N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h06m08s
SRv6 TLV Type  : SRv6 L2 Service TLV (6)
SRv6 SubTLV    : SRv6 SID Information (1)
Sid            : 2001:bbbb:4::
Full Sid       : 2001:bbbb:4:4007::
Behavior       : End.uDT2M (68)
SRv6 SubSubTLV : SRv6 SID Structure (1)
Loc-Block-Len  : 32                     Loc-Node-Len   : 16
Func-Len       : 16                     Arg-Len        : 16
Tpose-Len      : 16                     Tpose-offset   : 48
-------------------------------------------------------------------------------
PMSI Tunnel Attributes :
Tunnel-type    : Ingress Replication
Flags          : Type: RNVE(0) BM: 0 U: 0 Leaf: not required
MPLS Label     : 4196096
Tunnel-Endpoint: 2001:db8::2:4
-------------------------------------------------------------------------------
---snip---
```

PE-2 receives the following EVPN-AD routes from PE-4: the first route is an AD per-EVI route and the second one is an AD per-ES route:

```
*A:PE-2# show router bgp routes evpn auto-disc rd 192.0.2.4:2 detail
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Original Attributes

Network        : n/a
Nexthop        : 2001:db8::2:4
Path Id        : None
From           : 2001:db8::2:4
Res. Nexthop   : fe80::204:feff:fe00:0
Local Pref.    : 100                    Interface Name : int-PE-2-PE-4
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64500:2
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.4
Origin         : IGP
Flags          : Used Valid Best
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : AUTO-DISC
ESI            : 01:00:00:00:00:45:00:00:00:01
Tag            : 0
Route Dist.    : 192.0.2.4:2
MPLS Label     : 16389
Route Tag      : 0
Neighbor-AS    : n/a
DB Orig Val    : N/A                    Final Orig Val : N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 01h39m21s
SRv6 TLV Type  : SRv6 L2 Service TLV (6)
SRv6 SubTLV    : SRv6 SID Information (1)
Sid            : 2001:bbbb:4::
Full Sid       : 2001:bbbb:4:4005::
Behavior       : End.uDT2U (67)
SRv6 SubSubTLV : SRv6 SID Structure (1)
Loc-Block-Len  : 32                     Loc-Node-Len   : 16
Func-Len       : 16                     Arg-Len        : 0
Tpose-Len      : 16                     Tpose-offset   : 48
---snip---

-------------------------------------------------------------------------------
Original Attributes

Network        : n/a
Nexthop        : 2001:db8::2:4
Path Id        : None
From           : 2001:db8::2:4
Res. Nexthop   : fe80::204:feff:fe00:0
Local Pref.    : 100                    Interface Name : int-PE-2-PE-4
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 0
Connector      : None
Community      : target:64500:2 esi-label:1/Single-Active
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.4
Origin         : IGP
Flags          : Used Valid Best
Route Source   : Internal
AS-Path        : No As-Path
Add Paths Send : Default
Last Modified  : 01h39m21s
SRv6 TLV Type  : SRv6 L2 Service TLV (6)
```

```
SRv6 SubTLV    : SRv6 SID Information (1)
Sid            : ::
EVPN type      : AUTO-DISC
ESI            : 01:00:00:00:00:45:00:00:00:01
Tag            : MAX-ET
Route Dist.    : 192.0.2.4:2
MPLS Label     : 0
Route Tag      : 0
Neighbor-AS    : n/a
DB Orig Val    : N/A                    Final Orig Val : N/A
Source Class   : 0                      Dest Class     : 0
Full Sid       : ::
Behavior       : End.uDT2M (68)
SRv6 SubSubTLV : SRv6 SID Structure (1)
Loc-Block-Len  : 32                     Loc-Node-Len   : 16
Func-Len       : 16                     Arg-Len        : 16
Tpose-Len      : 16                     Tpose-offset   : 64

---snip---
```

ES "SA-ES-45" is configured in a single-active multihoming mode with the ESI label; therefore, PE-4 sends AD per-ES routes that carry the arg.fe2 value. The SID structure in the SRv6 sub-sub-TLV indicates that the argument length is 16 bits. The 16-bit length of the arg.fe2 value is transposed into the ESI label extended community label field. The transposition length is 16 (argument length 16) and the transposition offset is 64 (block length 32 + node length 16 + function length 16). The SRv6 SID is :: (0).

SRv6 functions and SIDs

The following command on PE-2 shows the SIDs for the End.uDT2U and End.uDT2M functions:

```
*A:PE-2# show router segment-routing-v6 micro-segment-local-sid udt2u udt2m

===============================================================================
Micro Segment Routing v6 Local SIDs
===============================================================================
SID                                               Type           Function
  Micro Segment Locator
  Context
-------------------------------------------------------------------------------
2001:bbbb:2:4006::                                uDT2U          16390
  PE2-mloc
  SvcId: 2 Name: VPLS-2
2001:bbbb:2:4007::                                uDT2M          16391
  PE2-mloc
  SvcId: 2 Name: VPLS-2
-------------------------------------------------------------------------------
SIDs : 2
-------------------------------------------------------------------------------
===============================================================================
```

The same SRv6 functions and SIDs are applied in the VPLS-2 service, as follows:

```
*A:PE-2# show service id "VPLS-2" segment-routing-v6 detail

===============================================================================
Micro Segment Routing v6 Instance 1 Service 2
===============================================================================
Micro Segment Locator
 Type          Function  SID                                   Status
```

```
               Oper Func
-------------------------------------------------------------------------------
PE2-mloc
 uDT2U         *-        2001:bbbb:2:4006::                      ok
               16390
 uDT2M         *-        2001:bbbb:2:4007::                      ok
               16391
===============================================================================
Legend: * - System allocated
```

The following command on PE-2 shows the SRv6 destinations in VPLS-2:

```
*A:PE-2# show service id "VPLS-2" segment-routing-v6 destinations

===============================================================================
TEP, SID (Instance 1)
===============================================================================
TEP Address                Segment Id                          Oper  Mcast Num
                                                               State       MACs
-------------------------------------------------------------------------------
2001:db8::2:3              2001:bbbb:3:4007::                  Up    BUM   0
2001:db8::2:4              2001:bbbb:4:4007::                  Up    BUM   0
2001:db8::2:5              2001:bbbb:5:4007::                  Up    BUM   0
-------------------------------------------------------------------------------
Number of TEP, SID: 3
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
Segment Routing v6 Ethernet Segment Dest (Instance 1)
===============================================================================
Eth SegId                                 Num. Macs     Last Update
-------------------------------------------------------------------------------
01:00:00:00:00:45:00:00:00:01             1             03/18/2024 15:44:04
-------------------------------------------------------------------------------
Number of entries: 1
-------------------------------------------------------------------------------
===============================================================================
```

The ping or traceroute commands are used to verify the connectivity toward the remote SRv6 SIDs, for example, for the End.uDT2U function, as follows:

```
*A:PE-2# ping 2001:bbbb:3:4006:: rapid
PING 2001:bbbb:3:4006:: 56 data bytes
!!!!!
---- 2001:bbbb:3:4006:: PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 2.37ms, avg = 2.54ms, max = 2.92ms, stddev = 0.191ms
```

```
*A:PE-2# traceroute 2001:bbbb:5:4006::
traceroute to 2001:bbbb:5:4006::, 30 hops max, 60 byte packets
  1  2001:db8::2:5 (2001:db8::2:5)    3.08 ms  2.74 ms  2.98 ms
```

Conclusion

EVPN VPLS services using SRv6 transport can be configured with locators as well as micro-segment locators. Both all-active and single-active multihoming modes are supported.

EVPN VPLS with MPLS to SRv6 or VXLAN to SRv6 Stitching

This chapter provides information about the EVPN VPLS with MPLS to SRv6 or VXLAN to SRv6 stitching.

Topics in this chapter include:

Applicability

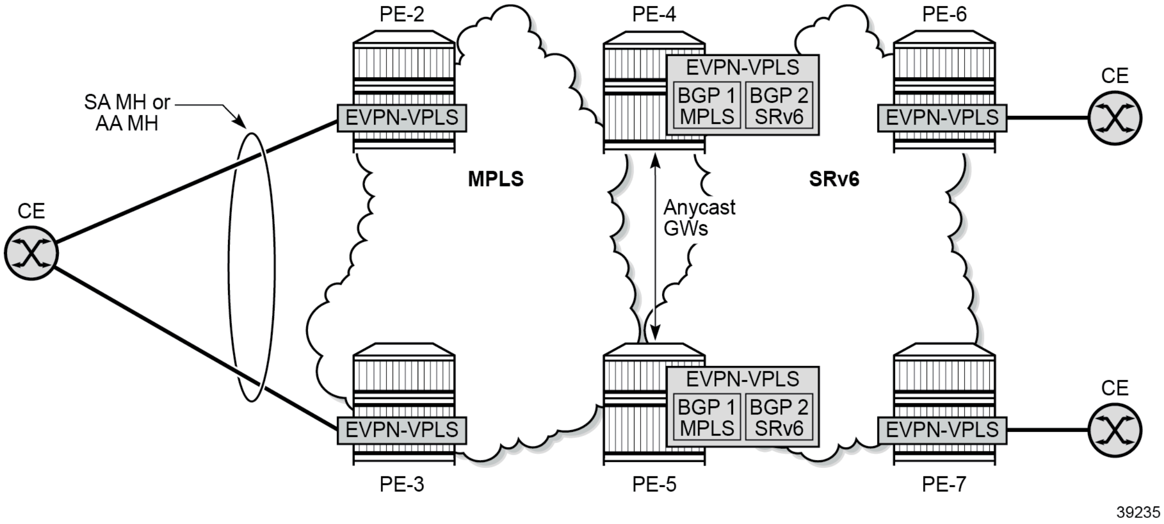
The information and configuration in this chapter are based on SR OS Release 23.10.R2. MPLS to SRv6 stitching within an EVPN VPLS is supported in SR OS Release 22.10.R1 and later; VXLAN to SRv6 stitching within an EVPN VPLS is supported in SR OS Release 22.10.R3 and later.

Overview

SRv6 to MPLS stitching or SRv6 to VXLAN stitching is required in hybrid networks where MPLS PEs and SRv6 PEs are both attached to the same EVPN VPLS service. This concept follows the RFC 9014 standard and it is implemented in SR OS by using two EVPN instances in the same EVPN VPLS service. Also, a migration from MPLS tunnels to SRv6 tunnels in EVPN VPLS services requires the support of an SRv6 instance and an MPLS instance in the same EVPN VPLS service. EVPN destinations of different transport types (MPLS, VXLAN, or SRv6) can be placed in the same Split Horizon Groups (SHGs) to avoid loops.

Figure 121: The need for MPLS to SRv6 stitching in an EVPN VPLS shows an EVPN VPLS service configured in all PEs where the gateways (GWs) PE-4 and PE-5 have two service instances in the EVPN VPLS: BGP instance 1 uses MPLS transport and BGP instance 2 uses SRv6. For GW redundancy on PE-4 and PE-5, the anycast multihoming concept is applied.

Figure 121: The need for MPLS to SRv6 stitching in an EVPN VPLS

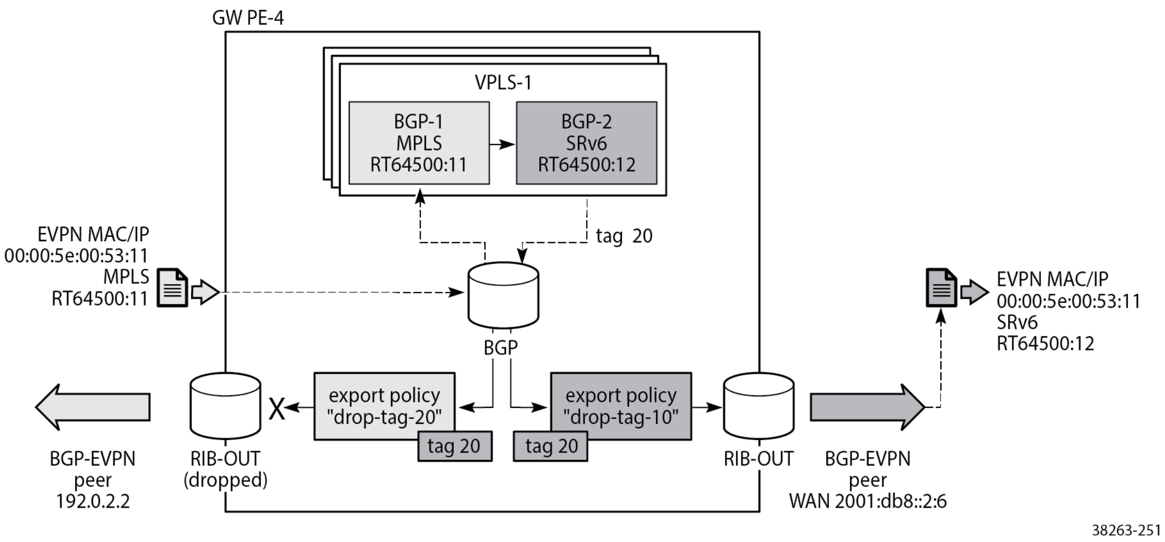


The following applies for multi-instance EVPN VPLS services with SRv6:

- SHGs are supported:
  - adding or removing an SHG to the EVPN MPLS instance or the EVPN SRv6 instance requires disabling the instance
  - EVPN VXLAN cannot be configured with an explicit SHG
  - SHG associated to the EVPN MPLS instance can be the same as or different from the SHG associated to the EVPN SRv6 instance:
    - when the same SHG is configured across two instances, no routes are redistributed between the instances
    - when different SHGs are configured across two instances, routes are redistributed between the instances
- when one of the instances is SRv6, SAPs can be configured only if the two instances are configured with the same explicit SHG
- no SDP bindings are supported on multi-instance VPLS services:
  - the configuration of spoke SDPs or mesh SDPs is blocked
  - BGP VPLS and BGP AD can be configured, but the spoke SDPs are not auto-created
- the **mh-mode access | network** command is supported to configure multihoming:
  - access or network mode can be configured in an SRv6 instance, with network being the default mode
  - the following combinations are supported:
    - SRv6 mh-mode network with MPLS instance mh-mode access
    - SRv6 mh-mode access with MPLS instance mh-mode network (Note: If provider-tunnel is needed on the multi-instance service, the MPLS instance must be configured as mh-mode network)
    - two mh-mode access instances in the same EVPN VPLS are allowed for the combinations VXLAN/MPLS, VXLAN/SRv6, and MPLS/SRv6 (but not for the combination MPLS/MPLS)

- Anycast multihoming can be applied:
  - two or more PEs can be configured with the same service parameters as part of the same redundancy group:
    - same Route Distinguisher (RD) for the same BGP instance
    - same Route Target (RT) for the same BGP instance
    - same inclusive multicast originator IP address
  - remote PEs set up EVPN destinations to only one PE in the anycast group for a service
  - no EVPN BUM destinations are established among the PEs in the anycast group because the received anycast peer inclusive multicast Ethernet tag (IMET) routes have the same inclusive multicast originator IP address
  - policies are applied on the GW PEs to avoid loops:
    - export policies add route target and site-of-origin (SOO) extended communities to the redistributed MAC/IP routes and the peer GW PEs drop the routes received with the group SOO
    - default route tags per service instance differentiate the allowed non-redistributed MAC/IP routes from the rest, so that these MAC/IP routes are not advertised between access and network peers, as shown in Figure 122: Default route tags per service instance avoid loops.

Figure 122: Default route tags per service instance avoid loops



The figure shows that an incoming MAC/IP route with RT 64500:11 is accepted in VPLS-1 on GW PE-4 by service instance 1 and passed on to service instance 2 where it gets RT 64500:12 and default route tag 20. Routes with route tag 20 are accepted for routes sent to peer 2001:db8::2:6 in the network, but not for routes sent back to access peer 192.0.2.2. Likewise (but not shown in the figure), routes coming from peer 2001:db8::2:6 with RT 64500:12 are accepted by service instance 2 and passed on to service instance 1 where default route tag 10 is added. Routes with route tag 10 are forwarded to peer 192.0.2.2, but are not sent back to the network peer 2001:db8::2:6.

Configuration

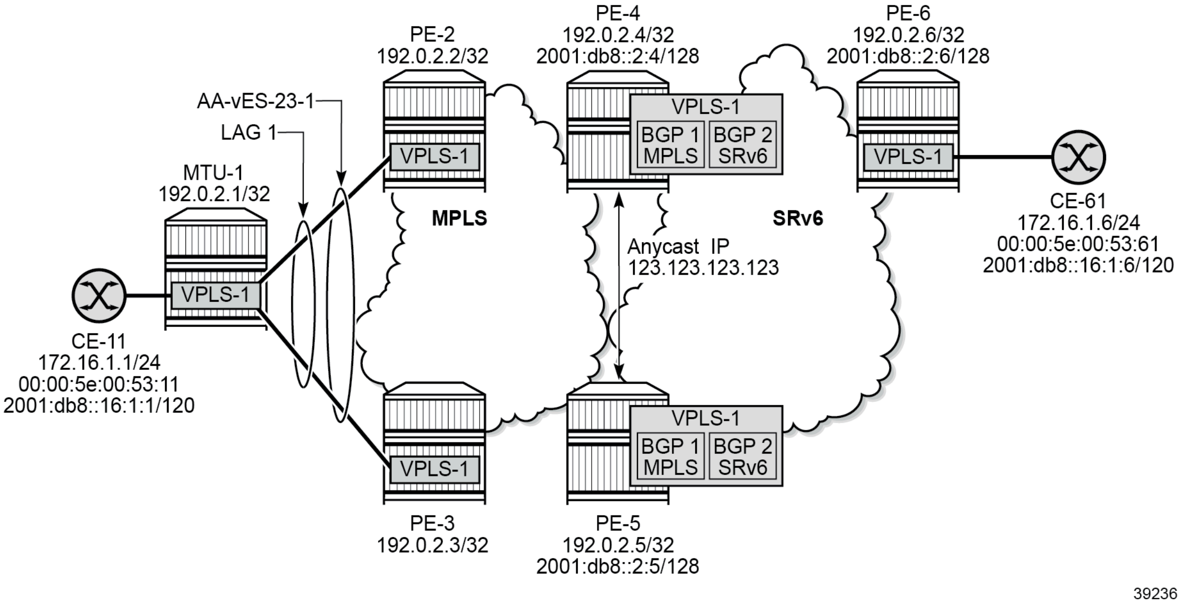
The following use cases are described in this section:

- EVPN VPLS with MPLS to SRv6 stitching
- EVPN VPLS with VXLAN to SRv6 stitching

EVPN VPLS with MPLS to SRv6 stitching

Figure 123: Example topology with VPLS-1 shows the example topology for VPLS-1 with MPLS to SRv6 stitching in the GWs PE-4 and PE-5.

Figure 123: Example topology with VPLS-1



The initial configuration on the nodes includes:

- cards, MDAs, ports, LAG
- router interfaces
- IS-IS as IGP
- LDP between PE-2, PE-3, PE-4, PE-5

SRv6 configuration

The SRv6 configuration on PE-4 is as follows:

```
# on PE-4:
configure
    card 1
```

```
        mda 1
            xconnect
                mac 1 create
                    loopback 1 create
                    exit
                    loopback 2 create
                    exit
                exit
            exit
            no shutdown
        exit
    exit
    port-xc
        pxc 1 create
            port 1/1/m1/1
            no shutdown
        exit
        pxc 2 create
            port 1/1/m1/2
            no shutdown
        exit
    exit
    port pxc-1.a
        ethernet
        exit
        no shutdown
    exit
    port pxc-1.b
        ethernet
        exit
        no shutdown
    exit
    port pxc-2.a
        ethernet
        exit
        no shutdown
    exit
    port pxc-2.b
        ethernet
        exit
        no shutdown
    exit
    port 1/1/m1/1
        no shutdown
    exit
    port 1/1/m1/2
        no shutdown
    exit
    fwd-path-ext
        fpe 1 create
            path pxc 1
            srv6 origination
                interface-a
                exit
                interface-b
                exit
            exit
        exit
        fpe 2 create
            path pxc 2
            srv6 termination
                interface-a
                exit
                interface-b
                exit
            exit
        exit
    exit
    router Base
        segment-routing
            segment-routing-v6
                origination-fpe 1
                source-address 2001:db8::2:4
                locator "PE4-loc"
                    block-length 48
                    termination-fpe 2
                    prefix
                        ip-prefix 2001:db8:aaaa:104::/64
                    exit
                    static-function
                    exit
                    no shutdown
                exit
                base-routing-instance
                    locator "PE4-loc"
                        function
                            end-x-auto-allocate srh-mode psp protection unprotected
                            end 1
                                srh-mode usp
                            exit
                        exit
                    exit
                exit
            exit
        exit
        isis 0
            level-capability level-2
            area-id 49.0001
            traffic-engineering
            traffic-engineering-options
                ipv6
                application-link-attributes
                exit
            exit
            advertise-passive-only
            advertise-router-capability as
            ipv6-routing native
            level 2
                wide-metrics-only
            exit
            segment-routing-v6
                locator "PE4-loc"
                    level-capability level-2
                    level 1
                    exit
                    level 2
                    exit
                exit
                no shutdown
            exit
            interface "system"
                passive
                no shutdown
            exit
---snip---
```

The SRv6 configuration on PE-5 and PE-6 is similar.

BGP configuration

In the MPLS access network, PE-2 acts as the route reflector (RR) with clients PE-3, PE-4, and PE-5. The BGP configuration is as follows:

```
# on RR PE-2:
configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "access-mpls"
                family evpn
                cluster 192.0.2.2
                peer-as 64500
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
```

The BGP configuration on PE-3 is as follows:

```
# on PE-3:
configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "access-mpls"
                family evpn
                peer-as 64500
                neighbor 192.0.2.2
                exit
            exit
```

In the SRv6 network, PE-6 acts as the RR. The BGP configuration on the GWs PE-4 and PE-5 is as follows. The export policy "drop-tag-10" is used to avoid loops within the core SRv6 network and the export policy "drop-tag-20" is used to avoid loops within the access MPLS network, as shown in Figure 122: Default route tags per service instance avoid loops.

```
# on GWs PE-4 and PE-5:
configure
    router Base
        autonomous-system 64500
        policy-options
            policy-statement "drop-tag-10"
```

```
                description "route tag 10 in instance 1 of VPLS to avoid loops"
                entry 10
                    from
                        tag 10
                    exit
                    action drop
                    exit
                exit
            exit
            policy-statement "drop-tag-20"
                description "route tag 20 in instance 2 of VPLS to avoid loops"
                entry 10
                    from
                        tag 20
                    exit
                    action drop
                    exit
                exit
            exit
            commit
        exit
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "core-srv6"
                family evpn
                export "drop-tag-10"
                peer-as 64500
                neighbor 2001:db8::2:6
                exit
            exit
            group "access-mpls"
                family evpn
                export "drop-tag-20"
                peer-as 64500
                neighbor 192.0.2.2
                exit
            exit
```

The BGP configuration on RR PE-6 is as follows:

```
# on RR PE-6:
configure
    router Base
        autonomous-system 64500
        bgp
            group "core-srv6"
                family evpn
                cluster 192.0.2.6
                peer-as 64500
                neighbor 2001:db8::2:4
                exit
                neighbor 2001:db8::2:5
                exit
            exit
```

Service configuration

VPLS-1 is configured on all nodes. On PE-2, PE-3, PE-4, and PE-5, service instance 1 of VPLS-1 uses MPLS tunnels. The service configuration on PE-2 and PE-3 is identical, except for the preference value in the all-active Ethernet segment (ES). Route target 64500:11 is accepted in service instance 1 of VPLS-1 on the GW PEs PE-4 and PE-5. The service configuration on PE-2 is as follows:

```
# on PE-2:
configure
    service
        system
            bgp-evpn
                ethernet-segment "AA-vES-23-1" virtual create
                    esi 01:00:00:00:00:23:00:00:01:01
                    es-activation-timer 3
                    service-carving
                        mode manual
                        manual
                            preference non-revertive create
                                value 100                    # on PE-3: preference 150
                            exit
                        exit
                    exit
                    multi-homing all-active
                    lag 1
                    dot1q
                        q-tag-range 1
                    exit
                    no shutdown
                exit
            exit
        exit
        vpls 1 name "VPLS-1" customer 1 create
            bgp
                route-target export target:64500:11 import target:64500:11
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap lag-1:1 create
                no shutdown
            exit
            no shutdown
        exit
```

On PE-6, VPLS-1 uses SRv6 transport and route target 64500:12 is accepted in service instance 2 of VPLS-1 on the GW PEs PE-4 and PE-5. SAP 1/1/c10/1:1 is connected to CE-61. The configuration on PE-6 is as follows.

```
# on PE-6:
configure
    service
        vpls 1 name "VPLS-1" customer 1 create
            segment-routing-v6 1 create
                locator "PE6-loc"
                    function
                        end-dt2u
                        end-dt2m
                    exit
                exit
            exit
            bgp
                route-target export target:64500:12 import target:64500:12
            exit
            bgp-evpn
                evi 1
                segment-routing-v6 bgp 1 srv6-instance 1 default-locator "PE6-loc" create
                    route-next-hop system-ipv6
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/1/c10/1:1 create
                no shutdown
            exit
            no shutdown
        exit
```

The following configuration on the anycast GW PE-4 shows that the EVPN VPLS is configured with two instances: service instance 1 uses MPLS transport and service instance 2 uses SRv6. The configuration on GW PE-5 is identical with only a different SRv6 locator name. The VSI policies are used to accept EVPN routes with the matching route target and to avoid loops between GWs PE-4 and PE-5 based on the SOO.

```
# on PE-4:
configure
    router Base
        policy-options
            begin
            community "SOO-45"
                members "origin:45:45"
            exit
            community "RT64500:11"
                members "target:64500:11"
            exit
            community "RT64500:12"
                members "target:64500:12"
            exit
            policy-statement "vsi-11-export"
                entry 10
                    action accept
                        community add "RT64500:11" "SOO-45"
                    exit
                exit
            exit
            policy-statement "vsi-11-import"
                entry 10
                    from
                        community "SOO-45"
                        family evpn
                    exit
                    action drop
                    exit
                exit
                entry 20
                    from
                        community "RT64500:11"
                        family evpn
                    exit
                    action accept
                    exit
                exit
            exit
            policy-statement "vsi-12-export"
                entry 10
                    action accept
                        community add "RT64500:12" "SOO-45"
                    exit
                exit
            exit
            policy-statement "vsi-12-import"
                entry 10
                    from
                        community "SOO-45"
                        family evpn
                    exit
                    action drop
                    exit
                exit
                entry 20
                    from
                        community "RT64500:12"
                        family evpn
                    exit
                    action accept
                    exit
                exit
            exit
            commit
        exit
    exit
    service
        vpls 1 name "VPLS-1" customer 1 create
            split-horizon-group "SHG-1" create
            exit
            split-horizon-group "SHG-2" create
            exit
            segment-routing-v6 1 create
                locator "PE4-loc"
                    function
                        end-dt2u
                        end-dt2m
                    exit
                exit
            exit
            bgp
                route-distinguisher 192.0.2.45:1
                vsi-export "vsi-11-export"
```

```
                vsi-import "vsi-11-import"
            exit
            bgp 2
                route-distinguisher 192.0.2.54:1
                vsi-export "vsi-12-export"
                vsi-import "vsi-12-import"
            exit
            bgp-evpn
                incl-mcast-orig-ip 145.145.145.145
                evi 1
                mpls bgp 1
                    mh-mode access
                    split-horizon-group "SHG-1"
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    default-route-tag 10
                    no shutdown
                exit
                segment-routing-v6 bgp 2 srv6-instance 1 default-locator "PE4-loc" create
                    default-route-tag 20
                    route-next-hop system-ipv6
                    split-horizon-group "SHG-2"
                    no shutdown
                exit
            exit
            no shutdown
```

The configuration of SHGs is optional. In this example, different SHGs are applied to the two service instances, so the routes can be redistributed between the instances.

In the anycast solution, the RDs, RTs, and the originator IP address must be identical on the GWs PE-4 and PE-5:

- originator IP address: 145.145.145.145
- for instance 1: RD 192.0.2.45:1, RT 64500:11
- for instance 2: RD 192.0.2.54:1, RT 64500:12

Service instance 1 has default route tag 10 and service instance 2 has default route tag 20. These route tags allow to differentiate routes and avoid loops as shown in Figure 122: Default route tags per service instance avoid loops.

The MPLS multihoming mode is access in this example; the SRv6 multihoming mode is network (default).

Show commands

After VPLS-1 is configured on all nodes, traffic is sent between CE-11 and CE-61. The FDB for VPLS-1 on PE-2 is as follows:

```
*A:PE-2# show service id "VPLS-1" fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:5e:00:53:11 sap:lag-1:1             L/150    12/21/23 08:37:33
```

```
1          00:00:5e:00:53:61 mpls-1:                Evpn     12/21/23 08:37:42
                             192.0.2.4:524279
           ldp:65538
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

PE-2 learns the remote MAC addresses via GW PE-4 (active) and GW PE-5 (backup). The following shows the received EVPN MAC routes for MAC address 00:00:5e:00:53:61 of CE-61. The anycast RD 192.0.2.45:1 is used.

```
*A:PE-2# show router bgp routes evpn mac mac-address 00:00:5e:00:53:61
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.45:1        00:00:5e:00:53:61 ESI-0
      0                   Seq:0             LABEL 524279
                          n/a
                          192.0.2.4

*>i   192.0.2.45:1        00:00:5e:00:53:61 ESI-0
      0                   Seq:0             LABEL 524279
                          n/a
                          192.0.2.5

-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

The detailed output for the first of these EVPN MAC routes shows the communities for RT, SOO, and the tunnel encapsulation MPLS.

```
*A:PE-2# show router bgp routes evpn mac mac-address 00:00:5e:00:53:61 detail
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Original Attributes

Network        : n/a
```

```
Nexthop        : 192.0.2.4
Path Id        : None
From           : 192.0.2.4
Res. Nexthop   : 192.168.24.2
Local Pref.    : 100                    Interface Name : int-PE-2-PE-4
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64500:11 origin:45:45 bgp-tunnel-encap:MPLS
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.4
Origin         : IGP
Flags          : Used Valid Best
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : MAC
ESI            : ESI-0
Tag            : 0
IP Address     : n/a
Route Dist.    : 192.0.2.45:1
Mac Address    : 00:00:5e:00:53:61
MPLS Label1    : LABEL 524279           MPLS Label2    : n/a
Route Tag      : 0
Neighbor-AS    : n/a
DB Orig Val    : N/A                    Final Orig Val : N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h03m32s
---snip---
```

The GWs PE-4 and PE-5 redistribute EVPN MAC routes between MPLS and SRv6 domains. The FDB on PE-4 shows that MAC address 00:00:5e:00:53:11 of CE-11 can be reached via the all-active ES with ESI 01:00:00:00:00:23:00:00:01:01, which is in the MPLS domain. MAC address 00:00:5e:00:53:61 of CE-61 can be reached using an SRv6 tunnel to PE-6. The FDB table on PE-5 is similar.

```
*A:PE-4# show service id "VPLS-1" fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:5e:00:53:11 eES:                    Evpn     12/21/23 08:37:27
                             01:00:00:00:00:23:00:00:01:01
1          00:00:5e:00:53:61 srv6-1:                 Evpn     12/21/23 08:37:42
                             2001:db8::2:6
           2001:db8:aaaa:106:8000::
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

The remote PEs, such as PE-2, set up EVPN destinations to only one GW PE in the anycast group for an EVPN VPLS service. The following command on PE-2 shows that two IMET routes with originator IP address 145.145.145.145 are valid, but only the IMET route from PE-4 is used:

```
*A:PE-2# show router bgp routes evpn incl-mcast originator-ip 145.145.145.145
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Flag  Route Dist.         OrigAddr
      Tag                 NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.45:1        145.145.145.145
      0                   192.0.2.4

*>i   192.0.2.45:1        145.145.145.145
      0                   192.0.2.5

-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

The details of the used IMET route from PE-4 are the following:

```
*A:PE-2# show router bgp routes evpn incl-mcast originator-ip 145.145.145.145 detail
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Original Attributes

Network        : n/a
Nexthop        : 192.0.2.4
Path Id        : None
From           : 192.0.2.4
Res. Nexthop   : 192.168.24.2
Local Pref.    : 100                    Interface Name : int-PE-2-PE-4
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64500:11 origin:45:45 bgp-tunnel-encap:MPLS
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.4
Origin         : IGP
Flags          : Used Valid Best
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : INCL-MCAST
Tag            : 0
Originator IP  : 145.145.145.145
Route Dist.    : 192.0.2.45:1
Route Tag      : 0
Neighbor-AS    : n/a
DB Orig Val    : N/A                    Final Orig Val : N/A
```

```
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h06m13s
-------------------------------------------------------------------------------
PMSI Tunnel Attributes :
Tunnel-type    : Ingress Replication
Flags          : Type: RNVE(0) BM: 0 U: 0 Leaf: not required
MPLS Label     : LABEL 524279
Tunnel-Endpoint: 192.0.2.4
-------------------------------------------------------------------------------
---snip---
```

In the MPLS domain, PE-2 sets up EVPN destinations to PE-3 and to GW PE-4, but not to GW PE-5:

```
*A:PE-2# show service id "VPLS-1" evpn-mpls

===============================================================================
BGP EVPN-MPLS Dest (Instance 1)
===============================================================================
TEP Address                     Transport:Tnl     Egr Label  Oper  Mcast  Num
                                                             State        MACs
-------------------------------------------------------------------------------
192.0.2.3                       ldp:65537         524279     Up    bum    0
192.0.2.4                       ldp:65538         524279     Up    bum    1
-------------------------------------------------------------------------------
Number of entries: 2
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

In the MPLS domain, PE-4 sets up EVPN destinations to PE-2, PE-3, and the all-active ES with ESI 01:00:00:00:00:23:00:00:01:01, but not to GW PE-5:

```
*A:PE-4# show service id "VPLS-1" evpn-mpls

===============================================================================
BGP EVPN-MPLS Dest (Instance 1)
===============================================================================
TEP Address                     Transport:Tnl     Egr Label  Oper  Mcast  Num
                                                             State        MACs
-------------------------------------------------------------------------------
192.0.2.2                       ldp:65538         524279     Up    bum    0
192.0.2.3                       ldp:65537         524279     Up    bum    0
-------------------------------------------------------------------------------
Number of entries: 2
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-MPLS Dest (Instance 2)
===============================================================================
TEP Address                     Transport:Tnl     Egr Label  Oper  Mcast  Num
                                                             State        MACs
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================

===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest (Instance 1)
===============================================================================
Eth SegId                       Num. Macs               Last Update
-------------------------------------------------------------------------------
```

```
01:00:00:00:00:23:00:00:01:01   1                       12/21/2023 08:37:27
-------------------------------------------------------------------------------
Number of entries: 1
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

In the SRv6 domain, PE-6 receives two IMET routes with originator IP address 145.145.145.145, but only the IMET route from PE-4 is used:

```
*A:PE-6# show router bgp routes evpn incl-mcast originator-ip 145.145.145.145
===============================================================================
 BGP Router ID:192.0.2.6        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Flag  Route Dist.         OrigAddr
      Tag                 NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.54:1        145.145.145.145
      0                   2001:db8::2:4

*>i   192.0.2.54:1        145.145.145.145
      0                   2001:db8::2:5

-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

```
*A:PE-6# show router bgp routes evpn incl-mcast originator-ip 145.145.145.145 detail
===============================================================================
 BGP Router ID:192.0.2.6        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Original Attributes

Network        : n/a
Nexthop        : 2001:db8::2:4
Path Id        : None
From           : 2001:db8::2:4
Res. Nexthop   : fe80::1a:1ff:fe01:1
Local Pref.    : 100                    Interface Name : int-PE-6-PE-4
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64500:12 origin:45:45
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.4
```

```
Origin         : IGP
Flags          : Used Valid Best
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : INCL-MCAST
Tag            : 0
Originator IP  : 145.145.145.145
Route Dist.    : 192.0.2.54:1
Route Tag      : 0
Neighbor-AS    : n/a
DB Orig Val    : N/A                    Final Orig Val : N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h09m32s
SRv6 TLV Type  : SRv6 L2 Service TLV (6)
SRv6 SubTLV    : SRv6 SID Information (1)
Sid            : 2001:db8:aaaa:104::
Full Sid       : 2001:db8:aaaa:104:7fff:9000::
Behavior       : End.DT2M (24)
SRv6 SubSubTLV : SRv6 SID Structure (1)
Loc-Block-Len  : 48                     Loc-Node-Len   : 16
Func-Len       : 20                     Arg-Len        : 0
Tpose-Len      : 20                     Tpose-offset   : 64
-------------------------------------------------------------------------------
PMSI Tunnel Attributes :
Tunnel-type    : Ingress Replication
Flags          : Type: RNVE(0) BM: 0 U: 0 Leaf: not required
MPLS Label     : 8388496
Tunnel-Endpoint: 2001:db8::2:4
-------------------------------------------------------------------------------
---snip---
```

In the SRv6 domain, PE-6 sets up SRv6 destinations to PE-4, but not to PE-5:

```
*A:PE-6# show service id "VPLS-1" segment-routing-v6 destinations

===============================================================================
TEP, SID (Instance 1)
===============================================================================
TEP Address                Segment Id                          Oper  Mcast Num
                                                               State       MACs
-------------------------------------------------------------------------------
2001:db8::2:4              2001:db8:aaaa:104:7fff:9000::       Up    BUM   0
2001:db8::2:4              2001:db8:aaaa:104:7fff:a000::       Up    -     1
-------------------------------------------------------------------------------
Number of TEP, SID: 2
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
Segment Routing v6 Ethernet Segment Dest (Instance 1)
===============================================================================
Eth SegId                                 Num. Macs     Last Update
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

On PE-6, the following information for the SRv6 instance 1 in VPLS-1 shows the End.DT2U and End.DT2M types with the corresponding SIDs, and status:

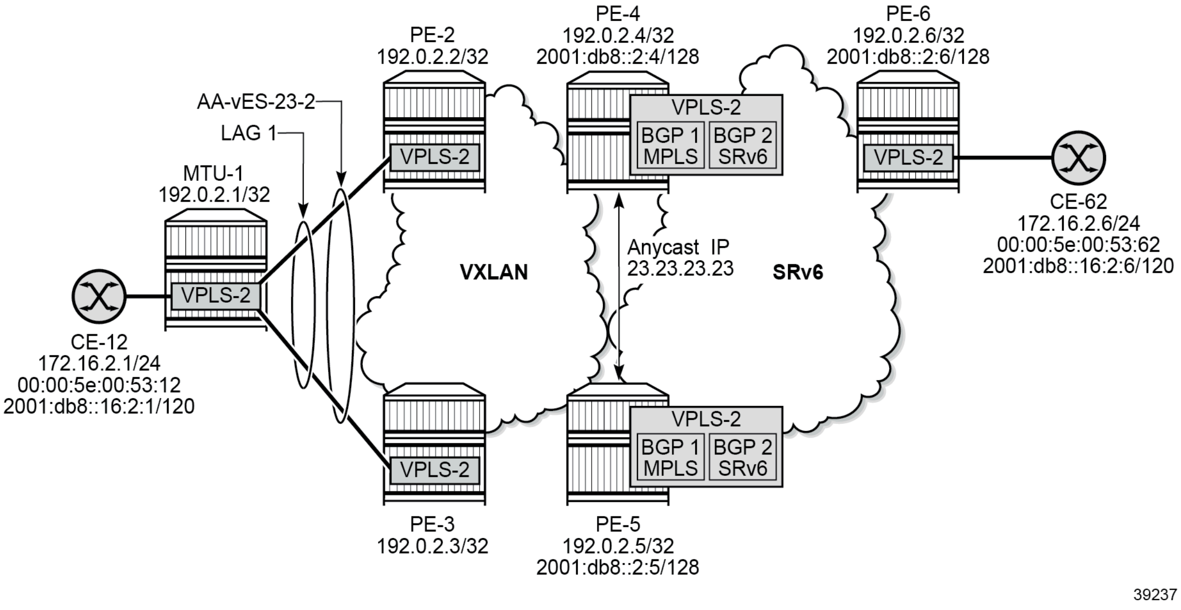
```
*A:PE-6# show service id "VPLS-1" segment-routing-v6 detail
```

```
===============================================================================
Segment Routing v6 Instance 1 Service 1
===============================================================================
Locator
 Type          Function  SID                                    Status
-------------------------------------------------------------------------------
PE6-loc
  End.DT2U     *524288   2001:db8:aaaa:106:8000::               ok
  End.DT2M     *524287   2001:db8:aaaa:106:7fff:f000::          ok
===============================================================================
Legend: * - System allocated
```

EVPN VPLS with VXLAN to SRv6 stitching

Figure 124: Example topology with VPLS-2 shows the example topology for VPLS-2 with VXLAN to SRv6 stitching in the GWs PE-4 and PE-5.

Figure 124: Example topology with VPLS-2



Initial configuration

The initial configuration is similar to the one for EVPN VPLS with MPLS to SRv6 stitching, but LDP is not required between PE-2, PE-3, PE-4, and PE-5. The BGP configuration remains the same with the export policies using route tags to avoid loops.

Service configuration

In the VXLAN domain, all-active multihoming is used between PE-2 and PE-3. The service configuration on PE-2 and PE-3 is identical. Route tag 64500:21 accepted in instance 1 of VPLS-2 on GWs PE-4 and PE-5:

```
#  on PE-2, PE-3:
configure
    service
        system
            bgp-evpn
                ethernet-segment "AA-vES-23-2" virtual create
                    esi 01:00:00:00:00:23:02:00:01:01
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 1
                    dot1q
                        q-tag-range 2
                    exit
                    no shutdown
                    info
                exit
            exit
        exit
        vpls 2 name "VPLS-2" customer 1 create
            vxlan instance 1 vni 2 create
            exit
            bgp
                route-target export target:64500:21 import target:64500:21
            exit
            bgp-evpn
                evi 2
                vxlan bgp 1 vxlan-instance 1
                    ecmp 2
                    auto-disc-route-advertisement        # required for VXLAN MH
                    mh-mode network                      # required for VXLAN MH
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap lag-1:2 create
                no shutdown
            exit
            no shutdown
        exit
```

In the SRv6 domain, no multihoming is used. The configuration of VPLS-2 on PE-6 is as follows:

```
#  on PE-6:
configure
    service
        vpls 2 name "VPLS-2" customer 1 create
            segment-routing-v6 1 create
                locator "PE6-loc"
                    function
                        end-dt2u
                        end-dt2m
                    exit
                exit
            exit
            bgp
                route-target export target:64500:22 import target:64500:22
            exit
            bgp-evpn
                evi 2
                segment-routing-v6 bgp 1 srv6-instance 1 default-locator "PE6-loc" create
                    route-next-hop system-ipv6
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/1/c10/1:2 create
                no shutdown
            exit
            no shutdown
        exit
```

On the GW PEs, service instance 1 uses VXLAN and service instance 2 uses SRv6 transport. VSI policies are used to accept routes with the matching RT and to avoid loops based on the SOO. In the VXLAN service instance, the multihoming mode is access (multihoming mode network is not supported on VXLAN) and no explicit SHG can be configured. For anycast, the same RD and RT values are used on both GW PEs and the originator IP address is 45.45.45.45. The configuration on PE-4 is as follows:

```
# on PE-4:
configure
    router Base
        policy-options
            begin
            community "SOO-45"
                members "origin:45:45"
            exit
            community "RT64500:21"
                members "target:64500:21"
            exit
            community "RT64500:22"
                members "target:64500:22"
            exit
            policy-statement "drop-tag-10"
                description "route tag in VPLSs to avoid loops"
                entry 10
                    from
                        tag 10
                    exit
                    action drop
                    exit
                exit
            exit
            policy-statement "drop-tag-20"
                description "route tag in VPLSs to avoid loops"
                entry 10
                    from
                        tag 20
                    exit
                    action drop
                    exit
                exit
            exit
            policy-statement "vsi-21-export"
```

```
                entry 10
                    action accept
                        community add "RT64500:21" "SOO-45"
                    exit
                exit
            exit
            policy-statement "vsi-21-import"
                entry 10
                    from
                        community "SOO-45"
                        family evpn
                    exit
                    action drop
                    exit
                exit
                entry 20
                    from
                        community "RT64500:21"
                        family evpn
                    exit
                    action accept
                    exit
                exit
            exit
            policy-statement "vsi-22-export"
                entry 10
                    action accept
                        community add "RT64500:22" "SOO-45"
                    exit
                exit
            exit
            policy-statement "vsi-22-import"
                entry 10
                    from
                        community "SOO-45"
                        family evpn
                    exit
                    action drop
                    exit
                exit
                entry 20
                    from
                        community "RT64500:22"
                        family evpn
                    exit
                    action accept
                    exit
                exit
            exit
            commit
        exit
    exit
    service
        vpls 2 name "VPLS-2" customer 1 create
            vxlan instance 1 vni 2 create
            exit
            segment-routing-v6 1 create
                locator "PE4-loc"
                    function
                        end-dt2u
                        end-dt2m
                    exit
                exit
            exit
            bgp
                route-distinguisher 192.0.2.45:2
                vsi-export "vsi-21-export"
                vsi-import "vsi-21-import"
            exit
            bgp 2
                route-distinguisher 192.0.2.54:2
                vsi-export "vsi-22-export"
                vsi-import "vsi-22-import"
            exit
            bgp-evpn
                incl-mcast-orig-ip 45.45.45.45
                evi 2
                vxlan bgp 1 vxlan-instance 1
                    # mh-mode access        # default
                    ecmp 2
                    default-route-tag 10
                    no shutdown
                exit
                segment-routing-v6 bgp 2 srv6-instance 1 default-locator "PE4-loc" create
                    mh-mode network
                    default-route-tag 20
                    route-next-hop system-ipv6
                    no shutdown
                exit
            exit
            no shutdown
```

The configuration on GW PE-5 is identical, but the locator name is different.

Show commands

When traffic is sent between CE-12 and CE-62, the FDB on PE-2 is as follows:

```
*A:PE-2# show service id "VPLS-2" fdb detail

===============================================================================
Forwarding Database, Service 2
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
2          00:00:5e:00:53:12 sap:lag-1:2             L/30     12/21/23 08:49:05
2          00:00:5e:00:53:62 vxlan-1:                Evpn     12/21/23 08:49:15
                             192.0.2.4:2
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

PE-2 receives two EVPN MAC routes for MAC address 00:00:5e:00:53:62 from CE-62, but only the route from GW PE-4 is used while the route from GW PE-5 is not.

```
*A:PE-2# show router bgp routes evpn mac mac-address 00:00:5e:00:53:62
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
```

```
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

================================================================================
BGP EVPN MAC Routes
================================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
--------------------------------------------------------------------------------
u*>i  192.0.2.45:2        00:00:5e:00:53:62 ESI-0
      0                   Seq:0             VNI 2
                          n/a
                          192.0.2.4

*>i   192.0.2.45:2        00:00:5e:00:53:62 ESI-0
      0                   Seq:0             VNI 2
                          n/a
                          192.0.2.5

--------------------------------------------------------------------------------
Routes : 2
================================================================================
```

The GWs redistribute the EVPN MAC routes between the VXLAN domain and the SRv6 domain. On PE-4, the FDB contains entries in the VXLAN domain and in the SRv6 domain. The FDB on PE-5 is similar.

```
*A:PE-4# show service id "VPLS-2" fdb detail

================================================================================
Forwarding Database, Service 2
================================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
--------------------------------------------------------------------------------
2          00:00:5e:00:53:12 eES:                    Evpn     12/21/23 08:49:05
                             01:00:00:00:00:23:02:00:01:01
2          00:00:5e:00:53:62 srv6-1:                 Evpn     12/21/23 08:49:15
                             2001:db8::2:6
           2001:db8:aaaa:106:7fff:e000::
--------------------------------------------------------------------------------
No. of MAC Entries: 2
--------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
================================================================================
```

On PE-4, the VXLAN destinations are the following:

```
*A:PE-4# show service id "VPLS-2" vxlan destinations

================================================================================
Egress VTEP, VNI (Instance 1)
================================================================================
VTEP Address                                        Egress VNI Oper  Mcast Num
                                                               State       MACs
--------------------------------------------------------------------------------
192.0.2.2                                           2          Up    BUM   0
192.0.2.3                                           2          Up    BUM   0
--------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 2
--------------------------------------------------------------------------------
```

```
===============================================================================

===============================================================================
Egress VTEP, VNI (Instance 2)
===============================================================================
VTEP Address                                        Egress VNI Oper  Mcast Num
                                                               State       MACs
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================

===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest (Instance 1)
===============================================================================
Eth SegId                               Num. Macs       Last Update
-------------------------------------------------------------------------------
01:00:00:00:00:23:02:00:01:01           1               12/21/2023 08:49:05
-------------------------------------------------------------------------------
Number of entries: 1
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest (Instance 2)
===============================================================================
Eth SegId                               Num. Macs       Last Update
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

PE-5 is not included in the list of VTEP addresses. Also, remote PEs such as PE-2 set up EVPN destinations to only one PE in the anycast group for VPLS-2:

```
*A:PE-2# show service id "VPLS-2" vxlan destinations

===============================================================================
Egress VTEP, VNI (Instance 1)
===============================================================================
VTEP Address                                        Egress VNI Oper  Mcast Num
                                                               State       MACs
-------------------------------------------------------------------------------
192.0.2.3                                           2          Up    BUM   0
192.0.2.4                                           2          Up    BUM   1
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 2
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
Egress VTEP, VNI (Instance 2)
===============================================================================
VTEP Address                                        Egress VNI Oper  Mcast Num
                                                               State       MACs
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================

===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest (Instance 1)
===============================================================================
Eth SegId                               Num. Macs       Last Update
-------------------------------------------------------------------------------
No Matching Entries
```

```
===============================================================================

===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest (Instance 2)
===============================================================================
Eth SegId                              Num. Macs       Last Update
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

PE-2 receives two IMET routes with originator IP address 45.45.45.45, but only the route with next hop PE-4 is used while the route with next hop PE-5 is not.

```
*A:PE-2# show router bgp routes evpn incl-mcast originator-ip 45.45.45.45
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Flag  Route Dist.         OrigAddr
      Tag                 NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.45:2        45.45.45.45
      0                   192.0.2.4

*>i   192.0.2.45:2        45.45.45.45
      0                   192.0.2.5

-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

The detailed information for the active IMET route is as follows:

```
*A:PE-2# show router bgp routes evpn incl-mcast originator-ip 45.45.45.45 detail
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Original Attributes

Network        : n/a
Nexthop        : 192.0.2.4
Path Id        : None
From           : 192.0.2.4
Res. Nexthop   : 192.168.24.2
Local Pref.    : 100                    Interface Name : int-PE-2-PE-4
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
```

```
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64500:21 origin:45:45 bgp-tunnel-encap:VXLAN
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.4
Origin         : IGP
Flags          : Used Valid Best
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : INCL-MCAST
Tag            : 0
Originator IP  : 45.45.45.45
Route Dist.    : 192.0.2.45:2
Route Tag      : 0
Neighbor-AS    : n/a
DB Orig Val    : N/A                    Final Orig Val : N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h04m04s
-------------------------------------------------------------------------------
PMSI Tunnel Attributes :
Tunnel-type    : Ingress Replication
Flags          : Type: RNVE(0) BM: 0 U: 0 Leaf: not required
MPLS Label     : VNI 2
Tunnel-Endpoint: 192.0.2.4
-------------------------------------------------------------------------------
---snip---
```

In the SRv6 domain, PE-6 receives IMET routes with originator IP address 45.45.45.45 from both GW PEs, but only the IMET route from PE-4 is used:

```
*A:PE-6# show router bgp routes evpn incl-mcast originator-ip 45.45.45.45
===============================================================================
 BGP Router ID:192.0.2.6        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Flag  Route Dist.         OrigAddr
      Tag                 NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.54:2        45.45.45.45
      0                   2001:db8::2:4

*>i   192.0.2.54:2        45.45.45.45
      0                   2001:db8::2:5

-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

The detailed output for the used IMET route from PE-4 is the following:

```
*A:PE-6# show router bgp routes evpn incl-mcast originator-ip 45.45.45.45 detail
===============================================================================
 BGP Router ID:192.0.2.6        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Original Attributes

Network        : n/a
Nexthop        : 2001:db8::2:4
Path Id        : None
From           : 2001:db8::2:4
Res. Nexthop   : fe80::1a:1ff:fe01:1
Local Pref.    : 100                    Interface Name : int-PE-6-PE-4
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64500:22 origin:45:45
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.4
Origin         : IGP
Flags          : Used Valid Best
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : INCL-MCAST
Tag            : 0
Originator IP  : 45.45.45.45
Route Dist.    : 192.0.2.54:2
Route Tag      : 0
Neighbor-AS    : n/a
DB Orig Val    : N/A                    Final Orig Val : N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h06m09s
SRv6 TLV Type  : SRv6 L2 Service TLV (6)
SRv6 SubTLV    : SRv6 SID Information (1)
Sid            : 2001:db8:aaaa:104::
Full Sid       : 2001:db8:aaaa:104:7fff:6000::
Behavior       : End.DT2M (24)
SRv6 SubSubTLV : SRv6 SID Structure (1)
Loc-Block-Len  : 48                     Loc-Node-Len   : 16
Func-Len       : 20                     Arg-Len        : 0
Tpose-Len      : 20                     Tpose-offset   : 64
-------------------------------------------------------------------------------
PMSI Tunnel Attributes :
Tunnel-type    : Ingress Replication
Flags          : Type: RNVE(0) BM: 0 U: 0 Leaf: not required
MPLS Label     : 8388448
Tunnel-Endpoint: 2001:db8::2:4
-------------------------------------------------------------------------------
---snip---
```

PE-6 sets up SRv6 destinations to PE-4, but not to PE-5:

```
*A:PE-6# show service id "VPLS-2" segment-routing-v6 destinations

===============================================================================
TEP, SID (Instance 1)
===============================================================================
TEP Address                Segment Id                                Oper  Mcast Num
```

```
                                                               State       MACs
-------------------------------------------------------------------------------
2001:db8::2:4              2001:db8:aaaa:104:7fff:6000::       Up    BUM   0
2001:db8::2:4              2001:db8:aaaa:104:7fff:7000::       Up    -     1
-------------------------------------------------------------------------------
Number of TEP, SID: 2
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
Segment Routing v6 Ethernet Segment Dest (Instance 1)
===============================================================================
Eth SegId                                 Num. Macs     Last Update
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

Conclusion

SRv6 to MPLS stitching and SRv6 to VXLAN stitching are required to interwork with non-SRv6 networks.

EVPN VPWS Multihoming on PW ports

This chapter provides information about EVPN-VPWS multihoming on PW ports.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

The information and configuration in this chapter are based on SR OS Release 23.10.R1.

Prerequisite reading: EVPN for MPLS Tunnels in Epipe Services (EVPN-VPWS) and EVPN VPWS Services with SRv6 Transport.

Overview

Service providers are migrating active/standby pseudowire (PW) aggregation networks to EVPN-VPWS. This architecture is commonly known as "Pseudowire Headend" architecture. In SR OS, the PW headend PE uses PW ports to map ingress traffic from the access into Layer 2 or Layer 3 services in the core. PW ports provide PW termination with the following characteristics:

- provide SAP-based capabilities to a PW which has traditionally been a network-port-based concept within SR OS. PW payload can be extracted onto PW-port-based SAPs with granular queuing capabilities (per SAP).
- look up dot1q and qinq VLAN tags underneath the PW labels and map the traffic to different services.
- terminate subscriber traffic carried within the PW on a Broadband Network Gateway (BNG): PW-port-based SAPs are instantiated under a group interface with Enhanced Subscriber Management (ESM).

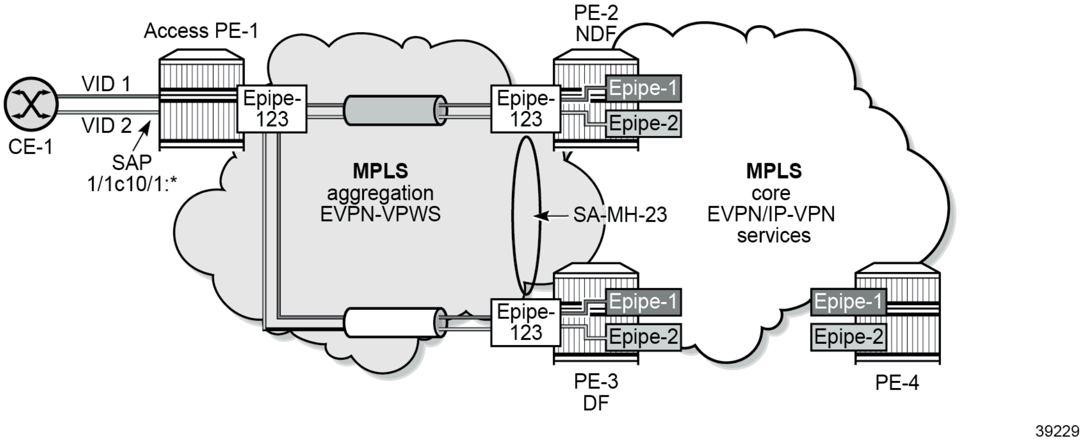
PW ports can operate in two modes:

- PW port bound to a specific physical input/output port (I/O port)
- PW port independent of the I/O port on which the PW is terminated: Forwarding Path Extension (FPE) based PW port

In this chapter, FPE-based PW ports are used. The benefit of FPE-based PW ports is that they can provide services when traffic within the PW is rerouted between I/O ports because of a network failure.

Both all-active and single-active EVPN multihoming modes are supported. Figure 125: EVPN-MPLS single-active multihoming on Epipe PW ports shows the example topology with stitching Epipe "Epipe-123" in the access network and single-active multihoming between PE-2 and PE-3. The single-active Ethernet segment (ES) has a PW port associated to it.

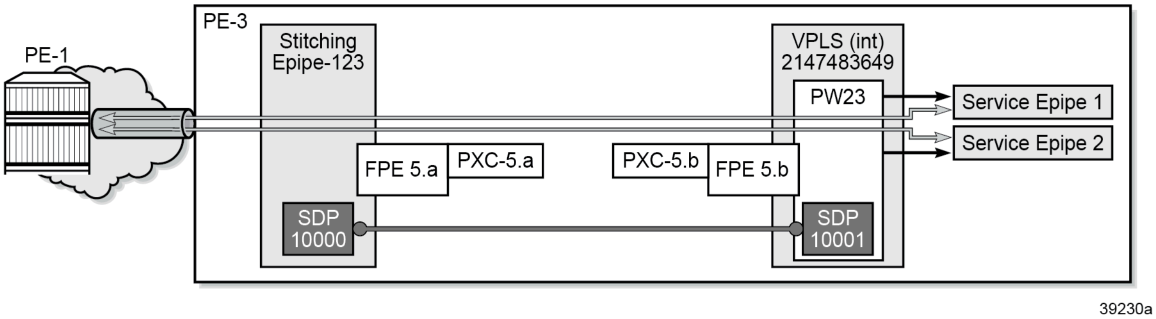
Figure 125: EVPN-MPLS single-active multihoming on Epipe PW ports



In the aggregation network, the stitching EVPN-VPWS "Epipe-123" backhauls the traffic. The SAP of the stitching Epipe on access node PE-1 is 1/1/c10/1:*, so it accepts packets with VID 1 for core service "Epipe-1" and packets with VID 2 for core service "Epipe-2". The stitching between Epipe "Epipe-123" and the service Epipes is done at the Designated Forwarder (DF) PE-3. PE-2 is the Non-Designated Forwarder (NDF) which brings the PW port operationally down due to the MHStandby flag (unless the **oper-up-on-mhstandby** option is enabled). On the NDF PE-2, the PW SAPs pw-23:1 and pw-23:2 (in the core services "Epipe-1" and "Epipe-2") are brought operationally down when the PW port is down. If the PW port is down only due to the MHStandby flag, the AD per-ES route and AD per-EVI route for the service Epipes are still advertised, so PE-1 receives EVPN-AD routes from DF PE-3 and NDF PE-2.

Figure 126: Internal connectivity between switching Epipe and service Epipes shows how internal VPLS "_tmnx_InternalVplsService" with ID 2147483649 is used for the internal cross-connect between the stitching Epipe and the service Epipes.

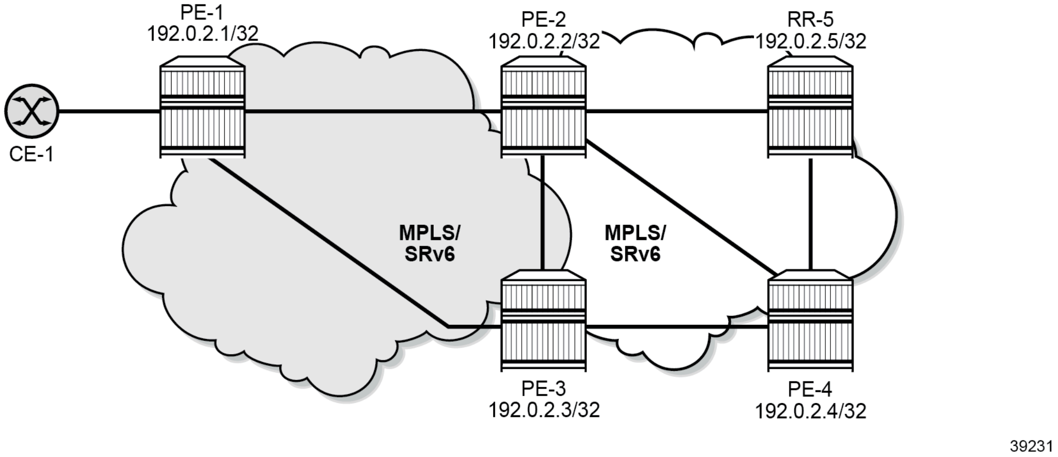
Figure 126: Internal connectivity between switching Epipe and service Epipes



Configuration

Figure 127: Example topology shows the example topology used throughout this section.

Figure 127: Example topology



The initial configuration on the PEs and RR includes:

- cards, MDAs, ports
- router interfaces
- IS-IS on the router interfaces (alternatively, OSPF can be used)

The following scenarios are described in this section:

- EVPN-MPLS all-active multihoming on Epipe PW ports
- EVPN-MPLS single-active multihoming on Epipe PW ports
- EVPN-SRv6 single-active multihoming on Epipe PW ports

In the EVPN-MPLS scenarios, MPLS is used in the aggregation network (between PE-1, PE-2, and PE-3) and in the core network (between PE-2, PE-3, and PE-4). In this example, LDP is applied in the aggregation network and SR-ISIS in the core network.

In the EVPN-MPLS scenarios, the BGP configuration is as follows:

```
# on PE-1, PE-2, PE-3, PE-4:
configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family evpn
                peer-as 64500
                neighbor 192.0.2.5
                exit
            exit
```

In the EVPN-MPLS scenarios, the BGP configuration on RR-5 is the following:

```
# on RR-5:
configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family evpn
                cluster 192.0.2.5
                peer-as 64500
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
            exit
```

In all scenarios, FPE-based PW ports are configured on PE-2 and PE-3, as follows:

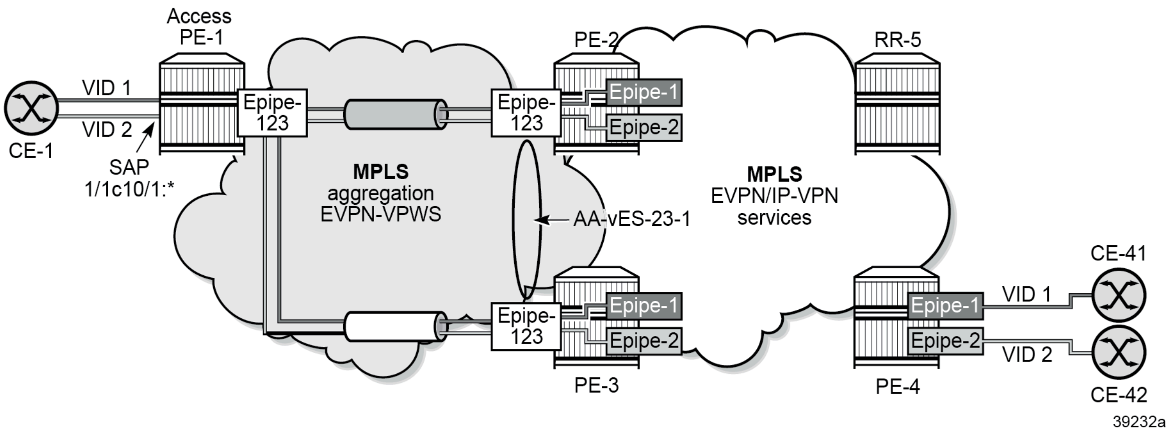
```
# on PE-2, PE-3:
configure
    port-xc
        pxc 5 create
            port 1/1/c5/1
            no shutdown
        exit
    exit
    port pxc-5.a
        ethernet
            encap-type dot1q
        exit
        no shutdown
    exit
    port pxc-5.b
        ethernet
            encap-type dot1q
        exit
        no shutdown
    exit
    port 1/1/c5/1
        ethernet
            dot1x
                tunneling
            exit
            mode hybrid
            encap-type dot1q
        exit
        no shutdown
    exit
    fwd-path-ext
        sdp-id-range from 10000 to 10127
        fpe 5 create
            path pxc 5
            pw-port-extension
                interface-a
                exit
                interface-b
                exit
            exit
        exit
    exit
    pw-port 23 create
        encap-type dot1q    # default
    exit
```

The preceding configuration is similar to the VXLAN FPE configuration, as described in the VXLAN Forwarding Path Extension chapter

EVPN-MPLS all-active multihoming on Epipe PW ports

Figure 128: EVPN-MPLS all-active multi-homing on Epipe PW ports shows the stitching Epipe in the aggregation network and the core Epipe services in an all-active multihoming scenario.

Figure 128: EVPN-MPLS all-active multi-homing on Epipe PW ports



The stitching EVPN-VPWS service is configured on PE-1, PE-2, and PE-3. The configuration on PE-1 is as follows:

```
# on PE-1:
configure
    service
        epipe 123 name "Epipe-123" customer 1 create
            bgp
            exit
            bgp-evpn
                local-attachment-circuit ac-1 create
                    eth-tag 1
                exit
                remote-attachment-circuit ac-23 create
                    eth-tag 23
                exit
                evi 123
                mpls bgp 1
```

```
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            sap 1/1/c10/1:* create
                description "SAP to CEs"
                no shutdown
            exit
            no shutdown
```

CE-1 is connected to this Epipe on PE-1. In this setup, CE-1 is emulated by a VPRN service: on one interface, it sends traffic with VID 1 and an another interface, it sends traffic with VID 2. Both are accepted on SAP 1/1/c10/1:* in EVPN-VPWS "Epipe-123" on PE-1.

The stitching EVPN-VPWS on PE-2 and PE-3 is configured with FPE-based PW port 23, as follows:

```
# on PE-2, PE-3:
configure
    service
        epipe 123 name "stitching-Epipe-123" customer 1 create
            bgp
            exit
            bgp-evpn
                local-attachment-circuit ac-23 create
                    eth-tag 23
                exit
                remote-attachment-circuit ac-1 create
                    eth-tag 1
                exit
                evi 123
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            pw-port 23 fpe 5 create
                no shutdown
            exit
            no shutdown
        exit
```

The aggregation network uses EVPN-VPWS to backhaul the traffic. The network nodes PE-2 and PE-3 apply the same Ethernet tag (23) on the local attachment circuit in the stitching Epipe. Optionally, PE-2 and PE-3 can use the same route distinguisher on the stitching service. AD per-EVI routes for the stitching service are advertised with ESI 0.

The FPE-based PW port is associated to a virtual all-active ES on PE-2 and PE-3. The configuration on PE-2 is as follows; the configuration on PE-3 is identical, but with a different preference value.

```
# on PE-2:
configure
    service
        system
            bgp-evpn
                ethernet-segment "AA-vES-23-1" virtual create
                    esi 01:00:00:00:00:23:00:00:01:01
                    es-activation-timer 3
                    service-carving
```

```
                        mode manual
                        manual
                            preference non-revertive create
                                value 100
                            exit
                        exit
                    exit
                    multi-homing all-active
                    pw-port 23
                    dot1q
                        q-tag-range 1 to 200
                    exit
                    no shutdown
                exit
```

PE-2 and PE-3 receive tagged traffic inside the EVPN-VPWS stitching Epipe and map each tag to a different service in the core, such as ESM services, Epipe services, or VPRN services. In this example, the following services are configured:

```
# on PE-2, PE-3:
configure
    service
        epipe 1 name "service-Epipe-1" customer 1 create
            bgp
            exit
            bgp-evpn
                local-attachment-circuit ac-23 create
                    eth-tag 23
                exit
                remote-attachment-circuit ac-4 create
                    eth-tag 4
                exit
                evi 1
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            sap pw-23:1 create
                no shutdown
            exit
            no shutdown
        exit
        epipe 2 name "service-Epipe-2" customer 1 create
            bgp
            exit
            bgp-evpn
                local-attachment-circuit ac-23 create
                    eth-tag 23
                exit
                remote-attachment-circuit ac-4 create
                    eth-tag 4
                exit
                evi 2
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            sap pw-23:2 create
                no shutdown
            exit
            no shutdown
        exit
```

These Epipe services are also configured on PE-4:

```
# on PE-4:
configure
    service
        epipe 1 name "Epipe-1" customer 1 create
            bgp
            exit
            bgp-evpn
                local-attachment-circuit ac-4 create
                    eth-tag 4
                exit
                remote-attachment-circuit ac-23 create
                    eth-tag 23
                exit
                evi 1
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            sap 1/1/c10/1:1 create
                description "SAP to CE-41"
                no shutdown
            exit
            no shutdown
        exit
        epipe 2 name "Epipe-2" customer 1 create
            bgp
            exit
            bgp-evpn
                local-attachment-circuit ac-4 create
                    eth-tag 4
                exit
                remote-attachment-circuit ac-23 create
                    eth-tag 23
                exit
                evi 2
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            sap 1/1/c10/1:2 create
                description "SAP to CE-42"
                no shutdown
            exit
            no shutdown
        exit
```

Forwarding from CE-1 to CE-41 or CE-42 works as follows:

- Access node PE-1 forwards traffic based on the best AD per-EVI route advertised by PE-2 and PE-3 for the stitching Epipe. This selection can be either BGP-based if PE-2 and PE-3 use the same route distinguisher (RD) in the stitching service or EVPN-based if different RDs are used. The EVPN-based selection when the RDs are different is based on the lowest IP address of the route. In the example, the RDs are auto-derived, such as 192.0.2.2:123, 192.0.2.3:123.

  Note:
  BGP-based selection is also possible when the RDs are different if the command **configure service system bgp-evpn ad-per-evi-routes bgp-path-selection** is configured. For example, in the case of regular BGP best path selection, it is possible to modify the local preference to influence which path is selected.

- When access node PE-1 selects the route to PE-2, PE-2 receives the traffic on the local PW SAP for Epipes "Epipe-1" or "Epipe-2" and forwards it based on the EVPN-VPWS rules in the network to PE-4.

Forwarding from CE-41 or CE-42 to CE-1 works as follows:

- PE-4 forwards the traffic based on the configuration of ECMP and aliasing rules for Epipes "Epipe-1" or "Epipe-2".
- PE-4 may send the traffic to PE-3 and PE-3 to the access node PE-1.

Traffic from the core to the access network may follow an asymmetric path because the multihoming procedures are run on the PW SAPs of the core services, not on the stitching Epipe service. The AD per-EVI routes advertised in the context of the stitching Epipe use ESI 0.

The following command shows that the all-active ES applies to EVI 1 and EVI 2, not to the stitching Epipe with EVI 123:

```
*A:PE-2# show service system bgp-evpn ethernet-segment name "AA-vES-23-1" all

===============================================================================
Service Ethernet Segment
===============================================================================
Name                     : AA-vES-23-1
Eth Seg Type             : Virtual
Admin State              : Enabled            Oper State         : Up
ESI                      : 01:00:00:00:00:23:00:00:01:01
Oper ESI                 : 01:00:00:00:00:23:00:00:01:01
Auto-ESI Type            : None
AC DF Capability         : Include
Multi-homing             : allActive          Oper Multi-homing  : allActive
ES Split Horizon Label   : 524277
ES Split Horizon Arg     : 1
Source BMAC LSB          : None
PW Port Id               : 23
ES Activation Timer      : 3 secs
Oper Group               : (Not Specified)
Svc Carving              : manual             Oper Svc Carving   : manual
Cfg Range Type           : lowest-pref

-------------------------------------------------------------------------------
DF Pref Election Information
-------------------------------------------------------------------------------
Preference     Preference     Last Admin Change        Oper Pref      Do No
Mode           Value                                   Value          Preempt
-------------------------------------------------------------------------------
non-revertive  100            11/28/2023 06:39:53      100            Enabled
-------------------------------------------------------------------------------
EVI Ranges: <none>
ISID Ranges: <none>
```

```
Vprn NextHop EVI Ranges : <none>
===============================================================================

===============================================================================
EVI Information
===============================================================================
EVI                 SvcId               Actv Timer Rem      DF
-------------------------------------------------------------------------------
1                   1                    0                  yes
2                   2                    0                  yes
-------------------------------------------------------------------------------
Number of entries: 2
===============================================================================

-------------------------------------------------------------------------------
DF Candidate list
-------------------------------------------------------------------------------
EVI                                     DF Address
-------------------------------------------------------------------------------
No entries found
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
---snip---
```

Both PE-2 and PE-3 are DF in all-active EVPN-VPWS services "Epipe-1" and "Epipe-2". The following commands are launched on PE-2, but the output on PE-3 is similar.

```
*A:PE-2# show service id 1 ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
pw-23:1               AA-vES-23-1                      DF
===============================================================================
No sdp entries
No vxlan instance entries


*A:PE-2# show service id 2 ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
pw-23:2               AA-vES-23-1                      DF
===============================================================================
No sdp entries
No vxlan instance entries


*A:PE-2# show service id 123 ethernet-segment
No sap entries
No sdp entries
No vxlan instance entries


*A:PE-2# tools dump service system bgp-evpn ethernet-segment "AA-vES-23-1" evi 123 df

[11/28/2023 07:40:39] Evi not active on ethernet-segment
```

```
*A:PE-2# tools dump service system bgp-evpn ethernet-segment "AA-vES-23-1" evi 1 df

[11/28/2023 07:40:39] All Active VPWS or IP-ALIASING - DF N/A

*A:PE-2# tools dump service system bgp-evpn ethernet-segment "AA-vES-23-1" evi 2 df

[11/28/2023 07:40:39] All Active VPWS or IP-ALIASING - DF N/A
```

DF election is not applicable for all-active multihoming in EVPN-VPWS services. For the stitching Epipe, the EVI 123 is not active on the Ethernet segment.

The following VPLS "_tmnx_InternalVplsService" with SDP 10001:100023 on FPE-5.a ensures the internal cross-connect between the stitching Epipe and the core Epipes in PE-2.

```
*A:PE-2# show service id "_tmnx_InternalVplsService" base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 2147483649          Vpn Id            : 0
Service Type      : intVpls
MACSec enabled    : no
Name              : _tmnx_InternalVplsService
Description       : VPLS Service for internal purposes only
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 11/28/2023 06:39:56
Last Mgmt Change  : 11/28/2023 06:39:53
Admin State       : Up                  Oper State        : Up
SAP Count         : 0                   SDP Bind Count    : 1

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sdp:10001:100023 SB(fpe_5.a)             Fpe          0       8910    Up   Up
===============================================================================
```

The following command on PE-2 shows that PW port 23 uses SDP 10001 with VC ID 100023 on SDP binding port pxc-5.b:

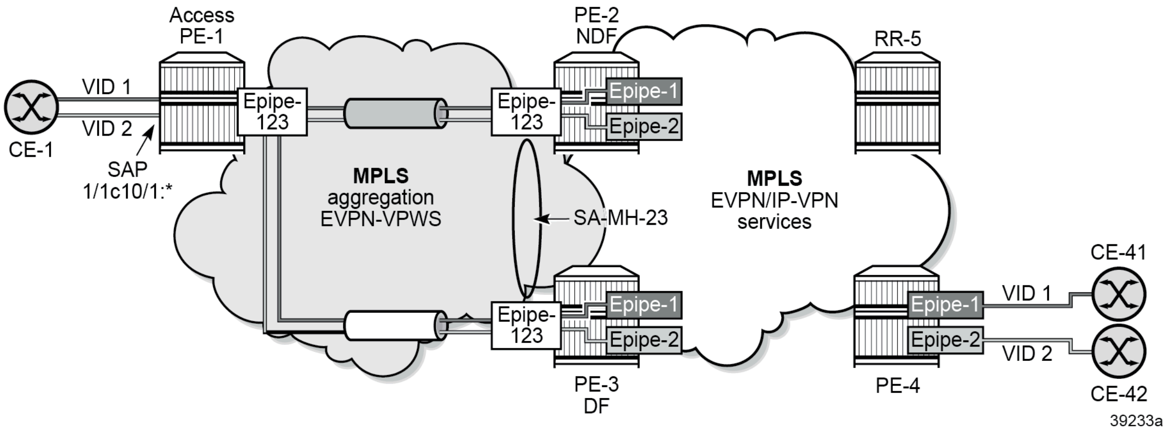
```
*A:PE-2# show pw-port 23 detail

===============================================================================
PW Port Information
===============================================================================

PW Port                           : 23
Encap                             : dot1q
SDP                               : 10001
IfIndex                           : 1526726679
VC-Id                             : 100023
Description                       : PW Port
Dot1Q Ethertype                   : 0x8100
Service Id                        : 123
Down on Peer Tldp PW Status Faults: No
Oper Up on MH Standby             : No
===============================================================================

===============================================================================
Service Destination Point (Sdp Id 10001 Pw-Port 23)
===============================================================================
SDP Binding port      : pxc-5.b
VC-Id                 : 100023                Admin Status       : up
Encap                 : dot1q                 Oper Status        : up
VC Type               : ether
Dot1Q Ethertype       : 0x8100
Control Word          : Not Preferred
Entropy Label         : Disabled
Service MTU           : default

Admin Ingress label   : 524278                Admin Egress label : 524279
Oper Flags            : (Not Specified)
Monitor Oper-Group    : (Not Specified)
===============================================================================
```

EVPN-MPLS single-active multihoming on Epipe PW ports

EVPN-MPLS single-active multihoming support on Epipe PW ports is supported in SR OS Release 22.2.R1 and later.

Figure 125: EVPN-MPLS single-active multihoming on Epipe PW ports shows the EVPN-MPLS single-active multihoming on Epipe PW ports.

Figure 129: EVPN-MPLS single-active multihoming on Epipe PW ports



Service configuration on PE-2 and PE-3

On PE-2 and PE-3, the stitching Epipe is configured as follows:

```
# on PE-2, PE-3:
configure
    service
        epipe 123 name "stitching-Epipe-123" customer 1 create
            bgp
            exit
            bgp-evpn
                local-attachment-circuit ac-23 create
                    eth-tag 23
                exit
                remote-attachment-circuit ac-1 create
                    eth-tag 1
                exit
                evi 123
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            pw-port 23 fpe 5 create
                no shutdown
            exit
            no shutdown
        exit
```

The following (non-virtual) single-active ES is configured on PE-2 and PE-3.

```
# on PE-2, PE-3:
configure
    service
        system
            bgp-evpn
                ethernet-segment "SA-ES-23" create
                    esi 01:00:00:00:00:23:00:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    pw-port 23 pw-headend
                    no shutdown
                exit
            exit
        exit
```

The **pw-headend** keyword allows PW ports to be associated with ESs in single-active mode. The **pw-headend** keyword ensures that the stitching Epipe is running the ES and DF election procedures similar to the mh-mode network in VPLS services. The NDF on the stitching Epipe brings the PW port down with reason MHStandby. The AD per-ES routes and AD per-EVI routes are advertised with the RD and RT of the service Epipe and the configured ESI of the ES associated with the PW port. If the PW port is down only due to MHStandby, the AD per-ES routes and AD per-EVI routes are still advertised. The **oper-up-on-mhstandby** option allows to keep the PW port up on the NDF, which can speed up convergence in case a large number of PW SAPs is configured on the same PW port.

The DF receives tagged traffic inside EVPN-VPWS circuits and maps each tag to a different service in the core network, such as ESM services, Epipe services, or VPRN services. In this example, the following Epipe services are configured with PW SAPs:

```
# on PE-2, PE-3:
configure
    service
        epipe 1 name "service-Epipe-1" customer 1 create
            bgp
            exit
            bgp-evpn
                local-attachment-circuit ac-23 create
                    eth-tag 23
                exit
                remote-attachment-circuit ac-4 create
```

```
                    eth-tag 4
                exit
                evi 1
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            sap pw-23:1 create
                no shutdown
            exit
            no shutdown
        exit
        epipe 2 name "service-Epipe-2" customer 1 create
            bgp
            exit
            bgp-evpn
                local-attachment-circuit ac-23 create
                    eth-tag 23
                exit
                remote-attachment-circuit ac-4 create
                    eth-tag 4
                exit
                evi 2
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            sap pw-23:2 create
                no shutdown
            exit
            no shutdown
        exit
```

The configuration on PE-1 and PE-4 is similar to the configuration in the all-active multihoming scenario.

ES and DF election procedures on stitching Epipe

The stitching Epipe associated with the PW port is running the ES and DF election procedures. The following ES command on PE-3 shows the state for the stitching Epipe EVI 123, not for the contained PW SAP services with EVI 1 and EVI 2. PE-3 is the DF.

```
*A:PE-3# show service system bgp-evpn ethernet-segment name "SA-ES-23" all

===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : SA-ES-23
Eth Seg Type            : None
Admin State             : Enabled            Oper State         : Up
ESI                     : 01:00:00:00:00:23:00:00:00:01
Oper ESI                : 01:00:00:00:00:23:00:00:00:01
Auto-ESI Type           : None
AC DF Capability        : Include
Multi-homing            : singleActive       Oper Multi-homing  : singleActive
ES Split Horizon Label  : None
ES Split Horizon Arg    : None
Source BMAC LSB         : None
PW Port Id              : 23
PW Port Headend         : enabled
ES Activation Timer     : 3 secs
Oper Group              : (Not Specified)
Svc Carving             : auto               Oper Svc Carving   : auto
Cfg Range Type          : primary
Vprn NextHop EVI Ranges : <none>
===============================================================================

===============================================================================
EVI Information
===============================================================================
EVI                 SvcId               Actv Timer Rem      DF
-------------------------------------------------------------------------------
123                 123                 0                   yes
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================

-------------------------------------------------------------------------------
DF Candidate list
-------------------------------------------------------------------------------
EVI                                     DF Address
-------------------------------------------------------------------------------
123                                     192.0.2.2
123                                     192.0.2.3
-------------------------------------------------------------------------------
Number of entries: 2
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
---snip---
```

The following command shows that PE-2 is NDF for the stitching Epipe:

```
*A:PE-2# show service id 123 ethernet-segment
No sap entries
No sdp entries
No vxlan instance entries

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
Pw-Port               Eth-Seg                          Status
-------------------------------------------------------------------------------
23                    SA-ES-23                         NDF
===============================================================================
```

The NDF PE-2 will bring the PW port down because of the MHStandby flag, but the AD per-ES route for the stitching Epipe service is still advertised (if MHStandby is the only reason for the PW port to be down). Therefore, PE-1 receives an AD per-ES route for the stitching Epipe from PE-2 and from PE-3:

```
# on PE-1:
8 2023/11/23 13:35:32.942 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 95
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-AD Len: 25 RD: 192.0.2.2:123 ESI: 01:00:00:00:00:23:00:00:00:01, tag: MAX-ET
 Label: 0 (Raw Label: 0x0) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.2
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        192.0.2.5
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:123
        esi-label:0/Single-Active
        bgp-tunnel-encap:MPLS
"
```

```
13 2023/11/23 13:35:32.947 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 95
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-AD Len: 25 RD: 192.0.2.3:123 ESI: 01:00:00:00:00:23:00:00:00:01, tag: MAX-ET
 Label: 0 (Raw Label: 0x0) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.3
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        192.0.2.5
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:123
        esi-label:0/Single-Active
        bgp-tunnel-encap:MPLS
"
```

Likewise, PE-1 receives an AD per-EVI route from both PE-2 and PE-3. DF PE-3 sends an AD per-EVI with primary bit P: 1, as follows:

```
# on PE-1:
36 2023/11/23 13:35:36.002 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 95
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-AD Len: 25 RD: 192.0.2.3:123 ESI: 01:00:00:00:00:23:00:00:00:01, tag: 23
 Label: 8388496 (Raw Label: 0x7fff90) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.3
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        192.0.2.5
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:123
        l2-attribute:MTU: 1514 F: 0 C: 0 P: 1 B: 0
        bgp-tunnel-encap:MPLS
"
```

NDF PE-2 sends an AD per-EVI with backup bit B: 1, as follows:

```
# on PE-1:
33 2023/11/23 13:35:35.910 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 95
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-AD Len: 25 RD: 192.0.2.2:123 ESI: 01:00:00:00:00:23:00:00:00:01, tag: 23
 Label: 8388480 (Raw Label: 0x7fff80) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.2
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        192.0.2.5
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:123
        l2-attribute:MTU: 1514 F: 0 C: 0 P: 0 B: 1
        bgp-tunnel-encap:MPLS
"
```

NDF behavior

The following command on NDF PE-2 shows that the PW port is operationally down with flag StandbyForMHProtocol:

```
*A:PE-2# show pw-port 23 detail

===============================================================================
PW Port Information
===============================================================================

PW Port                           : 23
Encap                             : dot1q
SDP                               : 10001
IfIndex                           : 1526726679
VC-Id                             : 100023
Description                       : PW Port
Dot1Q Ethertype                   : 0x8100
Service Id                        : 123
Down on Peer Tldp PW Status Faults: No
Oper Up on MH Standby             : No
===============================================================================

===============================================================================
Service Destination Point (Sdp Id 10001 Pw-Port 23)
===============================================================================
SDP Binding port     : pxc-5.b
VC-Id                : 100023                Admin Status       : up
Encap                : dot1q                 Oper Status        : down
VC Type              : ether
Dot1Q Ethertype      : 0x8100
Control Word         : Not Preferred
Entropy Label        : Disabled
Service MTU          : default
```

```
Admin Ingress label  : 524278                Admin Egress label : 524279
Oper Flags           : StandbyForMHProtocol
Monitor Oper-Group   : (Not Specified)
===============================================================================
```

On NDF PE-2, the PW SAPs contained in the PW port are also brought down:

```
*A:PE-2# show service id "service-Epipe-1" sap

===============================================================================
SAP(Summary), Service 1
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
pw-23:1                         1          1     none    1     none   Up   Down
-------------------------------------------------------------------------------
Number of SAPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

```
*A:PE-2# show service id "service-Epipe-2" sap

===============================================================================
SAP(Summary), Service 2
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
pw-23:2                         2          1     none    1     none   Up   Down
-------------------------------------------------------------------------------
Number of SAPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

The following command configures the **oper-up-on-mhstandby** option for the stitching Epipe on NDF PE-2:

```
# on PE-2:
configure
    service
        epipe "stitching-Epipe-123"
            pw-port 23 fpe 5 create
                oper-up-on-mhstandby
                no shutdown
            exit
```

With the **oper-up-on-mhstandby** option enabled, the PW port is operationally up on NDF PE-2:

```
*A:PE-2# show pw-port 23 detail

===============================================================================
PW Port Information
===============================================================================

PW Port                           : 23
Encap                             : dot1q
SDP                               : 10001
IfIndex                           : 1526726679
VC-Id                             : 100023
Description                       : PW Port
```

```
Dot1Q Ethertype                   : 0x8100
Service Id                        : 123
Down on Peer Tldp PW Status Faults: No
Oper Up on MH Standby             : Yes
===============================================================================

===============================================================================
Service Destination Point (Sdp Id 10001 Pw-Port 23)
===============================================================================
SDP Binding port     : pxc-5.b
VC-Id                : 100023                Admin Status       : up
Encap                : dot1q                 Oper Status        : up
VC Type              : ether
Dot1Q Ethertype      : 0x8100
Control Word         : Not Preferred
Entropy Label        : Disabled
Service MTU          : default

Admin Ingress label  : 524278                Admin Egress label : 524279
Oper Flags           : StandbyForMHProtocol
Monitor Oper-Group   : (Not Specified)
===============================================================================
```

Likewise, the PW SAPs in the service Epipes are operationally up:

```
*A:PE-2# show service id "service-Epipe-1" sap

===============================================================================
SAP(Summary), Service 1
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
pw-23:1                         1          1     none    1     none   Up   Up
-------------------------------------------------------------------------------
Number of SAPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

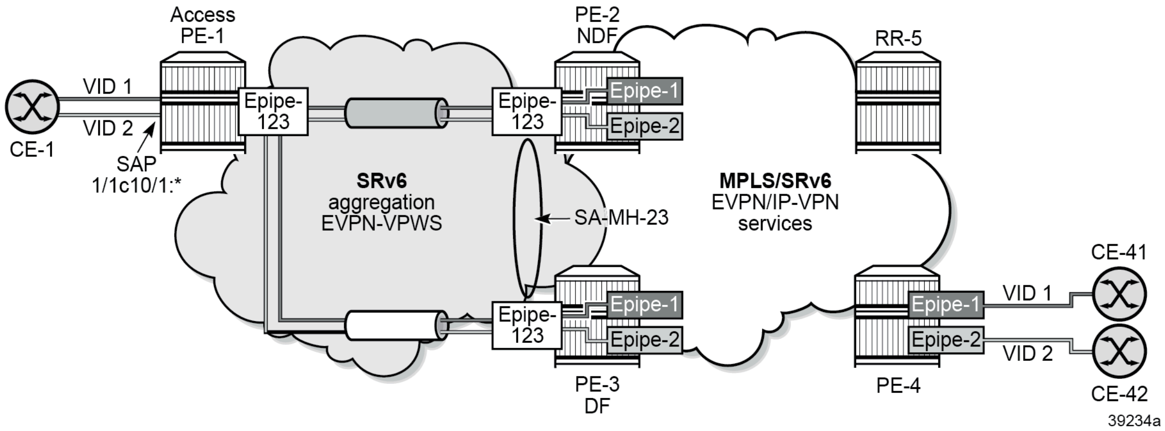
```
*A:PE-2# show service id "service-Epipe-2" sap

===============================================================================
SAP(Summary), Service 2
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
pw-23:2                         2          1     none    1     none   Up   Up
-------------------------------------------------------------------------------
Number of SAPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

EVPN-SRv6 single-active multihoming on Epipe PW ports

EVPN-SRv6 single-active multihoming support on Epipe PW ports is supported in SR OS Release 22.5.R1 and later.

Figure 130: EVPN-SRv6 single-active multihoming on Epipe PW ports shows the topology with EVPN-SRv6 in the aggregation network and both EVPN-MPLS and EVPN-SRv6 in the core network.

Figure 130: EVPN-SRv6 single-active multihoming on Epipe PW ports



The stitching Epipe in the aggregation network uses SRv6 transport. Epipe 1 in the core network uses SRv6 transport, while Epipe 2 in the core network uses MPLS transport.

SRv6 configuration

The SRv6 configuration is as described in the EVPN VPWS Services with SRv6 Transport chapter:

```
# on PE-2:
configure
    card 1
        mda 1
            xconnect
                mac 1 create
                    loopback 1 create
                    exit
                    loopback 2 create
                    exit
                exit
            exit
            no shutdown
            info
        exit
    exit
    port-xc
        pxc 1 create
            port 1/1/m1/1
            no shutdown
        exit
        pxc 2 create
            port 1/1/m1/2
            no shutdown
        exit
    exit
    port pxc-1.a
        ethernet
        exit
        no shutdown
    exit
    port pxc-1.b
        ethernet
        exit
        no shutdown
    exit
    port pxc-2.a
        ethernet
        exit
        no shutdown
    exit
    port pxc-2.b
        ethernet
        exit
        no shutdown
    exit
    port 1/1/m1/1
        no shutdown
    exit
    port 1/1/m1/2
        no shutdown
    exit
    fwd-path-ext
        fpe 1 create
            path pxc 1
            srv6 origination
                interface-a
                exit
                interface-b
                exit
            exit
        exit
        fpe 2 create
            path pxc 2
            srv6 termination
                interface-a
                exit
                interface-b
                exit
            exit
        exit
    exit
    router Base
        segment-routing
            segment-routing-v6
                origination-fpe 1
                source-address 2001:db8::2:2
                locator "loc_Epipe"
                    block-length 48
                    termination-fpe 2
                    prefix
                        ip-prefix 2001:db8:aaaa:102::/64
                    exit
                    static-function
                    exit
                    no shutdown
                exit
                base-routing-instance
                    locator "loc_Epipe"
                        function
                            end-x-auto-allocate srh-mode psp protection unprotected
                            end 1
                                srh-mode usp
                            exit
                        exit
                    exit
                exit
            exit
        exit
        isis 0
            area-id 49.0001
            traffic-engineering
            traffic-engineering-options
                ipv6
                application-link-attributes
                exit
            exit
            advertise-passive-only
            advertise-router-capability as
            ipv6-routing native
            no ipv6-multicast-routing
            segment-routing-v6
                locator "loc_Epipe"
                    level-capability level-2
                    level 1
                    exit
                    level 2
                    exit
                exit
                no shutdown
            exit
```

The configuration on the other PEs is similar. IPv6 addresses are configured on all interfaces. The BGP configuration uses IPv6 addresses, as follows:

```
# on PE-1, PE-2, PE-3, PE-4:
configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family evpn
                peer-as 64500
                neighbor 2001:db8::2:5
                exit
            exit
```

```
# on RR-5:
configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
```

```
                family evpn
                cluster 192.0.2.5
                peer-as 64500
                extended-nh-encoding ipv4 vpn-ipv4
                advertise-ipv6-next-hops evpn
                neighbor 2001:db8::2:1
                exit
                neighbor 2001:db8::2:2
                exit
                neighbor 2001:db8::2:3
                exit
                neighbor 2001:db8::2:4
                exit
            exit
```

Service configuration

On PE-1, the EVPN-VPWS "Epipe-123" is configured as follows:

```
# on access node PE-1:
configure
    service
        epipe 123 name "Epipe-123" customer 1 create
            segment-routing-v6 1 create
                locator "loc_Epipe"
                    function
                        end-dx2
                    exit
                exit
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit ac-1 create
                    eth-tag 1
                exit
                remote-attachment-circuit ac-23 create
                    eth-tag 23
                exit
                evi 123
                segment-routing-v6 bgp 1 srv6-instance 1 default-locator "loc_Epipe" create
                    no shutdown
                exit
            exit
            sap 1/1/c10/1:* create
                description "SAP to CEs"
                no shutdown
            exit
            no shutdown
        exit
```

On PE-2 and PE-3, the stitching Epipe and the single-active ES are configured as follows:

```
# on PE-2, PE-3:
configure
    service
        epipe 123 name "stitching-Epipe-123" customer 1 create
            segment-routing-v6 1 create
                locator "loc_Epipe"
                    function
```

```
                        end-dx2
                    exit
                exit
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit ac-23 create
                    eth-tag 23
                exit
                remote-attachment-circuit ac-1 create
                    eth-tag 1
                exit
                evi 123
                segment-routing-v6 bgp 1 srv6-instance 1 default-locator "loc_Epipe" create
                    no shutdown
                exit
            exit
            pw-port 23 fpe 5 create
                no shutdown
            exit
            no shutdown
        exit
        system
            bgp-evpn
                ethernet-segment "SA-ES-23" create
                    esi 01:00:00:00:00:23:00:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    pw-port 23 pw-headend
                    no shutdown
                exit
            exit
        exit
```

The core service "service-Epipe-1" uses SRv6 transport:

```
# on PE-2, PE-3:
configure
    service
        epipe 1 name "service-Epipe-1" customer 1 create
            segment-routing-v6 1 create
                locator "loc_Epipe"
                    function
                        end-dx2
                    exit
                exit
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit ac-23 create
                    eth-tag 23
                exit
                remote-attachment-circuit ac-4 create
                    eth-tag 4
                exit
                evi 1
                segment-routing-v6 bgp 1 srv6-instance 1 default-locator "loc_Epipe" create
                    no shutdown
                exit
            exit
            sap pw-23:1 create
                no shutdown
            exit
            no shutdown
```

The core service "service-Epipe-2" uses MPLS transport:

```
# on PE-2, PE-3:
configure
    service
        epipe 2 name "service-Epipe-2" customer 1 create
            bgp
            exit
            bgp-evpn
                local-attachment-circuit ac-23 create
                    eth-tag 23
                exit
                remote-attachment-circuit ac-4 create
                    eth-tag 4
                exit
                evi 2
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            sap pw-23:2 create
                no shutdown
            exit
            no shutdown
```

On PE-4, the corresponding Epipe services are configured as follows:

```
# on PE-4:
configure
    service
        epipe 1 name "Epipe-1" customer 1 create
            segment-routing-v6 1 create
                locator "loc_Epipe"
                    function
                        end-dx2
                    exit
                exit
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit ac-1 create
                    eth-tag 4
                exit
                remote-attachment-circuit ac-23 create
                    eth-tag 23
                exit
                evi 1
                segment-routing-v6 bgp 1 srv6-instance 1 default-locator "loc_Epipe" create
                    no shutdown
                exit
            exit
            sap 1/1/c10/1:1 create
```

```
                description "SAP to CE-41"
                no shutdown
            exit
            no shutdown
        exit
        epipe 2 name "Epipe-2" customer 1 create
            bgp
            exit
            bgp-evpn
                local-attachment-circuit ac-1 create
                    eth-tag 4
                exit
                remote-attachment-circuit ac-23 create
                    eth-tag 23
                exit
                evi 2
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            sap 1/1/c10/1:2 create
                description "SAP to CE-42"
                no shutdown
            exit
            no shutdown
        exit
```

Verification

The stitching Epipe associated with the PW port is running the ES and DF election procedures. The following service configuration output shows that the ES is applied for EVI 123 of the stitching Epipe, not for EVI 1 or EVI 2.

```
*A:PE-3# show service system bgp-evpn ethernet-segment name "SA-ES-23" all

===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : SA-ES-23
Eth Seg Type            : None
Admin State             : Enabled            Oper State         : Up
ESI                     : 01:00:00:00:00:23:00:00:00:01
Oper ESI                : 01:00:00:00:00:23:00:00:00:01
Auto-ESI Type           : None
AC DF Capability        : Include
Multi-homing            : singleActive       Oper Multi-homing  : singleActive
ES Split Horizon Label  : None
ES Split Horizon Arg    : None
Source BMAC LSB         : None
PW Port Id              : 23
PW Port Headend         : enabled
ES Activation Timer     : 3 secs
Oper Group              : (Not Specified)
Svc Carving             : auto               Oper Svc Carving   : auto
Cfg Range Type          : primary
Vprn NextHop EVI Ranges : <none>
===============================================================================
===============================================================================
EVI Information
===============================================================================
EVI                 SvcId               Actv Timer Rem      DF
-------------------------------------------------------------------------------
123                 123                 0                   yes
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================

-------------------------------------------------------------------------------
DF Candidate list
-------------------------------------------------------------------------------
EVI                                     DF Address
-------------------------------------------------------------------------------
123                                     192.0.2.2
123                                     192.0.2.3
-------------------------------------------------------------------------------
Number of entries: 2
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
---snip---
```

PE-2 is NDF for the stitching Epipe:

```
*A:PE-2# show service id 123 ethernet-segment
No sap entries
No sdp entries
No vxlan instance entries

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
Pw-Port               Eth-Seg                          Status
-------------------------------------------------------------------------------
23                    SA-ES-23                         NDF
===============================================================================
```

On NDF PE-2, the PW port is operationally down with flag StandbyForMHProtocol:

```
*A:PE-2# show pw-port 23 detail

===============================================================================
PW Port Information
===============================================================================

PW Port                           : 23
Encap                             : dot1q
SDP                               : 10001
IfIndex                           : 1526726679
VC-Id                             : 100023
Description                       : PW Port
Dot1Q Ethertype                   : 0x8100
Service Id                        : 123
Down on Peer Tldp PW Status Faults: No
Oper Up on MH Standby             : No
===============================================================================

===============================================================================
Service Destination Point (Sdp Id 10001 Pw-Port 23)
===============================================================================
SDP Binding port     : pxc-5.b
VC-Id                : 100023                Admin Status       : up
```

```
Encap                : dot1q                 Oper Status        : down
VC Type              : ether
Dot1Q Ethertype      : 0x8100
Control Word         : Not Preferred
Entropy Label        : Disabled
Service MTU          : default

Admin Ingress label  : 524278                Admin Egress label : 524279
Oper Flags           : StandbyForMHProtocol
Monitor Oper-Group   : (Not Specified)
===============================================================================
```

On the NDF, the PW SAPs are also brought down, as follows:

```
*A:PE-2# show service id "service-Epipe-1" sap

===============================================================================
SAP(Summary), Service 1
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
pw-23:1                         1          1     none    1     none   Up   Down
-------------------------------------------------------------------------------
Number of SAPs : 1
-------------------------------------------------------------------------------
===============================================================================


*A:PE-2# show service id "service-Epipe-2" sap

===============================================================================
SAP(Summary), Service 2
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
pw-23:2                         2          1     none    1     none   Up   Down
-------------------------------------------------------------------------------
Number of SAPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

When the PW port is operationally down only due to MHStandby, the NDF still advertises AD per-EVI and AD per-ES routes. The following shows that PE-1 receives two AD per-EVI routes and two AD per-ES routes: one from the DF PE-3 and another one from the NDF PE-2:

```
*A:PE-1# show router bgp routes evpn auto-disc
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Flag  Route Dist.         ESI                           NextHop
      Tag                                               Label
-------------------------------------------------------------------------------
```

```
u*>i  192.0.2.2:123       01:00:00:00:00:23:00:00:00:01 192.0.2.2
      23                                                524281

u*>i  192.0.2.2:123       01:00:00:00:00:23:00:00:00:01 192.0.2.2
      MAX-ET                                            0

u*>i  192.0.2.3:123       01:00:00:00:00:23:00:00:00:01 192.0.2.3
      23                                                524283

u*>i  192.0.2.3:123       01:00:00:00:00:23:00:00:00:01 192.0.2.3
      MAX-ET                                            0

-------------------------------------------------------------------------------
Routes : 4
===============================================================================
```

PE-1 receives the following AD per-ES route with RD 192.0.2.2:123 of the stitching Epipe on PE-2. This AS per-ES route contains:

- an ESI-label extended community with the multihomed mode (single-active) and an ESI label
- an SRv6 L2 service TLV with:
  - an SRv6 SID value of zero (the locator, function, and argument equal zero)
  - the used endpoint behavior code point 0x18 for End.DT2M

```
# on PE-1:
27 2023/11/23 06:34:24.487 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:5
"Peer 1: 2001:db8::2:5: UPDATE
Peer 1: 2001:db8::2:5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 127
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-AD Len: 25 RD: 192.0.2.2:123 ESI: 01:00:00:00:00:23:00:00:00:01, tag: MAX-ET
 Label: 0 (Raw Label: 0x0) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.2
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        192.0.2.5
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:123
        esi-label:3/Single-Active
    Flag: 0xc0 Type: 40 Len: 37 Prefix-SID-attr:
       SRv6 Services TLV (37 bytes):-
           Type: SRV6 L2 Service TLV (6)
           Length: 34 bytes, Reserved: 0x0
         SRv6 Service Information Sub-TLV (33 bytes)
             Type: 1 Len: 30 Rsvd1: 0x0
             SRv6 SID: ::
             SID Flags: 0x0 Endpoint Behavior: 0x18 Rsvd2: 0x0
             SRv6 SID Sub-Sub-TLV
                Type: 1 Len: 6
                BL:0 NL:0 FL:0 AL:0 TL:0 TO:0
"
```

PE-1 receives the following AD per-EVI with RD 192.0.2.3:123 of the stitching Epipe from primary (P: 1) node PE-3:

```
# on PE-1:
31 2023/11/23 06:34:27.499 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:5
"Peer 1: 2001:db8::2:5: UPDATE
Peer 1: 2001:db8::2:5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 127
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-AD Len: 25 RD: 192.0.2.3:123 ESI: 01:00:00:00:00:23:00:00:00:01, tag: 23
 Label: 8388528 (Raw Label: 0x7fffb0) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.3
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        192.0.2.5
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:123
        l2-attribute:MTU: 1514 F: 0 C: 0 P: 1 B: 0
    Flag: 0xc0 Type: 40 Len: 37 Prefix-SID-attr:
       SRv6 Services TLV (37 bytes):-
           Type: SRV6 L2 Service TLV (6)
           Length: 34 bytes, Reserved: 0x0
         SRv6 Service Information Sub-TLV (33 bytes)
             Type: 1 Len: 30 Rsvd1: 0x0
             SRv6 SID: 2001:db8:aaaa:103::
             SID Flags: 0x0 Endpoint Behavior: 0x15 Rsvd2: 0x0
             SRv6 SID Sub-Sub-TLV
                Type: 1 Len: 6
                BL:48 NL:16 FL:20 AL:0 TL:20 TO:64
"
```

PE-1 receives the following AD per-EVI with RD 192.0.2.2:123 of the stitching Epipe from backup (B: 1) node PE-2:

```
# on PE-1:
33 2023/11/23 06:34:27.522 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:5
"Peer 1: 2001:db8::2:5: UPDATE
Peer 1: 2001:db8::2:5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 127
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-AD Len: 25 RD: 192.0.2.2:123 ESI: 01:00:00:00:00:23:00:00:00:01, tag: 23
 Label: 8388496 (Raw Label: 0x7fff90) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.2
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        192.0.2.5
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:123
        l2-attribute:MTU: 1514 F: 0 C: 0 P: 0 B: 1
    Flag: 0xc0 Type: 40 Len: 37 Prefix-SID-attr:
       SRv6 Services TLV (37 bytes):-
           Type: SRV6 L2 Service TLV (6)
           Length: 34 bytes, Reserved: 0x0
         SRv6 Service Information Sub-TLV (33 bytes)
             Type: 1 Len: 30 Rsvd1: 0x0
             SRv6 SID: 2001:db8:aaaa:102::
             SID Flags: 0x0 Endpoint Behavior: 0x15 Rsvd2: 0x0
             SRv6 SID Sub-Sub-TLV
                Type: 1 Len: 6
                BL:48 NL:16 FL:20 AL:0 TL:20 TO:64
"
```

The AD per-EVI routes contain an SRv6 L2 service TLV with:

- an SRv6 SID value of 2001:db8:aaaa:103:: with:
  - block length (BL) 48
  - node length (NL) 16
  - function length (FL) 20
  - argument length (AL) 0
  - transposition length (TL) 20 (for EVPN and IP-VPN) - transposition of 20 bits of the function field to the ESI label field
  - transposition offset (TO) 0
- the used endpoint behavior code point 0x15 for End.DX2

Conclusion

EVPN-VPWS multihoming on PW ports is supported for all-active and for single-active multihoming. The transport on the stitching (and service) Epipe services can be MPLS or SRv6.

EVPN VPWS Services with SRv6 Transport

This chapter provides information about SRv6 support for EVPN-VPWS overlay services.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

The information and configuration in this chapter are based on SR OS Release 22.10.R1. SRv6 support for EVPN-VPWS overlay services is supported on FP-based platforms with FP4-based network ports in SR OS Release 22.7.R1 and later.

Chapter EVPN for MPLS Tunnels is prerequisite reading.

Overview

Service providers prefer an optimized, standardized, and unified control plane for VPNs. EVPN-VPWS is supported in SRv6 networks that may also run other EVPN-based services, such as EVPN-based VPLS services or Layer 3 EVPN IFL (interface-less) services. From a control plane perspective, EVPN-VPWS is a simplified point-to-point version of RFC 7432, because there is no need to advertise MAC/IP advertisement routes in VPWS. EVPN-VPWS is described in RFC 8214, and the signaling aspects to support SRv6 are specified in RFC 9252.

EVPN-VPWS supports all-active multihoming (per-flow load-balancing multihoming) as well as single-active multihoming (per-service load-balancing multihoming), using the same Ethernet segments (ESs) used for EVPN-based VPLS services. EVPN-VPWS uses route type 1 and route type 4; it does not use route types 2, 3, or 5, because MAC/IP routes, inclusive multicast routes, or IP-prefix routes are not required.

EVPN-VPWS uses AD per-EVI routes, and optionally, if multihoming is used, AD per-ES and ES routes are required:

- route type 1 - Auto-discovery per EVPN instance (AD per-EVI). This route type is used in all EVPN-VPWS scenarios, with or without multihoming. For EVPN-VPWS, the Ethernet tag field is encoded with the local attachment circuit (AC) of the advertising PE. This value is configured using the **configure service epipe** <*service-id*> **bgp-evpn local-attachment-circuit** <*ac-name*> **eth-tag** <*tag-value*> command. The route distinguisher (RD), label, and the Ethernet segment identifier (ESI) are encoded as for EVPN-based VPLS. The label field is used as service label. In case of multihoming, AD per-EVI routes containing the same ESI are used to provide aliasing and a backup path to the PEs part of the ES. The L2 MTU field is encoded with the service MTU configured in the Epipe. The flags used for EVPN-VPWS are:

  - Flag C: this flag is set if a control word is configured in the service; however, this does not apply if the transport is SRv6.
  - Flag P: this flag is set if the advertising PE is a primary PE.
    - If no multihoming is used, there is no primary PE (P = 0).
    - In all-active multihoming, all PEs in the ES are primary (P = 1).
    - In single-active multihoming, only one PE per-EVI in the ES is a primary (P = 1).
  - Flag B: this flag is set if the advertising PE is a backup PE.
    - Flag B is only set in case of single-active multihoming and only for one PE, even if more than two PEs are present in the same single-active ES. The backup PE is the winner of the second designated forwarder (DF) election (excluding the DF). The remaining non-DF PEs send B = 0.

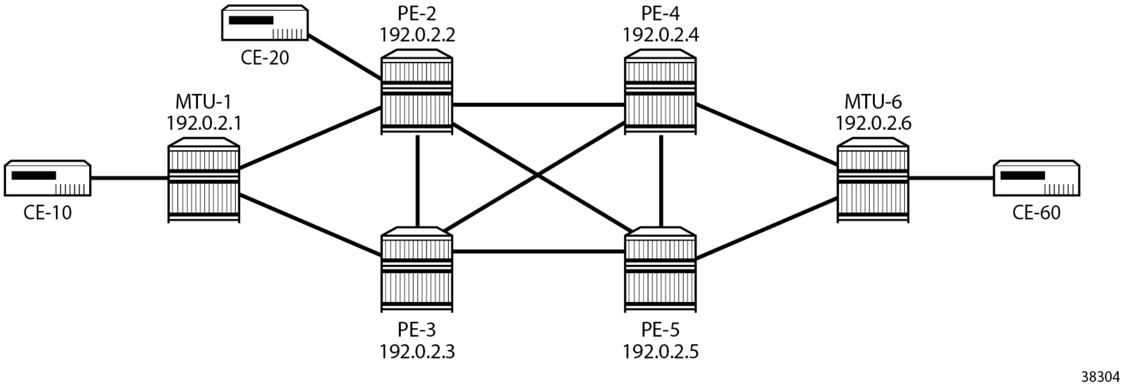
  If there is no multihoming, the ESI, flag P, and flag B are set to zero.

- route type 1 - Auto-discovery per Ethernet segment (AD per-ES). This route type has the same encoding as for EVPN-based VPLS. The AD per-ES route is only used in multihoming scenarios where it is advertised from the PE for each ES. This route type carries the ESI label (used for split-horizon, but only for VPLS services and not for Epipe services) and can affect procedures such as the DF election, as well as the aliasing on remote PEs.
- route type 4 - ES route. This route type has the same encoding as for EVPN-based VPLS. The ES route is only used in multihoming scenarios. This route type advertises a local configured ES. The exchange of this route type can discover remote PEs that are part of the same ES and the DF election algorithm among them.

Configuration

Figure 131: EVPN-VPWS example topology shows the example topology that is used throughout this chapter.

Figure 131: EVPN-VPWS example topology



The example topology consists of six SR OS nodes with the following initial configuration:

- Network (or hybrid) ports interconnect the core PEs with configured router interfaces.

- MTU-1 is a pure Ethernet aggregator. The ports toward the core PEs are access ports. Likewise, the ports on PE-2 and PE-3 toward MTU-1 are access ports.
- Core PEs and MTU-6 run IS-IS on all interfaces.
- Link LDP is configured between all PEs, and toward and from MTU-6.
- EVPN uses BGP for exchanging reachability information at the service level. Therefore, BGP peering sessions must be established among the core PEs for the EVPN family. Although a separate router is typically used, in this chapter, PE-2 is used as route reflector with the following BGP configuration:

```
*A:PE-2# configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "gr_v6_internal"
                family evpn
                cluster 1.1.1.1
                peer-as 64500
                extended-nh-encoding ipv4 vpn-ipv4
                advertise-ipv6-next-hops evpn
                neighbor 2001:db8::2:3
                exit
                neighbor 2001:db8::2:4
                exit
                neighbor 2001:db8::2:5
                exit
            exit
        exit all
```

  The BGP configuration on the other PEs is as follows:

```
*A:PE-3#, *A:PE-4#, *A:PE-5# configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "gr_v6_internal"
                family evpn
                peer-as 64500
                extended-nh-encoding ipv4 vpn-ipv4
                advertise-ipv6-next-hops evpn
                neighbor 2001:db8::2:2
                exit
            exit
        exit all
```

The following sections describe the EVPN-VPWS scenarios:

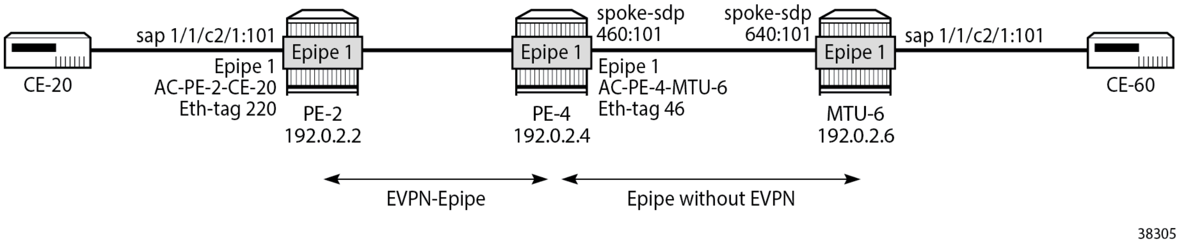
- SRv6 tunnels in EVPN-VPWS services without multihoming
- SRv6 tunnels in EVPN-VPWS services with all-active multihoming

- SRv6 tunnels in EVPN-VPWS services with single-active multihoming

SRv6 tunnels in EVPN-VPWS services without multihoming

BGP-EVPN can be enabled in Epipe services with either SAPs or spoke SDPs at the access, as shown in Figure 132: Example topology for EVPN-VPWS without multihoming.

Figure 132: Example topology for EVPN-VPWS without multihoming



On PE-2, Epipe 1 is configured as follows:

```
*A:PE-2# configure
    service
        epipe 1 name "Epipe-1" customer 1 create
            segment-routing-v6 1 create
                locator "loc_Epipe-1"
                    function
                        end-dx2
                    exit
                exit
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-PE-2-CE-20 create
                    eth-tag 220
                exit
                remote-attachment-circuit AC-PE-4-MTU-6 create
                    eth-tag 46
                exit
                evi 10
                segment-routing-v6 bgp 1 srv6-instance 1 default-locator "loc_Epipe-1" create
                    # source-address 2001:db8::2:2    # defined for SRv6 on router level
                    no shutdown
                exit
            exit
            sap 1/1/c2/1:101 create
                no shutdown
            exit
            no shutdown
        exit all
```

On PE-4, the service configuration is as follows:

```
*A:PE-4# configure
    service
        sdp 460 create
            far-end 192.0.2.6
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        epipe 1 name "Epipe-1" customer 1 create
            segment-routing-v6 1 create
                locator "loc_Epipe-1"
                    function
                        end-dx2
                    exit
                exit
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-PE-4-MTU-6 create
                    eth-tag 46
                exit
                remote-attachment-circuit AC-PE-2-CE-20 create
                    eth-tag 220
                exit
                evi 10
                segment-routing-v6 bgp 1 srv6-instance 1 default-locator "loc_Epipe-1" create
                    # source-address 2001:db8::2:4    # defined for SRv6 on router level
                    no shutdown
                exit
            exit
            spoke-sdp 460:101 create
                no shutdown
            exit
            no shutdown
        exit all
```

The following commands are relevant for the EVPN-VPWS configuration:

- the **bgp** command enables the context for the BGP configuration relevant to the service. The **bgp** context configures the common BGP parameters for all BGP families in the service, such as the RD and the route target (RT). Even if the general BGP parameters for the service are auto-derived, the **bgp** context must be enabled.

```
*A:PE-2# configure service epipe 1 bgp ?
  - bgp
  - no bgp

 [no] adv-service-mtu - Configure service-mtu to be advertised
 [no] pw-template-bi* + Configure pw-template bind policy
 [no] route-distingu* - Configure route distinguisher
 [no] route-target    - Configure route target
 [no] vsi-export      - VSI export route policies
 [no] vsi-import      - VSI import route policies
```

- The following commands can be configured in the **bgp-evpn** context:

```
*A:PE-2# configure service epipe 1 bgp-evpn ?
  - bgp-evpn
  - no bgp-evpn

 [no] evi             - EVPN Identifier
 [no] local-attachme* + Configure local attachment circuit information
 [no] mpls            + Configure BGP EVPN mpls
 [no] remote-attachm* + Configure remote attachment circuit information
```

```
[no] segment-routin* + Configure SRv6 instance
[no] vxlan           + Configure BGP EVPN vxlan
```

- The **evi** command configures a 2-byte or 3-byte EVPN identifier (EVI) used for auto-deriving the service RD, service RT, and for the service carving (or DF election) when multihoming is used. For 2-byte EVIs, the auto-derivation of RD and RT is as follows:

  - RD system-ip:evi
  - RT autonomous-system:evi

  The EVI values must be unique in the system, regardless of the type of service they are assigned to (Epipe or VPLS).

- The **local-attachment-circuit** and **remote-attachment-circuit** commands configure the two attachment circuits connected by the EVPN-VPWS service. The configured Ethernet tag for the local AC is advertised in the Ethernet tag field of the AD per-EVI route for the Epipe, along with the corresponding RD, RT, and label. Both local and remote Ethernet tags are necessary to bring up the Epipe service. If the received Ethernet tag for the Epipe service matches the configured remote AC Ethernet tag, an EVPN-SRv6 destination is created to the next hop.

  The local Ethernet tag cannot be modified without disabling **bgp-evpn segment-routing-v6** in the Epipe, as shown in the following output:

  ```
  *A:PE-2# configure service epipe "Epipe-1" bgp-evpn local-attachment-circuit AC-PE-2-CE-
  20 eth-tag 221
  MINOR: SVCMGR #8036 evpn-vpws ac eth-tag not allowed - cannot change while evpn mpls/
  vxlan/srv6 is enabled
  ```

  Unlike local Ethernet tags, remote Ethernet tags can be modified without disabling bgp-evpn.

- The following configuration options are available for Epipes in the **configure service epipe 1 bgp-evpn segment-routing-v6** context:

  ```
  *A:PE-2# configure service epipe 1 bgp-evpn segment-routing-v6 ?
    - no segment-routing-v6 [bgp <bgp-instance>]
    - segment-routing-v6 [bgp <bgp-instance>] [srv6-instance <[1..1]>] [default-locator
      <name>] [create]

   <bgp-instance>       : [1..1]
   <name>               : [64 chars max]
   <create>             : keyword


   [no] default-route-* - Configure default-route-tag to match against export policies
        ecmp            - Configure maximum ECMP routes information
   [no] evi-three-byte* - Enable/Disable evi-three-byte-auto-rt
   [no] force-qinq-vc-* - Forces qinq-vc-type forwarding in the data-path
   [no] force-vlan-vc-* - Forces vlan-vc-type forwarding in the data-path
   [no] oper-group      - Configure oper-group
        resolution      - Configure route resolution options
        route-next-hop  - Configure route next-hop
   [no] shutdown        - Enable/disable SRV6
   [no] source-address  - Configure source IPv6 address
  ```

  This output shows a subset of the options for VPLS services; see chapter EVPN for MPLS Tunnels for a longer list of options.

When the local AC (sap 1/1/c2/1:101) is up, PE-2 sends a BGP EVPN AD per-EVI route that contains Ethernet tag 220 for the local AC:

```
# on PE-2:
4 2022/11/30 09:46:56.704 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:4
"Peer 1: 2001:db8::2:4: UPDATE
Peer 1: 2001:db8::2:4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 113
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-AD Len: 25 RD: 192.0.2.2:10 ESI: ESI-0, tag: 220 Label: 8388448 (Raw Label:
 0x7fff60) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:10
        l2-attribute:MTU: 1514 C: 0 P: 0 B: 0
    Flag: 0xc0 Type: 40 Len: 37 Prefix-SID-attr:
       SRv6 Services TLV (37 bytes):-
           Type: SRV6 L2 Service TLV (6)
           Length: 34 bytes, Reserved: 0x0
         SRv6 Service Information Sub-TLV (33 bytes)
             Type: 1 Len: 30 Rsvd1: 0x0
             SRv6 SID: 2001:db8:aaaa:102::
             SID Flags: 0x0 Endpoint Behavior: 0x15 Rsvd2: 0x0
             SRv6 SID Sub-Sub-TLV
                Type: 1 Len: 6
                BL:48 NL:16 FL:20 AL:0 TL:20 TO:64
"
```

The auto-derived RD is 192.0.2.2:10 and the RT is 64500:10.

When the remote AC on PE-4 (spoke sdp 460:101) is up, PE-2 receives the following BGP update from PE-4:

```
# on PE-2:
5 2022/11/30 09:47:19.837 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:4
"Peer 1: 2001:db8::2:4: UPDATE
Peer 1: 2001:db8::2:4 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 113
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.4
        Type: EVPN-AD Len: 25 RD: 192.0.2.4:10 ESI: ESI-0, tag: 46 Label: 8388448 (Raw Label:
 0x7fff60) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:10
        l2-attribute:MTU: 1514 C: 0 P: 0 B: 0
    Flag: 0xc0 Type: 40 Len: 37 Prefix-SID-attr:
       SRv6 Services TLV (37 bytes):-
           Type: SRV6 L2 Service TLV (6)
           Length: 34 bytes, Reserved: 0x0
         SRv6 Service Information Sub-TLV (33 bytes)
             Type: 1 Len: 30 Rsvd1: 0x0
             SRv6 SID: 2001:db8:aaaa:104::
             SID Flags: 0x0 Endpoint Behavior: 0x15 Rsvd2: 0x0
             SRv6 SID Sub-Sub-TLV
                Type: 1 Len: 6
                BL:48 NL:16 FL:20 AL:0 TL:20 TO:64
"
```

When the received RT matches and the received Ethernet tag matches the configured remote AC Ethernet tag, the EVPN-SRv6 destination, which consists of a termination endpoint (TEP) and a SID) is created on PE-2 and PE-4:

```
*A:PE-2# show service id 1 segment-routing-v6 instance 1 destinations

===============================================================================
TEP, SID
===============================================================================
Instance  TEP Address                        Segment Id
-------------------------------------------------------------------------------
1         192.0.2.4                          2001:db8:aaaa:104:7fff:6000::
-------------------------------------------------------------------------------
Number of TEP, SID: 1
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
Segment Routing v6 Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

Note:
The egress label for the EVPN-SRv6 destination on PE-4 is 524278. The 24-bit label value in the BGP update debug is 16 (2^4) times as high:

524 278 * 16 = 8 388 448

because the debug message is shown before the router can parse the label field and determine if it corresponds to an MPLS label or a transposed function (20 bits), or to a VXLAN VNI (24 bits).

The BGP AD per-EVI routes for Ethernet tag 46 are shown with the following command:

```
*A:PE-2# show router bgp routes evpn auto-disc tag 46
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Flag  Route Dist.         ESI                           NextHop
      Tag                                               Label
-------------------------------------------------------------------------------
u*>i  192.0.2.4:10        ESI-0                         192.0.2.4
      46                                                524278

-------------------------------------------------------------------------------
Routes : 1
```

```
===============================================================================
```

The following command shows the BGP EVPN information for Epipe 1:

```
*A:PE-2# show service id 1 bgp-evpn

===============================================================================
BGP EVPN Table
===============================================================================
EVI                : 10                Creation Origin    : manual

-------------------------------------------------------------------------------
Local AC Name              Eth Tag  Endpoint                    Ingress Label
-------------------------------------------------------------------------------
AC-PE-2-CE-20              220                                  0
-------------------------------------------------------------------------------
Number of local ACs : 1

-------------------------------------------------------------------------------
Remote AC Name                     Eth Tag  Endpoint
-------------------------------------------------------------------------------
AC-PE-4-MTU-6                      46
-------------------------------------------------------------------------------
Number of Remote ACs : 1
===============================================================================

===============================================================================
Segment Routing v6 Instance 1 Service 1
===============================================================================
Admin State             : Enabled
Srv6 Instance           : 1
Default Locator         : loc_Epipe-1

Oper Group              : (Not Specified)
Default Route Tag       : 0x0
Source Address          : (Not Specified)
ECMP                    : 1
Force Vlan VC Fwd       : disabled
Next Hop Type           : system-ipv4
Evi 3-byte Auto-RT      : disabled
Route Resolution        : route-table
Force QinQ VC Fwd       : none
MH Mode                 : network
===============================================================================
```

Note:
Each PE sends its service MTU into the L2 MTU field in the l2-attribute in the AD per-EVI route for the Epipe service. The received L2 MTU is checked. In case of a mismatch between the received MTU and the configured service MTU, the router does not set up the EVPN destination and, therefore, the service does not come up.

SRv6 tunnels in EVPN-VPWS services with multihoming

SR OS supports EVPN multihoming as per RFC 8214.

The EVPN multihoming implementation is based on the concept of the ES. An ES is a logical structure that can be defined in one or more PEs and identifies the CE (or access network) multihoming to the EVPN PEs. An ES is associated with a port, LAG, or SDP object, and is shared by all the services defined on those objects. It can also be shared between Epipe and VPLS services.

Each ES has a unique ESI that is 10 bytes and is manually configured. The ESI is advertised in the control plane to all the PEs in an EVPN network; therefore, it is very important to ensure that the 10-byte ESI value is unique throughout the entire network. Single-homing CEs are assumed to be connected to an ES with ESI = 0 (single-homing ESs are not explicitly configured).

The ES is part of the base BGP-EVPN configuration and is not applied to any EVPN-based VPLS service by default. An ES can be shared by multiple services; a specific SAP or spoke SDP is automatically associated with an ES when the SAP is defined in the same LAG or port configured in the ES, or when the spoke SDP is defined in the same SDP configured in the ES.

Regardless of the multihoming mode, the local Ethernet tag values must match on all the PEs that are part of the same ES. The PEs in the ES use the AD per-EVI routes from the peer PEs to validate the PEs as DF election candidates for an EVI. The DF election is only relevant for single-active multihoming ESs. For Epipes defined in an all-active multihoming ES, there is no DF election required, because all PEs are forwarding traffic and all traffic is treated as unicast.

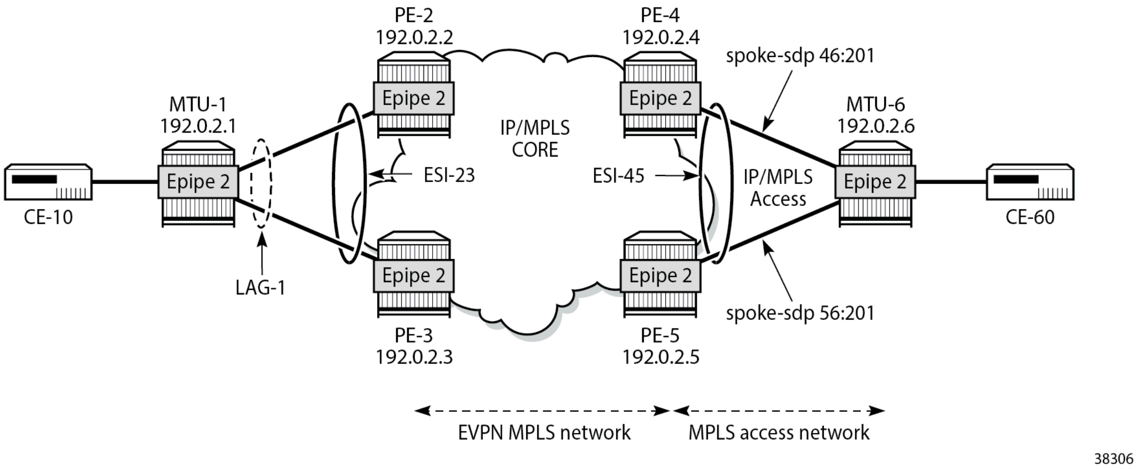
Aliasing is supported when sending traffic to an ES destination. Assuming ECMP is enabled on the ingress PE (and shared queuing or ingress policing are configured), per-flow load-balancing is performed among all the PEs that advertise P = 1. PEs advertising P = 0 are not considered as next hops for an ES destination.

The following sections show the configuration of:

- an all-active multihoming ES with a LAG associated with it
- a single-active multihoming ES linked to an SDP

Figure 133: Example topology EVPN-VPWS with multihoming shows the example topology has an all-active multihoming ES "ESI-23" with a LAG associated with it in PE-2 and PE-3. A single-active multihoming ES "ESI-45" with an SDP associated with it is configured in PE-4 and PE-5.

Figure 133: Example topology EVPN-VPWS with multihoming



SRv6 tunnels in EVPN-VPWS services with all-active multihoming

All-active multihoming allows for per-flow load-balancing. Unlike EVPN-based VPLS services, EVPN-VPWS has no DF election in all-active multihoming. All PEs in the ES are active and the remote PE performs per-flow load-balancing. ESI-23 is configured on PE-2 and PE-3 as all-active multihoming and is associated with LAG 1. This LAG is used as a SAP in Epipe 2 on both PE-2 and PE-3. The configuration of the ES and Epipe 2 is identical on PE-2 and PE-3, including the local AC and remote AC names and Ethernet tags:

```
*A:PE-2#, *A:PE-3# configure
    service
        system
            bgp-evpn
                ethernet-segment "ESI-23" create
                    esi 01:00:00:00:00:23:00:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 1
                    no shutdown
                exit
            exit
        exit
        epipe 2 name "Epipe-2" customer 1 create
            segment-routing-v6 1 create
                locator "loc_Epipe-2"
                    function
                        end-dx2
                    exit
                exit
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-ESI-23-MTU-1 create
                    eth-tag 231
                exit
                remote-attachment-circuit AC-ESI-45-MTU-6 create
                    eth-tag 456
                exit
                evi 20
                segment-routing-v6 bgp 1 srv6-instance 1 default-locator "loc_Epipe-2" create
                    ecmp 2
                    no shutdown
                exit
            exit
            sap lag-1:201 create
                no shutdown
            exit
            no shutdown
        exit
    exit all
```

See chapter EVPN for MPLS Tunnels for a detailed explanation of the configuration parameters of the ES.

In EVPN-VPWS multihoming scenarios, three route types are exchanged: AD per-EVI, AD per-ES, and ES routes. The following ES route (route type 4) for ESI 01:00:00:00:00:23:00:00:00:01, sent by PE-2, is imported at PE-3:

```
# on PE-3:
8 2022/11/30 10:02:59.056 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:2
"Peer 1: 2001:db8::2:2: UPDATE
Peer 1: 2001:db8::2:2 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 71
    Flag: 0x90 Type: 14 Len: 34 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-ETH-SEG Len: 23 RD: 192.0.2.2:0 ESI: 01:00:00:00:00:23:00:00:00:01, IP-Len:
 4 Orig-IP-Addr: 192.0.2.2
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        df-election::DF-Type:Auto/DP:0/DF-Preference:0/AC:1
        target:00:00:00:00:23:00
"
```

The target 00:00:00:00:23:00 in the extended community is derived from the ESI (bytes 2 to 7) and is only imported by the PEs that are part of the same ES; that is, PE-2 and PE-3 in this example.

At the same time, the following AD per-ES route (route type 1) with maximum Ethernet (MAX-ET) tag (all Fs) and label 0 is sent by RR PE-2 and imported by the rest of the PEs. The following two BGP updates with MAX-ET are received by PE-4:

```
# on PE-4:
15 2022/11/30 10:03:42.705 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:2
"Peer 1: 2001:db8::2:2: UPDATE
Peer 1: 2001:db8::2:2 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 113
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-AD Len: 25 RD: 192.0.2.2:20 ESI: 01:00:00:00:00:23:00:00:00:01, tag: MAX-ET
 Label: 0 (Raw Label: 0x0) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:20
        esi-label:3/All-Active
    Flag: 0xc0 Type: 40 Len: 37 Prefix-SID-attr:
       SRv6 Services TLV (37 bytes):-
           Type: SRV6 L2 Service TLV (6)
           Length: 34 bytes, Reserved: 0x0
         SRv6 Service Information Sub-TLV (33 bytes)
             Type: 1 Len: 30 Rsvd1: 0x0
                Type: 1 Len: 6
                BL:0 NL:0 FL:0 AL:0 TL:0 TO:0
"
```

```
13 2022/11/30 10:03:42.705 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:2
"Peer 1: 2001:db8::2:2: UPDATE
Peer 1: 2001:db8::2:2 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 127
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-AD Len: 25 RD: 192.0.2.3:20 ESI: 01:00:00:00:00:23:00:00:00:01, tag: MAX-ET
 Label: 0 (Raw Label: 0x0) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.3
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        1.1.1.1
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:20
        esi-label:3/All-Active
    Flag: 0xc0 Type: 40 Len: 37 Prefix-SID-attr:
       SRv6 Services TLV (37 bytes):-
           Type: SRV6 L2 Service TLV (6)
                Type: 1 Len: 6
                BL:0 NL:0 FL:0 AL:0 TL:0 TO:0
"
```

The ESI label is in the extended community, as well as the indication that the multihoming is all-active. Epipe services do not require ESI labels because BUM traffic is not recognized in EVPN-VPWS services. However, because the ES can be shared by Epipe and VPLS services, the AD per-ES route still includes a non-zero ESI label. In this case, the transport is SRv6, so there are no ESI labels. The label field in the ESI-label extended community is an implicit-null value (3) and the included SRv6 Services TLV encodes a SID with value 0.

The following two AD per-EVI routes (route type 1) with Ethernet tag 231 sent by RR PE-2 are received and imported on PE-4:

```
# on PE-4:
14 2022/11/30 10:03:42.705 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:2
"Peer 1: 2001:db8::2:2: UPDATE
Peer 1: 2001:db8::2:2 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 113
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-AD Len: 25 RD: 192.0.2.2:20 ESI: 01:00:00:00:00:23:00:00:00:01, tag: 231
 Label: 8388432 (Raw Label: 0x7fff50) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:20
        l2-attribute:MTU: 1514 C: 0 P: 1 B: 0
    Flag: 0xc0 Type: 40 Len: 37 Prefix-SID-attr:
       SRv6 Services TLV (37 bytes):-
           Type: SRV6 L2 Service TLV (6)
           Length: 34 bytes, Reserved: 0x0
         SRv6 Service Information Sub-TLV (33 bytes)
             Type: 1 Len: 30 Rsvd1: 0x0
                Type: 1 Len: 6
                BL:48 NL:16 FL:20 AL:0 TL:20 TO:64
"
```

```
12 2022/11/30 10:03:42.705 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:2
"Peer 1: 2001:db8::2:2: UPDATE
Peer 1: 2001:db8::2:2 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 127
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-AD Len: 25 RD: 192.0.2.3:20 ESI: 01:00:00:00:00:23:00:00:00:01, tag: 231
 Label: 8388432 (Raw Label: 0x7fff50) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
```

```
     Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.3
     Flag: 0x80 Type: 10 Len: 4 Cluster ID:
         1.1.1.1
     Flag: 0xc0 Type: 16 Len: 16 Extended Community:
         target:64500:20
         l2-attribute:MTU: 1514 C: 0 P: 1 B: 0
     Flag: 0xc0 Type: 40 Len: 37 Prefix-SID-attr:
        SRv6 Services TLV (37 bytes):-
            Type: SRV6 L2 Service TLV (6)
                 Type: 1 Len: 6
                 BL:48 NL:16 FL:20 AL:0 TL:20 TO:64
"
```

This route type contains the flags for control word (C), primary (P), and backup (B). In all-active multihoming, all nodes are primary (P = 1).

PE-4 learns AD per-EVI and AD per-ES routes for ESI-23 from PE-2 and PE-3, as shown in the following output:

```
*A:PE-4# show router bgp routes evpn auto-disc esi 01:00:00:00:00:23:00:00:00:01
===============================================================================
 BGP Router ID:192.0.2.4        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Flag  Route Dist.         ESI                           NextHop
      Tag                                               Label
-------------------------------------------------------------------------------
u*>i  192.0.2.2:20        01:00:00:00:00:23:00:00:00:01 192.0.2.2
      231                                               524277

u*>i  192.0.2.2:20        01:00:00:00:00:23:00:00:00:01 192.0.2.2
      MAX-ET                                            0

u*>i  192.0.2.3:20        01:00:00:00:00:23:00:00:00:01 192.0.2.3
      231                                               524277

u*>i  192.0.2.3:20        01:00:00:00:00:23:00:00:00:01 192.0.2.3
      MAX-ET                                            0

-------------------------------------------------------------------------------
Routes : 4
===============================================================================
```

For Epipe 2 on PE-4, the EVPN VPWS destination is not pointing at a specific TEP, but at ESI-23, as shown in the following output:

```
*A:PE-4# show service id 2 segment-routing-v6 instance 1 destinations

===============================================================================
TEP, SID
===============================================================================
Instance  TEP Address                       Segment Id
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
===============================================================================
Segment Routing v6 Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                        Num. Macs     Last Change
-------------------------------------------------------------------------------
1         01:00:00:00:00:23:00:00:00:01    0             11/30/2022 10:03:43
-------------------------------------------------------------------------------
Number of entries: 1
-------------------------------------------------------------------------------
===============================================================================
```

When ECMP is greater than 1 on the ingress PE, multiple TEPs can correspond to a specific ESI (aliasing). In this case, ECMP = 2 and PE-4 and PE-5 have two TEP addresses and SIDs for ESI 01:00:00:00:00:23:00:00:00:01, as shown for PE-4:

```
*A:PE-4# show service id 2 segment-routing-v6 esi 01:00:00:00:00:23:00:00:00:01

===============================================================================
Segment Routing v6 Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                        Num. Macs     Last Change
-------------------------------------------------------------------------------
1         01:00:00:00:00:23:00:00:00:01    0             11/30/2022 10:03:43
-------------------------------------------------------------------------------
Number of entries: 1
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
Segment Routing v6 Dest TEP Info
===============================================================================
Instance  TEP Address                    Segment Id          Last Change
-------------------------------------------------------------------------------
1         192.0.2.2                      2001:db8:aaaa:202:* 11/30/2022 10:03:43
1         192.0.2.3                      2001:db8:aaaa:203:* 11/30/2022 10:03:43
-------------------------------------------------------------------------------
Number of entries : 2
-------------------------------------------------------------------------------
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

Note:
Even if ECMP is configured, the ingress router (where a SAP is configured) does not load-balance the traffic unless shared queuing or ingress policing is configured in the SAP. This is not specific to EVPN, but is generic to the way Epipes forward traffic.

In all-active multihoming for EVPN-VPWS, there is no DF election and all PEs in the ES are active. For ESI-23, both PE-2 and PE-3 are active primary DF, but there are no DF candidates, because there is no DF election:

```
*A:PE-2# show service system bgp-evpn ethernet-segment name "ESI-23" evi 20

===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
20            2             0                   yes 11/30/2022 10:02:39
===============================================================================
===============================================================================
DF Candidates                           Time Added           Oper Pref  Do Not
                                                               Value    Preempt
-------------------------------------------------------------------------------
No entries found
===============================================================================
```

Similarly, on PE-3:

```
*A:PE-3# show service system bgp-evpn ethernet-segment name "ESI-23" evi 20

===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
20            2             0                   yes 11/30/2022 10:02:58
===============================================================================

===============================================================================
DF Candidates                           Time Added           Oper Pref  Do Not
                                                               Value    Preempt
-------------------------------------------------------------------------------
No entries found
===============================================================================
```

To confirm that all-active multihoming is working correctly, the following command shows all information related to a specific ESI; in this case, ESI-23 on PE-2:

```
*A:PE-2# show service system bgp-evpn ethernet-segment name "ESI-23" all

===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : ESI-23
Eth Seg Type            : None
Admin State             : Enabled            Oper State         : Up
ESI                     : 01:00:00:00:00:23:00:00:00:01
Oper ESI                : 01:00:00:00:00:23:00:00:00:01
Auto-ESI Type           : None
AC DF Capability        : Include
Multi-homing            : allActive          Oper Multi-homing  : allActive
ES SHG Label            : 524277
Source BMAC LSB         : None
Lag Id                  : 1
ES Activation Timer     : 3 secs
Oper Group              : (Not Specified)
Svc Carving             : auto               Oper Svc Carving   : auto
Cfg Range Type          : primary
Vprn NextHop EVI Ranges : <none>
===============================================================================

===============================================================================
EVI Information
===============================================================================
EVI                 SvcId               Actv Timer Rem      DF
-------------------------------------------------------------------------------
20                  2                   0                   yes
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
---snip---
===============================================================================
```

SRv6 tunnels in EVPN-VPWS services with single-active multihoming

Single-active multihoming allows for per-service load-balancing. Single-active multihoming is configured on PE-4 and PE-5 with ES "ESI-45". Both PEs have an SDP to MTU-6, which is associated with the ES and to the Epipe service. The configuration of the local and remote AC names and Ethernet tags is identical on PE-4 and PE-5.

On PE-4, the service configuration is as follows:

```
*A:PE-4# configure
    service
        sdp 46 mpls create
            far-end 192.0.2.6
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        system
            bgp-evpn
                ethernet-segment "ESI-45" create
                    esi 01:00:00:00:00:45:00:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 46
                    no shutdown
                exit
            exit
        exit
        epipe 2 name "Epipe-2" customer 1 create
            segment-routing-v6 1 create
                locator "loc_Epipe-2"
                    function
                        end-dx2
                    exit
                exit
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-ESI-45-MTU-6 create
                    eth-tag 456
                exit
                remote-attachment-circuit AC-ESI-23-MTU-1 create
                    eth-tag 231
                exit
                evi 20
                segment-routing-v6 bgp 1 srv6-instance 1 default-locator "loc_Epipe-2" create
                    # source-address 2001:db8::2:4    # defined for SRv6 on router level
                    ecmp 2
                    no shutdown
                exit
            exit
            spoke-sdp 46:201 create
```

```
                no shutdown
            exit
            no shutdown
        exit
    exit all
```

On PE-5, the configuration is similar, but with a different SDP:

```
*A:PE-5# configure
    service
        sdp 56 mpls create
            far-end 192.0.2.6
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        system
            bgp-evpn
                ethernet-segment "ESI-45" create
                    esi 01:00:00:00:00:45:00:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 56
                    no shutdown
                exit
            exit
        exit
        epipe 2 name "Epipe-2" customer 1 create
            segment-routing-v6 1 create
                locator "loc_Epipe-2"
                    function
                        end-dx2
                    exit
                exit
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-ESI-45-MTU-6 create
                    eth-tag 456
                exit
                remote-attachment-circuit AC-ESI-23-MTU-1 create
                    eth-tag 231
                exit
                evi 20
                segment-routing-v6 bgp 1 srv6-instance 1 default-locator "loc_Epipe-2" create
                    # source-address 2001:db8::2:5    # defined for SRv6 on router level
                    ecmp 2
                    no shutdown
                exit
            exit
            spoke-sdp 56:201 create
                no shutdown
            exit
            no shutdown
        exit
    exit all
```

The core PEs exchange three route types: AD per-EVI, AD per-ES, and ES routes.

As an example, the following is the ES route with originator PE-4 sent by RR PE-2 to PE-5. It contains a target 00:00:00:00:45:00 in the extended community that is derived from the ESI:

```
# on PE-2:
56 2022/11/30 10:04:09.636 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:5
"Peer 1: 2001:db8::2:5: UPDATE
Peer 1: 2001:db8::2:5 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 85
    Flag: 0x90 Type: 14 Len: 34 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.4
        Type: EVPN-ETH-SEG Len: 23 RD: 192.0.2.4:0 ESI: 01:00:00:00:00:45:00:00:00:01, IP-Len:
 4 Orig-IP-Addr: 192.0.2.4
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.4
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        1.1.1.1
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        df-election::DF-Type:Auto/DP:0/DF-Preference:0/AC:1
        target:00:00:00:00:45:00
"
```

The AD per-ES route has a MAX-ET tag and an ESI label in the extended community. The multihoming mode is single-active. As in the case of all-active multihoming, the ESI label is not used in Epipe services. The following BGP update with originator PE-5 is sent by RR PE-2 to its client PE-4:

```
# on PE-2:
53 2022/11/30 10:04:09.634 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:4
"Peer 1: 2001:db8::2:4: UPDATE
Peer 1: 2001:db8::2:4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 127
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.5
        Type: EVPN-AD Len: 25 RD: 192.0.2.5:20 ESI: 01:00:00:00:00:45:00:00:00:01, tag: MAX-ET
 Label: 0 (Raw Label: 0x0) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.5
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        1.1.1.1
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:20
        esi-label:3/Single-Active
    Flag: 0xc0 Type: 40 Len: 37 Prefix-SID-attr:
       SRv6 Services TLV (37 bytes):-
           Type: SRV6 L2 Service TLV (6)
                Type: 1 Len: 6
                BL:0 NL:0 FL:0 AL:0 TL:0 TO:0
"
```

The AD per-EVI route contains flags for primary and backup, which are different for routes received from PE-4 and PE-5. In this case, PE-4 is the primary in the single-active multihoming ES (P = 1):

```
# on PE-2:
67 2022/11/30 10:04:13.745 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:5
"Peer 1: 2001:db8::2:5: UPDATE
Peer 1: 2001:db8::2:5 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 127
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.4
        Type: EVPN-AD Len: 25 RD: 192.0.2.4:20 ESI: 01:00:00:00:00:45:00:00:00:01, tag: 456
 Label: 8388400 (Raw Label: 0x7fff30) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.4
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        1.1.1.1
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:20
        l2-attribute:MTU: 1514 C: 0 P: 1 B: 0
    Flag: 0xc0 Type: 40 Len: 37 Prefix-SID-attr:
       SRv6 Services TLV (37 bytes):-
           Type: SRV6 L2 Service TLV (6)
                Type: 1 Len: 6
                BL:48 NL:16 FL:20 AL:0 TL:20 TO:64
"
```

PE-5 is the backup in the single-active multihoming ES (B = 1):

```
# on PE-2:
69 2022/11/30 10:04:13.820 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:5
"Peer 1: 2001:db8::2:5: UPDATE
Peer 1: 2001:db8::2:5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 113
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.5
        Type: EVPN-AD Len: 25 RD: 192.0.2.5:20 ESI: 01:00:00:00:00:45:00:00:00:01, tag: 456
 Label: 8388432 (Raw Label: 0x7fff50) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:20
        l2-attribute:MTU: 1514 C: 0 P: 0 B: 1
    Flag: 0xc0 Type: 40 Len: 37 Prefix-SID-attr:
       SRv6 Services TLV (37 bytes):-
           Type: SRV6 L2 Service TLV (6)
           Length: 34 bytes, Reserved: 0x0
         SRv6 Service Information Sub-TLV (33 bytes)
             Type: 1 Len: 30 Rsvd1: 0x0
                Type: 1 Len: 6
                BL:48 NL:16 FL:20 AL:0 TL:20 TO:64
"
```

The BGP EVPN AD routes are shown with the following command:

```
*A:PE-2# show router bgp routes evpn auto-disc esi 01:00:00:00:00:45:00:00:00:01
```

```
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Flag  Route Dist.         ESI                           NextHop
      Tag                                               Label
-------------------------------------------------------------------------------
u*>i  192.0.2.4:20        01:00:00:00:00:45:00:00:00:01 192.0.2.4
      456                                               524275

u*>i  192.0.2.4:20        01:00:00:00:00:45:00:00:00:01 192.0.2.4
      MAX-ET                                            0

u*>i  192.0.2.5:20        01:00:00:00:00:45:00:00:00:01 192.0.2.5
      456                                               524277

u*>i  192.0.2.5:20        01:00:00:00:00:45:00:00:00:01 192.0.2.5
      MAX-ET                                            0

-------------------------------------------------------------------------------
Routes : 4
===============================================================================
```

For each PE in the single-active ES, there are two AD routes: the routes with MAX-ET are AD per-ES routes and the routes with a configured Ethernet tag are AD per-EVI routes.

The EVPN VPWS destination for Epipe 2 on PE-2 is ESI-45, as shown in the following output:

```
*A:PE-2# show service id 2 segment-routing-v6 instance 1 destinations

===============================================================================
TEP, SID
===============================================================================
Instance  TEP Address                        Segment Id
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================

===============================================================================
Segment Routing v6 Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
1         01:00:00:00:00:45:00:00:00:01   0             11/30/2022 10:04:14
-------------------------------------------------------------------------------
Number of entries: 1
-------------------------------------------------------------------------------
===============================================================================
```

The ESI is resolved to the TEP address of the primary (DF) PE-4, as follows:

```
*A:PE-2# show service id 2 segment-routing-v6 esi 01:00:00:00:00:45:00:00:00:01

===============================================================================
Segment Routing v6 Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
1         01:00:00:00:00:45:00:00:00:01   0             11/30/2022 10:04:14
-------------------------------------------------------------------------------
Number of entries: 1
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
Segment Routing v6 Dest TEP Info
===============================================================================
Instance  TEP Address                   Segment Id          Last Change
-------------------------------------------------------------------------------
1         192.0.2.4                     2001:db8:aaaa:204:* 11/30/2022 10:04:14
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

The DF election is key for the forwarding and backup functions in single-active multihoming ESs. The PE elected as DF is the primary for the ES in the Epipe and unblocks its SAP and spoke SDP for upstream and downstream traffic. The rest of the PEs in the ES bring their ES SAPs or spoke SDPs operationally down.

PE-5 is a non-DF, as follows:

```
*A:PE-5# show service system bgp-evpn ethernet-segment name "ESI-45" evi 20

===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
20            2             0                   no  11/30/2022 10:03:57
===============================================================================

===============================================================================
DF Candidates                           Time Added           Oper Pref  Do Not
                                                               Value    Preempt
-------------------------------------------------------------------------------
192.0.2.4                               11/30/2022 10:04:10  0          Disabl*
192.0.2.5                               11/30/2022 10:04:11  0          Disabl*
-------------------------------------------------------------------------------
Number of entries: 2
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

In single-active multihoming, the service SAP or spoke SDP is brought operationally down on the non-DF, as shown in the following output:

```
*A:PE-5# show service id 2 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
56:201           Spok     192.0.2.6       Up      Down      524275    524275
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
```

```
===============================================================================
```

The spoke sdp 56:201 is operationally down with a StandbyForMHProtocol flag:

```
*A:PE-5# show service id 2 sdp 56:201 detail | match Flag
Flags              : StandbyForMHProtocol
```

Two consecutive DF elections take place: the first DF election includes all PEs in the ES for that Epipe and determines which PE is the primary PE (flags P = 1, B = 0). The second DF election excludes this DF and determines which PE is the backup (P = 0, B = 1). All other PEs signal flags P = 0 and B = 0.

When the primary PE fails, AD per-ES and AD per-EVI withdrawal messages are sent to the remote PE, which updates its next hop to the backup. The backup PE takes over immediately without waiting for the ES activation timer (configured with the **es-activation-timer** command) to bring up its SAP and spoke SDP.

ES failures

When the SDP toward the primary (DF) fails, the backup PE needs to take over. An SDP failure is emulated and log 99 on PE-4 shows that SDP 46 is operationally down and PE-4 is no longer the DF:

```
155 2022/11/30 10:11:25.583 UTC MINOR: SVCMGR #2303 Base
"Status of SDP 46 changed to admin=up oper=down"

157 2022/11/30 10:11:25.584 UTC MINOR: SVCMGR #2094 Base
"Ethernet Segment:ESI-45, EVI:20, Designated Forwarding state changed to:false"
```

Remote PEs receive route withdrawal updates (unreachable NLRI) from the former DF PE-4, for example on PE-2:

```
# on PE-2:
2 2022/11/30 10:11:25.585 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:4
"Peer 1: 2001:db8::2:4: UPDATE
Peer 1: 2001:db8::2:4 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 34
    Flag: 0x90 Type: 15 Len: 30 Multiprotocol Unreachable NLRI:
        Address Family EVPN
        Type: EVPN-AD Len: 25 RD: 192.0.2.4:20 ESI: 01:00:00:00:00:45:00:00:00:01, tag: MAX-ET
 Label: 0 (Raw Label: 0x0) PathId:
"

1 2022/11/30 10:11:25.585 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:4
"Peer 1: 2001:db8::2:4: UPDATE
Peer 1: 2001:db8::2:4 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 59
    Flag: 0x90 Type: 15 Len: 55 Multiprotocol Unreachable NLRI:
        Address Family EVPN
        Type: EVPN-AD Len: 25 RD: 192.0.2.4:20 ESI: 01:00:00:00:00:45:00:00:00:01, tag: 456
 Label: 0 (Raw Label: 0x0) PathId:
        Type: EVPN-ETH-SEG Len: 23 RD: 192.0.2.4:0 ESI: 01:00:00:00:00:45:00:00:00:01, IP-Len:
 4 Orig-IP-Addr: 192.0.2.4
"
```

The backup PE-5 is promoted to primary (P = 1, B = 0) and sends BGP updates accordingly. The following AD per-EVI is received on PE-2:

```
# on PE-2:
5 2022/11/30 10:11:25.589 UTC MINOR: DEBUG #2001 Base Peer 1: 2001:db8::2:5
"Peer 1: 2001:db8::2:5: UPDATE
Peer 1: 2001:db8::2:5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 113
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.5
        Type: EVPN-AD Len: 25 RD: 192.0.2.5:20 ESI: 01:00:00:00:00:45:00:00:00:01, tag: 456
 Label: 8388432 (Raw Label: 0x7fff50) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:20
        l2-attribute:MTU: 1514 C: 0 P: 1 B: 0
    Flag: 0xc0 Type: 40 Len: 37 Prefix-SID-attr:
       SRv6 Services TLV (37 bytes):-
           Type: SRV6 L2 Service TLV (6)
           Length: 34 bytes, Reserved: 0x0
             SRv6 SID Sub-Sub-TLV
                Type: 1 Len: 6
                BL:48 NL:16 FL:20 AL:0 TL:20 TO:64
"
```

PE-5 brings up its spoke SDP without waiting for the ES activation timer and takes over immediately. It is now the only DF candidate, and therefore the DF, as follows:

```
*A:PE-5# show service system bgp-evpn ethernet-segment name "ESI-45" evi 20

===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
20            2             0                   yes 11/30/2022 10:03:57
===============================================================================

===============================================================================
DF Candidates                           Time Added           Oper Pref  Do Not
                                                               Value    Preempt
-------------------------------------------------------------------------------
192.0.2.5                               11/30/2022 10:04:11  0          Disabl*
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

BGP updates are exchanged and the remote PEs resolve the ESI to the TEP address 192.0.2.5. For example, on PE-2:

```
*A:PE-2# show service id 2 segment-routing-v6 esi 01:00:00:00:00:45:00:00:00:01

===============================================================================
Segment Routing v6 Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
```

```
-------------------------------------------------------------------------------
1         01:00:00:00:00:45:00:00:00:01   0             11/30/2022 10:11:26
-------------------------------------------------------------------------------
Number of entries: 1
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
Segment Routing v6 Dest TEP Info
===============================================================================
Instance  TEP Address                     Segment Id          Last Change
-------------------------------------------------------------------------------
1         192.0.2.5                       2001:db8:aaaa:205:* 11/30/2022 10:11:26
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

Because of the default DF election algorithm, this process is revertive; as soon as the SDP 46 is operationally up again, a new DF election is triggered with two DF candidates and PE-4 is elected as DF. A non-revertive mode is also available if preference-based DF election is configured.

Troubleshooting and debugging

The following **show** and **debug** commands can be used in EVPN-VPWS:

- **show redundancy bgp-evpn-multi-homing**
- **show router bgp routes evpn** (and filters)
- **show service segment-routing-v6 [**<*ip-address*>**]**
- **show service id** <*service-id*> **bgp-evpn**
- **show service system bgp-evpn**
- **show service system bgp-evpn ethernet-segment** (and modifiers)
- **debug router bgp update**
- **show log log-id** *99*

Most of these commands have been shown in the preceding sections; some commands are shown in this section.

Information about the configured boot timers (before DF election) and ES activation timer (after the system has been elected DF) is shown as follows:

```
*A:PE-2# show redundancy bgp-evpn-multi-homing

===============================================================================
Redundancy BGP EVPN Multi-homing Information
===============================================================================
Boot-Timer              : 10 secs
Boot-Timer Remaining    : 0 secs
ES Activation Timer     : 3 secs
===============================================================================
```

See chapter EVPN for MPLS Tunnels for a description of these timers.

The following command shows that the BGP route type 4 (ES route) messages are only imported by the PEs in the same ES; for example, on PE-3:

```
*A:PE-3# show router bgp routes evpn eth-seg
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Eth-Seg Routes
===============================================================================
Flag  Route Dist.         ESI                           NextHop
      OrigAddr
-------------------------------------------------------------------------------
u*>i  192.0.2.2:0         01:00:00:00:00:23:00:00:00:01 192.0.2.2
      192.0.2.2

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

On PE-4:

```
*A:PE-4# show router bgp routes evpn eth-seg
===============================================================================
 BGP Router ID:192.0.2.4        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Eth-Seg Routes
===============================================================================
Flag  Route Dist.         ESI                           NextHop
      OrigAddr
-------------------------------------------------------------------------------
u*>i  192.0.2.5:0         01:00:00:00:00:45:00:00:00:01 192.0.2.5
      192.0.2.5

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

The following command shows all the EVPN-SRv6 destinations toward TEP 192.0.2.4. Epipe 1 has an EVPN-SRv6 destination toward TEP 192.0.2.4 directly and Epipe 2 has an EVPN-SRv6 destination to ESI-45, which is resolved to TEP 192.0.2.4. This is shown in the following output:

```
*A:PE-2# show service segment-routing-v6 192.0.2.4

===============================================================================
SRV6 Tunnel Endpoint: 192.0.2.4
===============================================================================
Service Id               Segment Id     Type                     Srv6 Instance
-------------------------------------------------------------------------------
1                        2001:db8:aaaa* evpn                     1
```

```
-------------------------------------------------------------------------------
===============================================================================
* indicates that the corresponding row element may have been truncated.

===============================================================================
BGP EVPN SRV6 Ethernet Segment Dest
===============================================================================
Instance  Service Id     Eth Seg Id                   Segment Id
-------------------------------------------------------------------------------
1        2             01:00:00:00:45:00:00:00:01 2001:db8:aaaa:204:7fff:*
-------------------------------------------------------------------------------
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

The following command lists all configured ESs on the system:

```
*A:PE-2# show service system bgp-evpn ethernet-segment

===============================================================================
Service Ethernet Segment
===============================================================================
Name                             ESI                          Admin     Oper
-------------------------------------------------------------------------------
ESI-23                           01:00:00:00:23:00:00:00:01 Enabled   Up
-------------------------------------------------------------------------------
Entries found: 1
===============================================================================
```

In addition to the preceding commands, the following **tools dump** commands may be useful:

- **tools dump service evpn usage** - This command shows the number of EVPN-SRv6 (and EVPN-MPLS and EVPN-VXLAN) destinations in the system.
- **tools dump service system bgp-evpn ethernet-segment** <*name*> **evi** <*value*> **df** - This command computes the DF election for a specific ESI and EVI. For all-active multihoming, there is no DF election and all PEs forward traffic. For single-active multihoming, one PE is active for a service while another PE is a backup. This command shows the DF (primary), even if it is not the local PE.

The usage of EVPN resources is shown as follows:

```
*A:PE-2# tools dump service evpn usage

vxlan-srv6-evpn-mpls usage statistics at 11/30/2022 10:08:31:

MPLS-TEP                                         :            0
VXLAN-TEP                                        :            0
SRV6-TEP                                         :            2
Total-TEP                                        :      2/ 16383

Mpls Dests (TEP, Egress Label + ES + ES-BMAC)    :            0
Mpls Etree Leaf Dests                            :            0
Vxlan Dests (TEP, Egress VNI + ES)               :            0
Srv6 Dests (TEP, SID  + ES)                      :            2
Total-Dest                                       :      2/196607

Sdp Bind +  Evpn Dests                           :      2/245759
ES L2/L3 PBR                                     :      0/ 32767
Evpn Etree Remote BUM Leaf Labels                :            0
```

On PE-2, there is one SRv6 TEP (192.0.2.4 in Epipe 1 and in Epipe 2) and there are two SRv6 destinations: 192.0.2.4 and ESI 01:00:00:00:00:45:00:00:00:01. PE-5 is not an SRv6 TEP for PE-2 because it is not a primary and, therefore, is not forwarding any traffic.

In all-active multihoming, the DF election is not applicable:

```
*A:PE-2# tools dump service system bgp-evpn ethernet-segment "ESI-23" evi 20 df

[11/30/2022 10:08:31] All Active VPWS or IP-ALIASING - DF N/A
```

In single-active multihoming, the following command shows which PE is the DF:

```
*A:PE-5# tools dump service system bgp-evpn ethernet-segment "ESI-45" evi 20 df

[11/30/2022 10:08:36] Computed DF: 192.0.2.4 (Remote) (Boot Timer Expired: Yes)
[11/30/2022 10:08:36] Computed Backup: 192.0.2.5 (This Node)
```

The command is launched on PE-5, which is a backup. The computed DF is PE-4 and the boot timer has expired, meaning there is no DF re-election pending.

Conclusion

EVPN-VPWS is a simplified point-to-point version of RFC 7432. EVPN provides a unified control plane mechanism that simplifies the network deployment and operation. Single-active and all-active multihoming can be used in Epipes; EVPN-VPWS is a differentiator of EVPN compared to traditional TLDP or BGP Epipe redundancy mechanisms.

EVPN-IFF BGP Attribute Propagation Between Families

This chapter provides information about EVPN-IFF BGP attribute propagation between families .

Topics in this chapter include:

Applicability

The information and configuration in this chapter are based on SR OS Release 22.7.R1. EVPN Interface-ful (EVPN-IFF) BGP attribute propagation between BGP families based on uniform propagation is supported in SR OS Release 21.2.R1 and later.

For more information on routed VPLS in EVPN, see chapters EVPN for VXLAN Tunnels (Layer 3) and EVPN for MPLS Tunnels in Routed VPLS .

Overview

SR OS allows multiple BGP owners in the same VPRN service to receive or advertise IP prefixes contained in the VPRN route table. A VPRN route table can simultaneously install and process IPv4 or IPv6 prefixes for the following owners:

- EVPN Interface-ful (EVPN-IFF)
- EVPN Interface-less (EVPN-IFL)
- VPN-IP (also referred to as IP-VPN routes)
- IP (also referred to as BGP PE-CE routes)

EVPN-IFF routes are EVPN IP-prefix routes, otherwise known as route type 5 (RT-5) routes, that are imported and exported based on the configuration of the R-VPLS services attached to the VPRN. To enable the EVPN-IFF model, the command **config>service>vpls>bgp-evpn>ip-route-advertisement** needs to be configured. By default, BGP attributes are re-originated when a prefix is propagated to and from an EVPN-IFF route. However, BGP attributes can be used to influence routing (for example, local preference, Autonomous System (AS) path, communities, and so on), and therefore, SR OS supports EVPN-IFF BGP attribute propagation to other BGP families (uniform propagation), as described in *draft-ietf-bess-evpn-ipvpn-interworking*.

The following CLI command is used to enable EVPN-IFF BGP attribute propagation and EVPN-IFF best path selection:

```
*A:PE-4>config>service>system>bgp-evpn# ip-prefix-routes ?
  - ip-prefix-routes
```

```
[no] iff-attribute-* - Enable attribute uniform propagation
[no] iff-bgp-path-s* - Enable bgp path selection
```

The **iff-bgp-path-selection** command cannot be enabled when **iff-attribute-uniform-propagation** is disabled.

When **iff-attribute-uniform-propagation** is enabled on a node:

- the following BGP path attributes are propagated:
  - AS path
  - domain path (D-PATH), supported in SR OS Release 21.10.R1 and later
  - IBGP-only attributes, when advertising to an IBGP neighbor: local preference, originator ID, cluster ID
  - Multiple Exit Discriminator (MED)
  - communities, large communities, extended communities
- the following BGP path attributes are not propagated across families:
  - any type 0x06 extended communities supported by RT-5 routes:
    - MAC mobility extended community
    - EVPN router MAC extended community
  - BGP encapsulation extended community
  - Route Target extended community
  - BGP tunnel encapsulation attribute
  - BGP prefix-SID attribute used in RT-5 routes and VPN-IP routes for Segment Routing over IPv6 dataplane (SRv6) services
- IBGP-only attributes are only propagated to IBGP neighbors; EBGP-only attributes only to EBGP neighbors
- routes received with well-known communities, such as no-advertise or no-export(-subconfed), are sent or not sent depending on the community values
- BGP path attributes are propagated even when doing route leaking between routing instances

If multiple EVPN-IFF routes for the same prefix are received for the same VPRN, they are by default ordered and selected based on the lowest R-VPLS Ifindex, Route Distinguisher (RD), and Ethernet tag.

When **iff-bgp-path-selection** is enabled, EVPN-IFF routes with the same or different RD are selected based on regular BGP path selection rules in the following order:

1. valid route wins over invalid route (invalid routes are looped routes or routes where the originator ID matches the receiving router)
2. lowest origin validation state (origin validation state: valid is preferred to origin validation state: not found; origin validation state: not found is preferred to origin validation state: invalid) – applicable to IPv4, IPv6, or BGP Labeled Unicast (BGP-LU) routes
3. lowest Routing Table Manager (RTM) preference
4. highest local preference
5. shortest D-PATH

6. lowest Accumulated Interior Gateway Protocol (AIGP) metric (AIGP is not supported for EVPN-IFL, EVPN-IFF, or IP-VPN routes)
7. shortest AS path
8. lowest origin (origin: IGP is preferred to origin: EGP; origin: EGP is preferred to origin: incomplete)
9. lowest MED (routes without MED are considered as zero or infinity based on the configuration of the **always-compare-med** command)
10. lowest owner type (owner type: BGP-label is preferred to owner type: BGP; owner type: BGP is preferred to owner type: BGP-VPN) with BGP-VPN referring to VPN-IP and EVPN-IFL
11. EBGP wins over IBGP
12. lowest route-table or tunnel-table cost to the next-hop

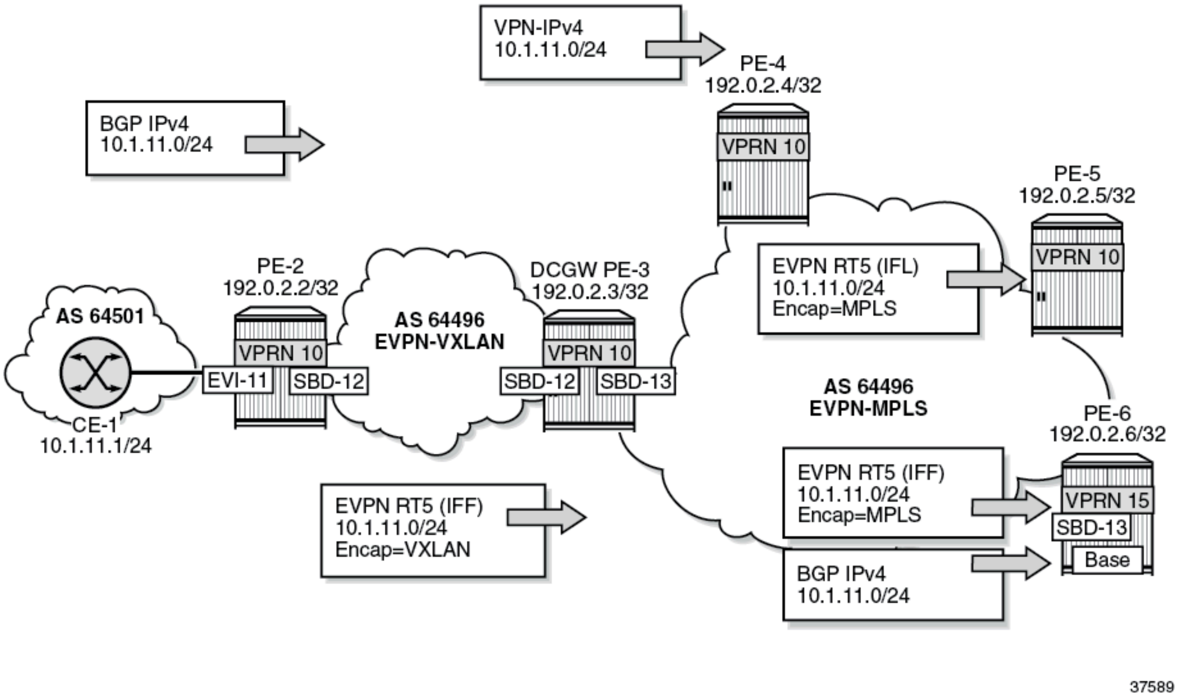
    Note: The **ignore-nh-metric** command is not supported for EVPN-IFF.

13. lowest next-hop type – a next-hop resolved to a tunnel-table entry is considered as a lower type than a next-hop resolved to a route-table entry
14. lowest router ID – applicable to IBGP peers
15. shortest cluster list length – applicable to IBGP peers
16. lowest IP address – IP address refers to the peer that advertised the route
17. EVPN-IFL wins over IPVPN
18. next-hop check (IPv4 next-hop wins over IPv6, then lowest next-hop wins) - The next-hop check is a tiebreaker if BGP receives the same prefix for VPN-IPv6 and EVPN-IFL. An IPv6 prefix received as VPN-IPv6 has an IPv6 next-hop whereas the same IPv6 prefix received as EVPN-IFL can have an IPv4 next-hop.
19. lowest RD for route-table selection
20. lowest path ID (add-path)

Configuration

Figure 134: Example topology shows the example topology with PE-3 as Data Center Gateway (DCGW) between an EVPN-VXLAN network and an EVPN-MPLS network. Routed VPLS is configured on PE-2, PE-3, and PE-6. Supplementary broadcast domain "SBD-12" is configured in the EVPN-VXLAN network between PE-2 and PE-3; "SBD-13" in the EVPN-MPLS network between PE-3 and PE-6. On PE-2, Ethernet VPN instance "EVI-11" is configured toward CE-1.

Figure 134: Example topology



CE-1 advertises prefix 10.1.11.0/24 to BGP neighbor 10.0.0.2 in VPRN 10 on PE-2. PE-2 sends an EVPN-IFF route to DCGW PE-3. PE-3 forwards the prefix 10.1.11.0/24 as VPN-IPv4 route to PE-4, as EVPN-IFL route to PE-5, as EVPN-IFF route to PE-6, and as IPv4 route to PE-6.

The initial configuration includes the following:

- Cards, MDAs, ports
- Router interfaces on all PEs
- IS-IS on the router interfaces
- LDP on the router interfaces on PE-3, PE-4, PE-5, and PE-6

On the PEs, BGP is configured for the EVPN address family. Between PE-3 and PE-4, both the VPN-IPv4 and the EVPN address family are configured. The configuration on PE-3 is as follows:

```
# on PE-3:
configure
    router Base
        autonomous-system 64496
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            rapid-update evpn
            group "internal1"
                family evpn
                peer-as 64496
                neighbor 192.0.2.2
                exit
            exit
            group "internal"
                peer-as 64496
                neighbor 192.0.2.4
                    family vpn-ipv4 evpn
                exit
                neighbor 192.0.2.5
                    family evpn
                exit
                neighbor 192.0.2.6
                    family evpn
                exit
            exit
        exit
```

On CE-1, BGP is configured in VPRN 11 for the IPv4 address family. The export policy adds communities "1:1" and "2:2" and sets the MED to a value of 81.

```
# on CE-1:
configure
    router Base
        policy-options
            begin
            community "1:1_2:2"
                members "1:1" "2:2"
            exit
            policy-statement "export-vnf-to-all"
                entry 10
                    from
                        protocol direct direct-interface
                    exit
                    action accept
                        community add "1:1_2:2"
                        bgp-med set 81
                    exit
                exit
            exit
            commit
        exit
    exit
    service
        vprn 11 name "VPRN 11" customer 1 create
            autonomous-system 64501
            interface "int-CE-1-PE-2" create
                address 10.0.0.1/24
                sap 1/1/2:11 create
                exit
            exit
            interface "test" create
                address 10.1.11.1/24
                sap 1/1/2:12 create
                exit
            exit
            bgp
                export "export-vnf-to-all"
                split-horizon
                group "CE-1-PE-2"
                    type external
                    peer-as 64496
                    neighbor 10.0.0.2
                    exit
                exit
            exit
            no shutdown
```

On PE-2, VPRN 10 has R-VPLS interface "int-EVI-11" toward CE-1 and R-VPLS interface "int-SBD-12" toward PE-3. BGP is configured toward neighbor 10.0.0.1 on CE-1 and the import policy sets the local preference (LP) to 200, as follows:

```
# on PE-2:
configure
    router Base
        policy-options
            begin
            policy-statement "local-preference-200"
                entry 10
                    action accept
                        local-preference 200
                    exit
                exit
            exit
            commit
        exit
    exit
    service
        vprn 10 name "VPRN 10" customer 1 create
            autonomous-system 64496
            interface "int-SBD-12" create
                vpls "SBD-12"
                    evpn-tunnel
                exit
            exit
            interface "int-EVI-11" create
                address 10.0.0.2/24
                vrrp 1 owner passive
                    backup 10.0.0.2
                exit
                vpls "EVI-11"
                exit
            exit
            bgp
                import "local-preference-200"
                local-as 64496
                split-horizon
                group "PE-2-CE-1"
                    type external
                    peer-as 64501
                    neighbor 10.0.0.1
                    exit
                exit
            exit
            no shutdown
        exit
        vpls 11 name "EVI-11" customer 1 create
            allow-ip-int-bind
            exit
            stp
                shutdown
            exit
            sap 1/1/1:11 create
                no shutdown
            exit
            no shutdown
        exit
        vpls 12 name "SBD-12" customer 1 create
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 12 create
            exit
            bgp-evpn
                no mac-advertisement
                ip-route-advertisement
                evi 12
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

On PE-3, VPRN 10 is configured with:

- three interfaces:
  - R-VPLS interface "int-SBD-12" toward PE-2
  - R-VPLS interface "int-SBD-13" toward PE-6
  - interface "int-VPRN10-PE-3-to-PE-6" to the base router of PE-6.
- BGP-IPVPN for the exchange of VPN-IPv4 routes with PE-4
- BGP-EVPN to propagate EVPN-IFL routes to PE-5 and EVPN-IFF routes to PE-6
- BGP to propagate BGP IPv4 routes to the base router on PE-6. The export policy is only required in the BGP configuration.

```
# on PE-3:
configure
    router Base
        policy-options
            begin
            prefix-list "10.1.0.0"
                prefix 10.1.0.0/16 longer
            exit
            policy-statement "export-bgp"
                entry 10
                    from
                        prefix-list "10.1.0.0"
                    exit
                    action accept
                    exit
                exit
            exit
            commit
        exit
    exit
    service
        vprn 10 name "VPRN 10" customer 1 create
            autonomous-system 64496
            interface "int-SBD-12" create
                vpls "SBD-12"
                    evpn-tunnel
                exit
            exit
            interface "int-SBD-13" create
                vpls "SBD-13"
                    evpn-tunnel
                exit
            exit
            interface "int-VPRN10-PE-3-to-PE-6" create
```

```
                address 10.15.16.3/24
                sap 1/1/3:13 create
                exit
            exit
            bgp-ipvpn
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    route-distinguisher 192.0.2.3:10
                    vrf-target target:64496:10
                    no shutdown
                exit
            exit
            bgp-evpn
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    route-distinguisher 192.0.2.3:10
                    vrf-target target:64496:10
                    no shutdown
                exit
            exit
            bgp
                export "export-bgp"
                rapid-withdrawal
                group "base router - PE-6"
                    family ipv4
                    neighbor 10.15.16.6
                        type internal
                        peer-as 64496
                    exit
                exit
            exit
            no shutdown
        exit
        vpls 12 name "SBD-12" customer 1 create
            description "EVPN-VXLAN VPLS for EVPN tunnel to PE-2"
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 12 create
            exit
            bgp-evpn
                no mac-advertisement
                ip-route-advertisement
                evi 12
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            no shutdown
        exit
        vpls 13 name "SBD-13" customer 1 create
            description "EVPN-MPLS VPLS for EVPN tunnel to PE-6"
            allow-ip-int-bind
            exit
            bgp
            exit
            bgp-evpn
                no mac-advertisement
                ip-route-advertisement
                evi 13
                mpls bgp 1
```

```
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

On PE-4, VPRN 10 is configured with BGP-IPVPN, as follows. BGP between PE-3 and PE-4 is configured for the VPN-IPv4 address family.

```
# on PE-4:
configure
    service
        vprn 10 name "VPRN 10" customer 1 create
            bgp-ipvpn
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    route-distinguisher 192.0.2.4:10
                    vrf-target target:64496:10
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

On PE-5, VPRN 10 is configured with BGP-EVPN, as follows:

```
# on PE-5:
configure
    service
        vprn 10 name "VPRN 10" customer 1 create
            bgp-evpn
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    route-distinguisher 192.0.2.5:10
                    vrf-target target:64496:10
                    no shutdown
                exit
            exit
            bgp
                no shutdown
            exit
            no shutdown
        exit
```

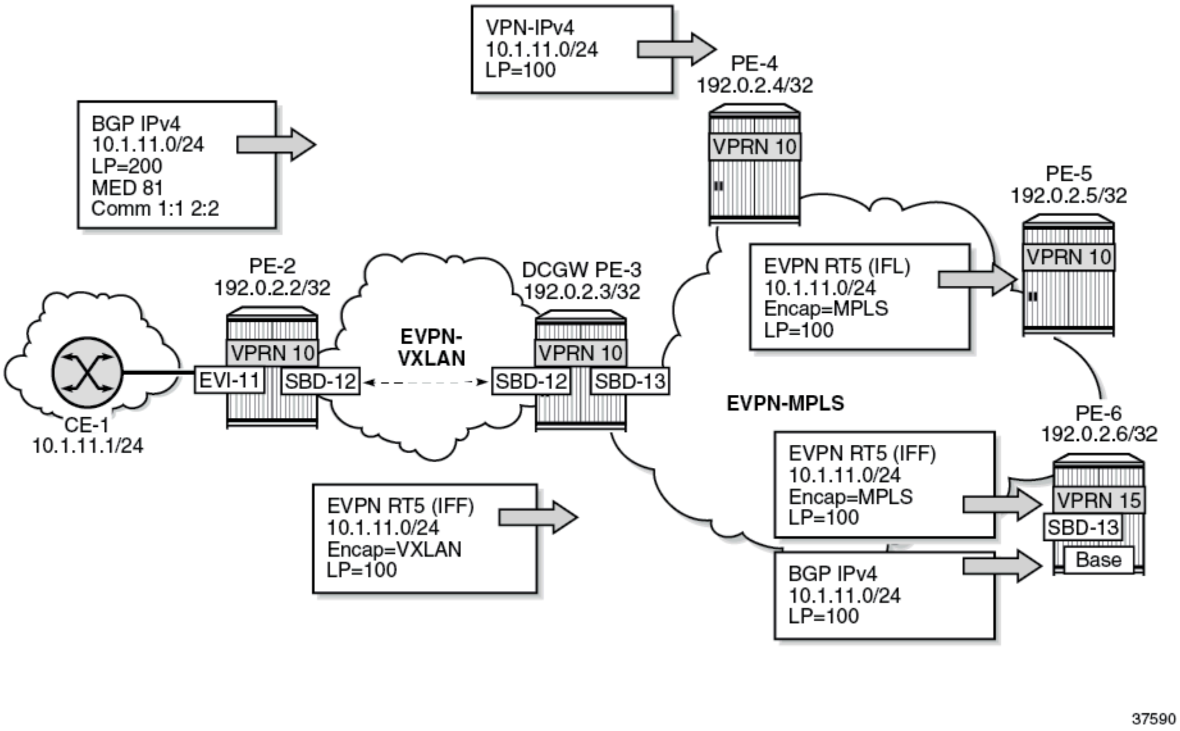
In the base router of PE-6, BGP is configured to neighbor 10.15.16.3 on PE-3. VPRN 15 is configured with R-VPLS interface "int-SBD-13" toward PE-3. The configuration is as follows:

```
# on PE-6:
configure
    router Base
        interface "int-PE-6-to-VPRN10-PE-3"
            address 10.15.16.6/24
            port 1/1/1:13
        exit
        bgp
            group "PE-6-CE"
                family ipv4
                neighbor 10.15.16.3
                    type internal
                    local-as 64496
                    peer-as 64496
                exit
            exit
        exit
    exit
    service
        vprn 15 name "VPRN 15" customer 1 create
            autonomous-system 64502
            interface "int-SBD-13" create
                vpls "SBD-13"
                    evpn-tunnel
                exit
            exit
            no shutdown
        exit
        vpls 13 name "SBD-13" customer 1 create
            allow-ip-int-bind
            exit
            bgp
            exit
            bgp-evpn
                no mac-advertisement
                ip-route-advertisement
                evi 13
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

Default behavior

By default, BGP path attributes are re-originated when a prefix is propagated to and from an EVPN-IFF route. Figure 135: EVPN-IFF BGP path attributes are re-originated by PE-2 and PE-3 shows that PE-2 receives an IPv4 route for prefix 10.1.11.0/24 with non-default BGP path attributes, whereas PE-2 propagates the prefix as an EVPN-IFF route with default path attributes.

Figure 135: EVPN-IFF BGP path attributes are re-originated by PE-2 and PE-3



VPRN 10 on PE-2 received a BGP IPv4 route for prefix 10.1.11.0/24 with LP 200, MED 81, and communities "1:1" and "2:2":

```
*A:PE-2# show router 10 bgp routes 10.1.11.0/24 hunt
===============================================================================
 BGP Router ID:192.0.2.2        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 10.1.11.0/24
Nexthop        : 10.0.0.1
Path Id        : None
From           : 10.0.0.1
Res. Protocol  : LOCAL                  Res. Metric    : 0
Res. Nexthop   : 10.0.0.1
Local Pref.    : 200                    Interface Name : int-EVI-11
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 81
AIGP Metric    : None                   IGP Cost       : 0
Connector      : None
Community      : 1:1 2:2
```

```
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 255.0.0.0
Fwd Class      : None                   Priority       : None
Flags          : Used Valid Best IGP In-RTM
Route Source   : External
AS-Path        : 64501
Route Tag      : 0
Neighbor-AS    : 64501
Orig Validation: NotFound
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
RIB Priority   : Normal
Last Modified  : 00h03m24s

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

PE-2 propagates prefix 10.1.11.0/24 as an EVPN-IFF route to PE-3 with default BGP attributes: LP 100, no MED, and without the communities "1:1" and "2:2":

```
*A:PE-2# show router bgp routes evpn ip-prefix prefix 10.1.11.0/24 hunt
===============================================================================
 BGP Router ID:192.0.2.2        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
Network        : n/a
Nexthop        : 192.0.2.2
Path Id        : None
To             : 192.0.2.3
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : n/a
Connector      : None
Community      : target:64496:12 mac-nh:02:13:ff:ff:ff:49
                 bgp-tunnel-encap:VXLAN
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.3
Origin         : IGP
AS-Path        : No As-Path
EVPN type      : IP-PREFIX
ESI            : ESI-0
Tag            : 0
Gateway Address: 02:13:ff:ff:ff:49
```

```
Prefix         : 10.1.11.0/24
Route Dist.    : 192.0.2.2:12
MPLS Label     : VNI 12
Route Tag      : 0
Neighbor-AS    : n/a
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

Uniform propagation for EVPN-IFF BGP path attributes to different BGP families

Enabling **iff-attribute-uniform-propagation** is not allowed when there are services enabled with **bgp-evpn ip-route-advertisement**:

```
*A:PE-2>config>service>system>bgp-evpn>ip-prefix-routes# iff-attribute-uniform-propagation
MINOR: SVCMGR #1003 Inconsistent value - iff-attribute-uniform-propagation cannot be enabled/
disabled when there are "bgp-evpn ip-route-advertisement" enabled services
```

To enable **iff-attribute-uniform-propagation** and **iff-best-path-selection** on PE-2, **ip-route-advertisement** must be temporarily disabled in VPLS "SBD-12", as follows:

```
# on PE-2
configure
    service
        vpls "SBD-12"
            bgp-evpn
                no ip-route-advertisement
            exit
        exit
        system
            bgp-evpn
                ip-prefix-routes
                    iff-attribute-uniform-propagation
                    iff-bgp-path-selection
                exit
            exit
        exit
        vpls "SBD-12"
            bgp-evpn
                ip-route-advertisement
            exit
        exit
```

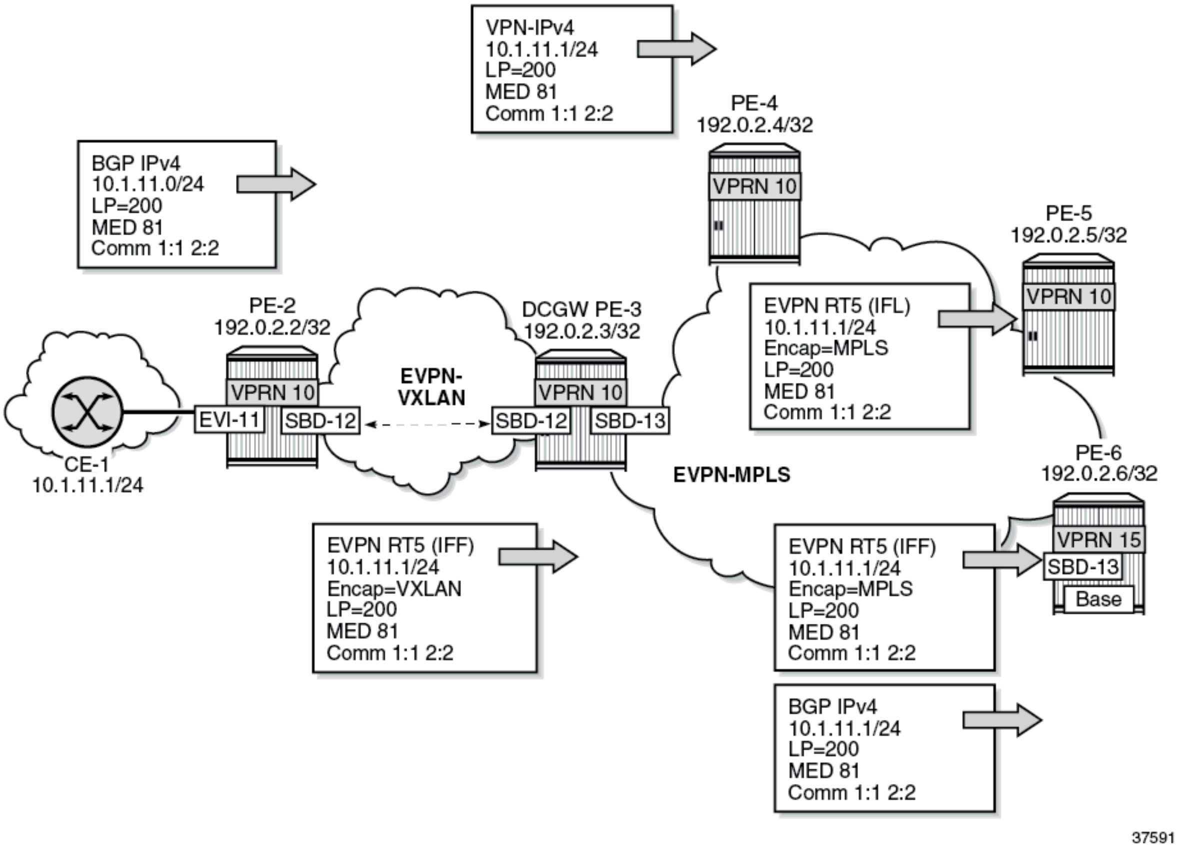
In a similar configuration, **iff-attribute-uniform-propagation** and **iff-bgp-path-selection** are enabled on the other PEs.

The following command shows that uniform propagation for EVPN-IFF BGP path attributes and BGP path selection are enabled:

```
*A:PE-2# show service system bgp-evpn

===============================================================================
System BGP EVPN Information
===============================================================================
Eth Seg Route Dist.               : <none>
Eth Seg Oper Route Dist.          : <none>
Eth Seg Oper Route Dist Type      : none
Ad Per ES Route Target            : evi-rt
Etree
  Leaf                            : Disabled
Mcast Leave Sync Prop             : 5
Attribute Uniform Prop            : Enabled
BGP Path Selection                : Enabled
D-Path Length Ignore              : Disabled
===============================================================================
```

Figure 136: Uniform propagation for EVPN-IFF BGP path attributes between families shows the uniform propagation for EVPN-IFF BGP path attributes between families in the same Virtual Routing and Forwarding (VRF).

Figure 136: Uniform propagation for EVPN-IFF BGP path attributes between families



With the uniform propagation for EVPN-IFF BGP path attributes enabled, PE-2 propagates EVPN-IFF route 10.1.11.0/24 to PE-3 with LP 200, MED 81, and communities "1:1" and "2:2". The following EVPN-IFF route is received at PE-3:

```
*A:PE-3# show router bgp routes evpn ip-prefix prefix 10.1.11.0/24 hunt
===============================================================================
 BGP Router ID:192.0.2.3        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : n/a
Nexthop        : 192.0.2.2
Path Id        : None
From           : 192.0.2.2
Res. Nexthop   : 192.168.23.1
Local Pref.    : 200                    Interface Name : int-PE-3-PE-2
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 81
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : 1:1 2:2 target:64496:12 mac-nh:02:13:ff:ff:ff:49
                 bgp-tunnel-encap:VXLAN
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.2
Flags          : Used Valid Best IGP
Route Source   : Internal
AS-Path        : 64501
EVPN type      : IP-PREFIX
ESI            : ESI-0
Tag            : 0
Gateway Address: 02:13:ff:ff:ff:49
Prefix         : 10.1.11.0/24
Route Dist.    : 192.0.2.2:12
MPLS Label     : VNI 12
Route Tag      : 0
Neighbor-AS    : 64501
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h01m30s

-------------------------------------------------------------------------------
---snip---
```

With the uniform propagation for EVPN-IFF BGP path attributes enabled, PE-3 propagates VPN-IPv4 route 10.1.11.0/24 to PE-4 with LP 200, MED 81, and communities "1:1" and "2:2". The following VPN-IPv4 route is received at PE-4:

```
*A:PE-4# show router bgp routes 10.1.11.0/24 vpn-ipv4 hunt
===============================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP VPN-IPv4 Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 10.1.11.0/24
Nexthop        : 192.0.2.3
```

```
Route Dist.    : 192.0.2.3:10           VPN Label      : 524283
Path Id        : None
From           : 192.0.2.3
Res. Nexthop   : n/a
Local Pref.    : 200                    Interface Name : int-PE-4-PE-3
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 81
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : 1:1 2:2 target:64496:10
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.3
Fwd Class      : None                   Priority       : None
Flags          : Used Valid Best IGP
Route Source   : Internal
AS-Path        : 64501
Route Tag      : 0
Neighbor-AS    : 64501
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h01m44s
VPRN Imported  :  10

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

PE-3 propagates EVPN-IFL route 10.1.11.0/24 to PE-5 with LP 200, MED 81, and communities "1:1" and "2:2". The following EVPN-IFL route is received at PE-5:

```
*A:PE-5# show router bgp routes evpn ip-prefix prefix 10.1.11.0/24 hunt
===============================================================================
 BGP Router ID:192.0.2.5        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : n/a
Nexthop        : 192.0.2.3
Path Id        : None
From           : 192.0.2.3
Res. Nexthop   : 192.168.35.1
Local Pref.    : 200                    Interface Name : int-PE-5-PE-3
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 81
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : 1:1 2:2 target:64496:10 bgp-tunnel-encap:MPLS
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.3
Flags          : Used Valid Best IGP
```

```
Route Source   : Internal
AS-Path        : 64501
EVPN type      : IP-PREFIX
ESI            : ESI-0
Tag            : 0
Gateway Address: 00:00:00:00:00:00
Prefix         : 10.1.11.0/24
Route Dist.    : 192.0.2.3:10
MPLS Label     : LABEL 524282
Route Tag      : 0
Neighbor-AS    : 64501
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h02m36s

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

PE-3 propagates EVPN-IFF route 10.1.11.0/24 to PE-6 with LP 200, MED 81, and communities "1:1" and "2:2". The following EVPN-IFF route is received at PE-6:

```
*A:PE-6# show router bgp routes evpn ip-prefix prefix 10.1.11.0/24 hunt
===============================================================================
 BGP Router ID:192.0.2.6        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : n/a
Nexthop        : 192.0.2.3
Path Id        : None
From           : 192.0.2.3
Res. Nexthop   : 192.168.36.1
Local Pref.    : 200                    Interface Name : int-PE-6-PE-3
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 81
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : 1:1 2:2 target:64496:13 mac-nh:02:17:ff:ff:ff:4a
                 bgp-tunnel-encap:MPLS
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.3
Flags          : Used Valid Best IGP
Route Source   : Internal
AS-Path        : 64501
EVPN type      : IP-PREFIX
ESI            : ESI-0
Tag            : 0
Gateway Address: 02:17:ff:ff:ff:4a
Prefix         : 10.1.11.0/24
```

```
Route Dist.    : 192.0.2.3:13
MPLS Label     : LABEL 524281
Route Tag      : 0
Neighbor-AS    : 64501
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h03m01s

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

PE-3 propagates BGP IPv4 route 10.1.11.0/24 to PE-6 with LP 200, MED 81, and communities "1:1" and "2:2". The following IPv4 route is received at PE-6:

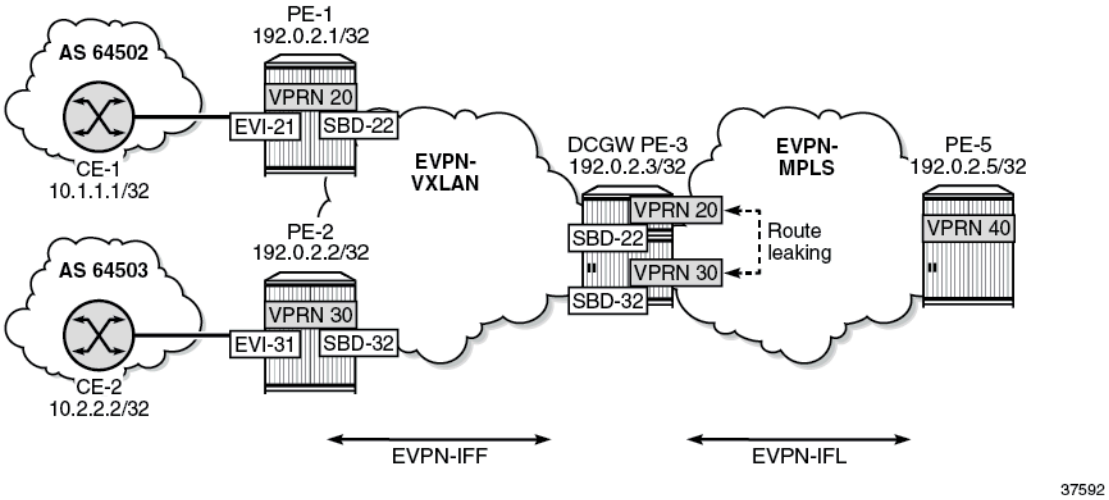
```
*A:PE-6# show router bgp routes 10.1.11.0/24 hunt
===============================================================================
 BGP Router ID:192.0.2.6        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 10.1.11.0/24
Nexthop        : 10.15.16.3
Path Id        : None
From           : 10.15.16.3
Res. Protocol  : LOCAL                  Res. Metric    : 0
Res. Nexthop   : 10.15.16.3
Local Pref.    : 200                    Interface Name : int-PE-6-to-VPRN10-PE*
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 81
AIGP Metric    : None                   IGP Cost       : 0
Connector      : None
Community      : 1:1 2:2
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.3
Fwd Class      : None                   Priority       : None
Flags          : Used Valid Best IGP In-RTM
Route Source   : Internal
AS-Path        : 64501
Route Tag      : 0
Neighbor-AS    : 64501
Orig Validation: NotFound
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
RIB Priority   : Normal
Last Modified  : 00h03m17s

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
```

```
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

EVPN-IFF BGP path attributes exported to leaked EVPN routes

Figure 137: Example topology shows the example topology with two VPRNs on DCGW PE-3 where routes are leaked.

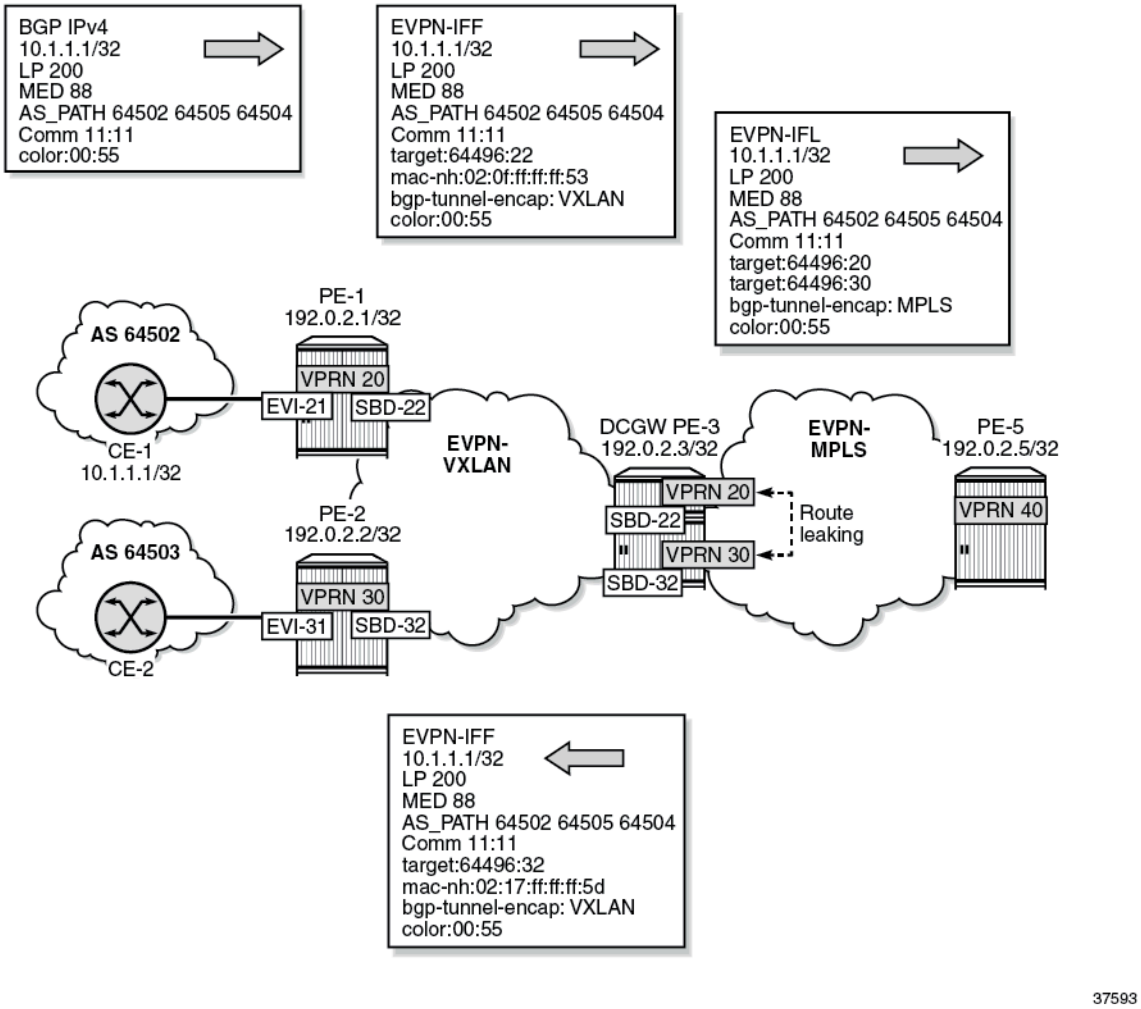
Figure 137: Example topology



The uniform propagation for EVPN-IFF BGP path attributes is enabled on all PEs.

Figure 138: BGP path attributes are propagated in leaked EVPN routes shows that CE-1 exports an IPv4 route for prefix 10.1.1.1/32 to PE-1. This route has non-default BGP attributes; for example, MED 88, AS path 64502 64505 64504, and community "11:11" "color:00:55". PE-1 exports this route as an EVPN-IFF route to PE-3. PE-3 forwards this route as EVPN-IFL route to PE-5. On PE-3, the route is leaked from VPRN 20 to VPRN 30. The BGP path attributes are propagated to the leaked EVPN routes, except those attributes that are not expected to be propagated, such as the router's MAC extended community. PE-3 advertises an EVPN-IFF route for prefix 10.1.1.1/32 to PE-2.

Figure 138: BGP path attributes are propagated in leaked EVPN routes



In a similar way, CE-2 exports IPv4 prefix 10.2.2.2/32 to PE-2 with non-default BGP path attributes. PE-2 advertises this prefix as an EVPN-IFF route with the same BGP path attributes. PE-3 leaks the route from VPRN 30 to VPRN 20 while preserving the BGP path attributes. PE-3 advertises an EVPN-IFF route for prefix 10.2.2.2/32 to PE-1 with the same BGP path attributes. PE-3 also advertises the prefix as EVPN-IFL route to PE-5 with the same BGP path attributes. For brevity, the routes for prefix 10.2.2.2/32 are not shown here.

In this example, VPRN "CE-1" is configured as follows. The export policy sets the MED, prepends some AS numbers to the AS path, and adds the communities "11:11" and "color:00:55".

```
# CE-1:
configure
    router Base
        policy-options
            begin
            community "11:11"
                members "11:11"
            exit
            community "color:00:55"
                members "color:00:55"
            exit
            policy-statement "export-vnf-to-all-2"
                entry 10
                    from
                        protocol direct direct-interface
                    exit
                    action next-entry
                        community add "11:11" "color:00:55"
                        as-path-prepend 64504
                        bgp-med set 88
                    exit
                exit
                entry 20
                    from
                        protocol direct direct-interface
                    exit
                    action accept
                        as-path-prepend 64505
                    exit
                exit
            exit
            commit
        exit
    exit
    service
        vprn 23 name "CE-1" customer 1 create
            autonomous-system 64502
            interface "int-CE-1-PE-1" create
                address 10.2.0.1/24
                sap 1/2/2:21 create
                exit
            exit
            interface "loopback" create
                address 10.1.1.1/32
                loopback
            exit
            bgp
                export "export-vnf-to-all-2"
                local-as 64502
                group "PE-1-CE-1"
                    neighbor 10.2.0.254
                        type external
                        peer-as 64496
                    exit
                exit
            exit
            no shutdown
        exit
```

On PE-1, an import policy sets the LP to a value of 200. VPRN 20 has R-VPLS interface "int-EVI-21" toward CE-1 and R-VPLS interface "int-SBD-22" toward PE-2.

```
# on PE-1:
configure
    router Base
        policy-options
            begin
            policy-statement "local-preference-200"
                entry 10
                    action accept
                        local-preference 200
                    exit
                exit
            exit
            commit
        exit
    exit
    service
        vprn 20 name "VPRN 20" customer 1 create
            autonomous-system 64496
            interface "int-SBD-22" create
                vpls "SBD-22"
                    evpn-tunnel
                exit
            exit
            interface "int-EVI-21" create
                address 10.2.0.254/24
                vrrp 1 owner passive
                    backup 10.2.0.254
                exit
                vpls "EVI-21"
                exit
            exit
            bgp
                import "local-preference-200"
                local-as 64496
                group "PE-1-CE"
                    type external
                    peer-as 64502
                    neighbor 10.2.0.1
                    exit
                exit
            exit
            no shutdown
        exit
        vpls 21 name "EVI-21" customer 1 create
            allow-ip-int-bind
            exit
            stp
                shutdown
            exit
            sap 1/2/1:21 create
            exit
            no shutdown
        exit
        vpls 22 name "SBD-22" customer 1 create
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 22 create
            exit
            bgp
            exit
            bgp-evpn
                no mac-advertisement
                ip-route-advertisement
                evi 22
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
```

The configuration on PE-2 is similar with VPRN 30, R-VPLS "EVI-31", and R-VPLS "SBD-32".

PE-3 has two VPRNs: "VPRN 20" and "VPRN 30". Export policy "leak-color-55-into-30" is used to leak routes with color community "color:00:55" from VPRN 20 to VPRN 30. The configuration is as follows:

```
# on PE-3:
configure
    router Base
        policy-options
            begin
            community "color:00:55"
                members "color:00:55"
            exit
            community "RT64496:20"
                members "target:64496:20"
            exit
            community "RT64496:30"
                members "target:64496:30"
            exit
            policy-statement "leak-color-55-into-20"
                entry 10
                    from
                        community "color:00:55"
                    exit
                    action accept
                        community add "RT64496:20" "RT64496:30"
                    exit
                exit
            exit
            policy-statement "leak-color-55-into-30"
                entry 10
                    from
                        community "color:00:55"
                    exit
                    action accept
                        community add "RT64496:20" "RT64496:30"
                    exit
                exit
            exit
            commit
        exit
    exit
    service
        vpls 22 name "SBD-22" customer 1 create
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 22 create
            exit
            bgp-evpn
                no mac-advertisement
                ip-route-advertisement
                evi 22
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
        vprn 20 name "VPRN 20" customer 1 create
            autonomous-system 64496
            interface "int-SBD-22" create
```

```
                vpls "SBD-22"
                    evpn-tunnel
                exit
            exit
            bgp-evpn
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    route-distinguisher 192.0.2.3:20
                    vrf-export "leak-color-55-into-30"
                    vrf-target import target:64496:20
                    no shutdown
                exit
            exit
            no shutdown
        exit
        vpls 32 name "SBD-32" customer 1 create
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 32 create
            exit
            bgp-evpn
                no mac-advertisement
                ip-route-advertisement
                evi 32
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
        vprn 30 name "VPRN 30" customer 1 create
            autonomous-system 64496
            interface "int-SBD-32" create
                vpls "SBD-32"
                    evpn-tunnel
                exit
            exit
            bgp-evpn
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    route-distinguisher 192.0.2.3:30
                    vrf-export "leak-color-55-into-20"
                    vrf-target import target:64496:30
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

PE-3 exports the prefix route as EVPN-IFL to PE-5. On PE-5, VPRN 40 is configured as follows:

```
# on PE-5:
configure
    router Base
        policy-options
            begin
            community "RT64496:20"
                members "target:64496:20"
            exit
            community "RT64496:30"
                members "target:64496:30"
            exit
            policy-statement "vrf-40-import"
                entry 10
                    from
                        community "RT64496:20"
                    exit
                    action accept
                    exit
                exit
                entry 20
                    from
                        community "RT64496:30"
                    exit
                    action accept
                    exit
                exit
            exit
            policy-statement "vrf-40-export"
                entry 10
                    from
                        protocol direct direct-interface
                    exit
                    action accept
                        community add "RT64496:20" "RT64496:30"
                    exit
                exit
            exit
            commit
        exit
    exit
    service
        vprn 40 name "VPRN 40" customer 1 create
            autonomous-system 64496
            interface "loopback" create
                address 10.5.5.5/32
                loopback
            exit
            bgp-evpn
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    route-distinguisher 192.0.2.5:40
                    vrf-export "vrf-40-export"
                    vrf-import "vrf-40-import"
                    no shutdown
                exit
            exit
            no shutdown
```

CE-1 exports an IPv4 route for prefix 10.1.1.1/32 to PE-1 with community "color:00:55" and other non-default BGP path attributes. The route table for VPRN 20 on PE-1 includes an BGP IPv4 route for prefix 10.1.1.1/32:

```
*A:PE-1# show router 20 route-table 10.1.1.1/32

===============================================================================
Route Table (Service: 20)
```

```
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.1.1.1/32                                   Remote  BGP       00h01m57s  170
       10.2.0.1                                                     0
-------------------------------------------------------------------------------
No. of Routes: 1
```

PE-1 propagates prefix 10.1.1.1/32 in an EVPN-IFF route. On PE-3, the route table includes an EVPN-IFF route for prefix 10.1.1.1/32:

```
*A:PE-3# show router 20 route-table 10.1.1.1/32

===============================================================================
Route Table (Service: 20)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.1.1.1/32                                   Remote  EVPN-IFF  00h01m58s  169
       int-SBD-22 (ET-02:0f:ff:ff:ff:53)                            0
-------------------------------------------------------------------------------
No. of Routes: 1
```

PE-3 forwards prefix 10.1.1.1/32 as an EVPN-IFL to PE-5. On PE-5, the route table includes an EVPN-IFL route for prefix 10.1.1.1/32:

```
*A:PE-5# show router 40 route-table

===============================================================================
Route Table (Service: 40)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.1.1.1/32                                   Remote  EVPN-IFL  00h02m07s  170
       192.0.2.3 (tunneled)                                         10
10.2.2.2/32                                   Remote  EVPN-IFL  00h02m39s  170
       192.0.2.3 (tunneled)                                         10
10.5.5.5/32                                   Local   Local     00h03m24s  0
       loopback                                                     0
-------------------------------------------------------------------------------
No. of Routes: 3
```

In a similar way, PE-5 received an EVPN-IFL route for prefix 10.2.2.2/32. Prefix 10.5.5.5/32 is local to VPRN 40 on PE-5 and is advertised to PE-3 as EVPN-IFL route.

On PE-3, routes with community "color:00:55" are leaked between VPRN 20 and VPRN 30. PE-1 and PE-3 have forwarded the route with the original BGP path attributes, so this community is preserved and the route for prefix 10.1.1.1/32 is leaked to VPRN 30, as shown in the following route table. The next hop is R-VPLS "SBD-22" in local VPRN 20.

```
*A:PE-3# show router 30 route-table

===============================================================================
Route Table (Service: 30)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
```

```
10.1.1.1/32                                   Remote  EVPN-IFL  00h02m19s  169
       Local VRF [20:int-SBD-22]                                    0
10.2.2.2/32                                   Remote  EVPN-IFF  00h02m52s  169
       int-SBD-32 (ET-02:13:ff:ff:ff:5d)                            0
10.3.0.0/24                                   Remote  EVPN-IFF  00h03m42s  169
       int-SBD-32 (ET-02:13:ff:ff:ff:5d)                            0
10.5.5.5/32                                   Remote  EVPN-IFL  00h03m36s  170
       192.0.2.5 (tunneled)                                         10
-------------------------------------------------------------------------------
No. of Routes: 4
```

PE-3 propagates prefix 10.1.1.1/32 as an EVPN-IFF route to PE-2, so the route table for VPRN 30 on PE-2 includes an entry for prefix 10.1.1.1/32 with next hop "SBD-32" toward VPRN 30 on PE-3:

```
*A:PE-2# show router 30 route-table

===============================================================================
Route Table (Service: 30)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.1.1.1/32                                   Remote  EVPN-IFF  00h02m30s  169
       int-SBD-32 (ET-02:17:ff:ff:ff:5d)                            0
10.2.2.2/32                                   Remote  BGP       00h03m02s  170
       10.3.0.1                                                     0
10.3.0.0/24                                   Local   Local     00h04m00s  0
       int-EVI-31                                                   0
10.5.5.5/32                                   Remote  EVPN-IFF  00h03m47s  169
       int-SBD-32 (ET-02:17:ff:ff:ff:5d)                            0
-------------------------------------------------------------------------------
No. of Routes: 4
```

The following show commands illustrate that the BGP path attributes are propagated. VPRN 20 on PE-1 receives an IPv4 route for prefix 10.1.1.1/32 from CE-1 with LP 200, MED 88, AS path 64502 64505 64504, and communities "1:1" "color:00:55", as follows:

```
*A:PE-1# show router 20 bgp routes 10.1.1.1/32 hunt
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : 10.1.1.1/32
Nexthop        : 10.2.0.1
Path Id        : None
From           : 10.2.0.1
Res. Protocol  : LOCAL                  Res. Metric    : 0
Res. Nexthop   : 10.2.0.1
Local Pref.    : 200                    Interface Name : int-EVI-21
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 88
AIGP Metric    : None                   IGP Cost       : 0
```

```
Connector      : None
Community      : 11:11 color:00:55
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.1
Fwd Class      : None                   Priority       : None
Flags          : Used Valid Best IGP In-RTM
Route Source   : External
AS-Path        : 64502 64505 64504
Route Tag      : 0
Neighbor-AS    : 64502
Orig Validation: NotFound
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
RIB Priority   : Normal
Last Modified  : 00h02m35s

-------------------------------------------------------------------------------
---snip---
```

PE-1 forwards an EVPN-IFF route to PE-3 for prefix 10.1.1.1/32 with the original BGP path attributes, as follows:

```
*A:PE-1# show router bgp routes 10.1.1.1/32 evpn ip-prefix hunt
===============================================================================
 BGP Router ID:192.0.2.1        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
---snip---
-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
---snip---

Network        : n/a
Nexthop        : 192.0.2.1
Path Id        : None
To             : 192.0.2.3
Res. Nexthop   : n/a
Local Pref.    : 200                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 88
AIGP Metric    : None                   IGP Cost       : n/a
Connector      : None
Community      : 11:11 target:64496:22 mac-nh:02:0f:ff:ff:ff:53
                 bgp-tunnel-encap:VXLAN color:00:55
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.3
Origin         : IGP
AS-Path        : 64502 64505 64504
EVPN type      : IP-PREFIX
ESI            : ESI-0
Tag            : 0
Gateway Address: 02:0f:ff:ff:ff:53
Prefix         : 10.1.1.1/32
Route Dist.    : 192.0.2.1:22
MPLS Label     : VNI 22
```

```
Route Tag      : 0
Neighbor-AS    : 64502
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
---snip---
```

PE-3 forwards an EVPN-IFL route for prefix 10.1.1.1/32 to PE-5, so PE-5 receives the following route with the original BGP path attributes:

```
*A:PE-5# show router bgp routes evpn ip-prefix prefix 10.1.1.1/32 hunt
===============================================================================
 BGP Router ID:192.0.2.5        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : n/a
Nexthop        : 192.0.2.3
Path Id        : None
From           : 192.0.2.3
Res. Nexthop   : 192.168.35.1
Local Pref.    : 200                    Interface Name : int-PE-5-PE-3
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 88
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : 11:11 target:64496:20 target:64496:30
                 bgp-tunnel-encap:MPLS color:00:55
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.3
Flags          : Used Valid Best IGP
Route Source   : Internal
AS-Path        : 64502 64505 64504
EVPN type      : IP-PREFIX
ESI            : ESI-0
Tag            : 0
Gateway Address: 00:00:00:00:00:00
Prefix         : 10.1.1.1/32
Route Dist.    : 192.0.2.3:20
MPLS Label     : LABEL 524280
Route Tag      : 0
Neighbor-AS    : 64502
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h03m09s

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

On PE-3, the route for prefix 10.1.1.1/32 is leaked from VPRN 20 to VPRN 30. Prefix 10.1.1.1/32 is then advertised to PE-2 in the new context but preserves the BGP path attributes, so PE-2 receives the following route:

```
*A:PE-2# show router bgp routes evpn ip-prefix prefix 10.1.1.1/32 hunt
===============================================================================
 BGP Router ID:192.0.2.2        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
---snip---
Network        : n/a
Nexthop        : 192.0.2.3
Path Id        : None
From           : 192.0.2.3
Res. Nexthop   : 192.168.23.2
Local Pref.    : 200                    Interface Name : int-PE-2-PE-3
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 88
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : 11:11 target:64496:32 mac-nh:02:17:ff:ff:ff:5d
                 bgp-tunnel-encap:VXLAN color:00:55
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.3
Flags          : Used Valid Best IGP
Route Source   : Internal
AS-Path        : 64502 64505 64504
EVPN type      : IP-PREFIX
ESI            : ESI-0
Tag            : 0
Gateway Address: 02:17:ff:ff:ff:5d
Prefix         : 10.1.1.1/32
Route Dist.    : 192.0.2.3:32
MPLS Label     : VNI 32
Route Tag      : 0
Neighbor-AS    : 64502
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h02m50s
---snip---
```

Conclusion

SR OS nodes can be configured to propagate EVPN-IFF BGP path attributes between families to influence the path selection, as per *draft-ietf-bess-evpn-ipvpn-interworking*.

EVPN-MPLS E-Tree

This chapter provides information about EVPN-MPLS E-Tree.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS Release 15.0.R6, but the CLI in the current edition is based on SR OS Release 23.7.R1. VPLS E-Tree without EVPN is supported in SR OS Release 12.0.R4, and later. EVPN-MPLS E-Tree is supported in SR OS Release 15.0.R1, and later.

Overview

Ethernet Tree (E-Tree) is a rooted multipoint Ethernet service defined by the Metro Ethernet Forum (MEF). E-Tree can be implemented based on the following:

- RFC 7796, *Ethernet-Tree Support in Virtual Private LAN Services* (VPLS E-Tree without EVPN)
- RFC 8317, *E-Tree Support in EVPN and PBB-EVPN* (EVPN-MPLS E-Tree)

VPLS E-Tree without EVPN

The E-Tree implementation is based on RFC 7796 and is supported for unicast and broadcast, unknown unicast, and multicast (BUM) traffic. Interfaces can be defined as root attachment circuit (AC) or leaf AC, or both, as described in Table 7: Interfaces in E-Tree . A VPLS E-Tree can have multiple root ACs. Access and network interfaces are both supported on SAPs and SDP bindings.

Table 7: Interfaces in E-Tree

Interface	Tag
Access interface (user-to-network interface - UNI)	Root tag
	Leaf tag
Network interface (network-to-network interface - NNI)	Root-leaf tag

On the ingress access interfaces, all frames are tagged and forwarded. On the network interfaces, no traffic is dropped based on the root or leaf tag. On the egress access interfaces, all traffic toward a root AC is forwarded, whereas traffic toward a leaf AC is only forwarded when it originates from a root AC, as summarized in Table 8: E-Tree Forwarding on Access Interfaces . Traffic from leaf AC to leaf AC is blocked.

Table 8: E-Tree Forwarding on Access Interfaces

	To root AC	To leaf AC
From root AC	Allowed	Allowed
From leaf AC	Allowed	Not allowed

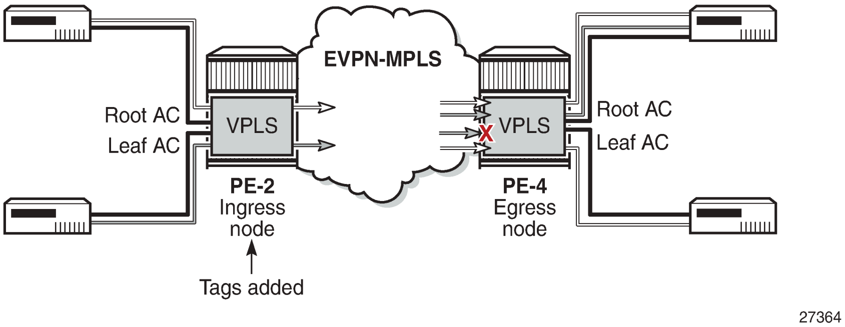
Within an E-Tree, the split horizon group capability is inherent for leaf SAPs and leaf SDP bindings and extends to all the remote nodes that are part of the same VPLS E-Tree service.

Ingress Tagging and Egress Filtering

Figure 139: Frame Forwarding in a VPLS E-Tree without EVPN shows how frames are forwarded in an E-Tree. The ingress node PE-2 knows whether the frame comes from a leaf AC or a root AC and adds a tag indicating "from root" or "from leaf". Specific VLAN IDs are used to indicate "from root" or "from leaf". The egress node PE-4 forwards the frame based on the "from root" or "from leaf" tag, as follows:

- A frame with the "from root" tag can be forwarded to any AC, leaf or root.
- A frame with the "from leaf" tag can only be forwarded to a root AC, not to a leaf AC.

Figure 139: Frame Forwarding in a VPLS E-Tree without EVPN



SAPs and SDP bindings are considered as root AC automatically (in the following example, SAP 1/2/c1/1:4 is a root AC); leaf ACs get the keyword **leaf-ac**, and NNI SAPs and SDP bindings get the keyword **root-leaf-tag**. The root tag equals the service delimiting VLAN ID (VID) in the SAP and the leaf tag can only be configured with a different value.

```
On PE-2:
configure
    service
        vpls 4 name "VPLS 4" customer 1 etree create
            sap 1/2/c1/1:4 create
            exit
            sap 1/2/c3/1:4 leaf-ac create
            exit
            sap 1/2/c5/1:4 root-leaf-tag leaf-tag 44 create
            exit
            spoke-sdp 24:4 vc-type vlan root-leaf-tag create
            exit
            spoke-sdp 210:4 leaf-ac create
            exit
            no shutdown
        exit
    exit
exit
```

VLAN ranges are not allowed in a VPLS E-Tree, as shown for the following connection profile VLAN, which is configured on PE-2:

```
On PE-2:
configure
    connection-profile-vlan 10 create
        vlan-range 10 to 19
        vlan-range 110
    exit
exit
```

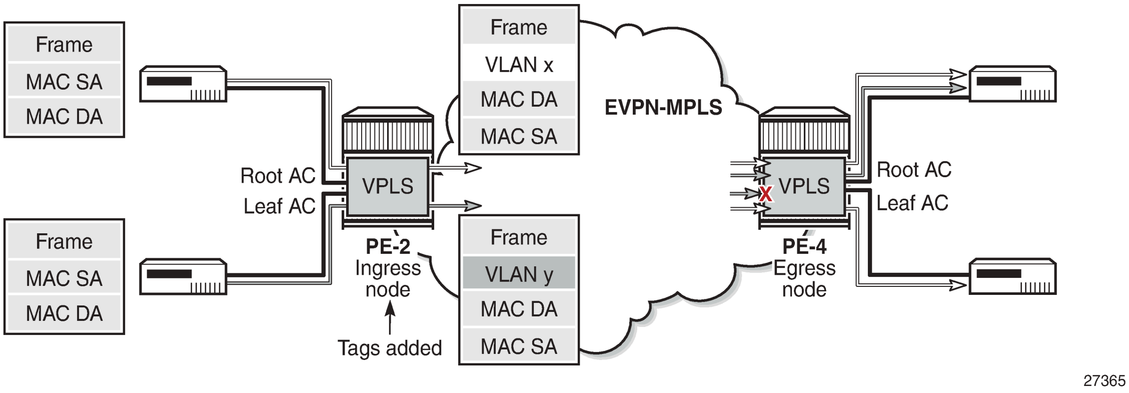
The following error is raised when attempting to configure a SAP with VLAN range cp-10:

```
configure service vpls 4 sap 1/2/c3/1:cp-10 create
MINOR: SVCMGR #8303 vlan-range not allowed - etree configured

configure service vpls 4 sap 1/2/c3/1:cp-10.* leaf-ac create
MINOR: CLI SAP-id has an invalid port number or encapsulation value.
```

All incoming frames on a SAP or SDP binding in a VPLS have their dot1q/qinq encapsulation removed by the local PE. In a VPLS E-Tree, the local PE then adds a VLAN tag with a dedicated VID indicating whether the frame originates from a root AC or a leaf AC.

- For dot1q/qinq-based L2 services, a VLAN tag with VID x is added for root and VID y for leaf. Frames with VID x are forwarded to any type of AC, while frames with VID y are only forwarded to root ACs at the remote node, as shown in Figure 140: VLAN Tags Added by Ingress Node and Filtered by Egress Node in VPLS E-Tree.
- For pseudowire-based L2 services, a VLAN tag with VID 1 is hard-coded for frames received on a root AC and a VLAN tag with VID 2 for frames received on a leaf AC.

Figure 140: VLAN Tags Added by Ingress Node and Filtered by Egress Node in VPLS E-Tree



EVPN-MPLS E-Tree

Operators migrate their regular VPLS services to EVPN services because of the advantages offered by EVPN, such as all-active multi-homing, scalability, and easy provisioning. EVPN-MPLS E-Trees block leaf-to-leaf traffic, while allowing all traffic from and to root ACs. The following is a configuration example of an EVPN-MPLS E-Tree. The **evpn-etree-leaf-label** command is only relevant for EVPN E-Tree services and allocates an E-Tree leaf label on the system, which is used for egress filtering of BUM traffic.

```
configure
    service
        system
            bgp-evpn
                evpn-etree-leaf-label
            exit
        exit
        vpls 1 name "VPLS 1" customer 1 etree create
            bgp
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    ingress-replication-bum-label
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            sap 1/2/c1/1:1 create
            exit
            sap 1/2/c3/1:1 leaf-ac create
            exit
            spoke-sdp 210:1 leaf-ac create
            exit
            no shutdown
        exit
    exit
exit
```

SAPs or SDP bindings are by default root AC objects. MAC addresses learned on root AC objects are advertised as usual, while MAC addresses learned on a SAP or SDP binding configured as leaf AC are advertised with an BGP EVPN E-Tree extended community with leaf indication bit L=1.

BGP EVPN VXLAN is not supported in E-Tree services; only EVPN-MPLS E-Tree is supported. The following error is raised when attempting to configure VXLAN in an E-Tree enabled service:

```
configure
    service
        vpls 3 name "VPLS 3" customer 1 etree create
            vxlan vni 3 create
MINOR: SVCMGR #7890 Cannot configure vxlan - not supported on etree enabled services
```

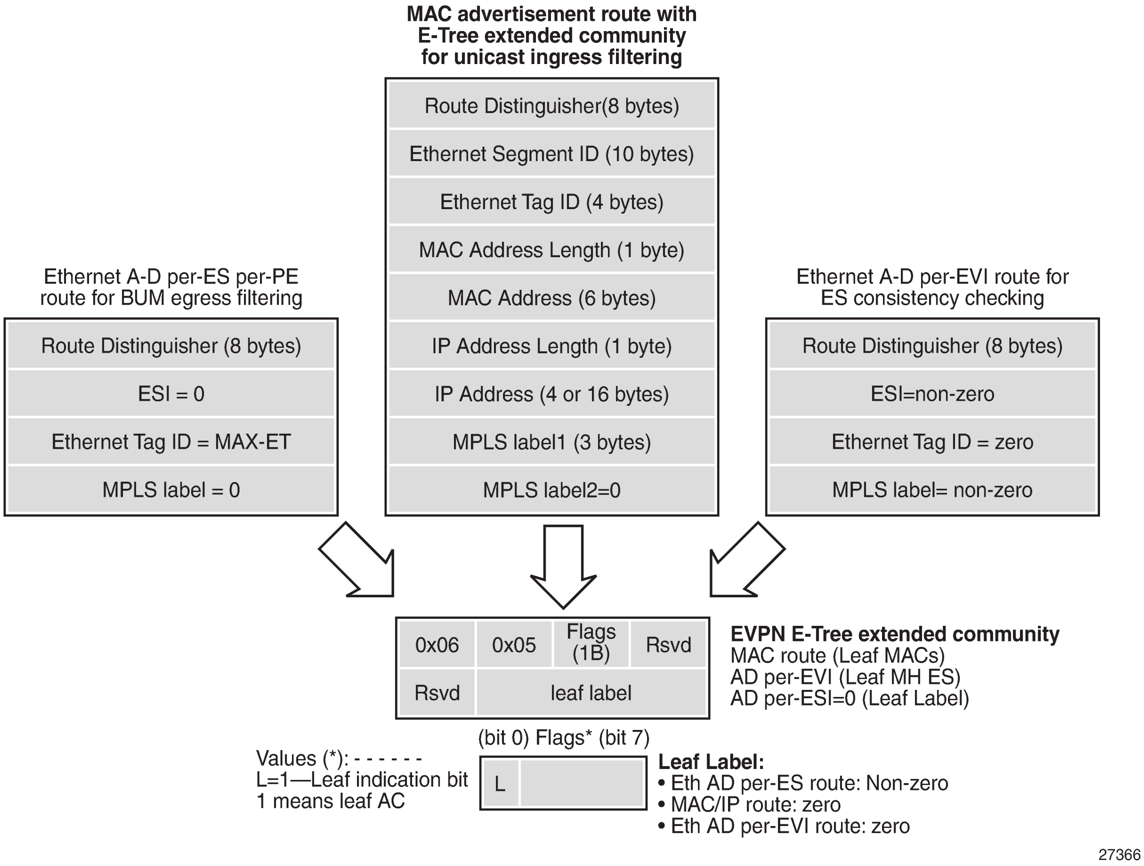
In an EVPN-MPLS E-Tree, it is not required and not even possible to configure the **root-leaf-tag** option on interfaces. The following error is raised when attempting to configure a spoke SDP or SAP with **root-leaf-tag** option:

```
configure
    service
        vpls 1
            spoke-sdp 24:1 vc-type vlan root-leaf-tag create
MINOR: SVCMGR #7883 evpn configured in service

configure
    service
        vpls 1
            sap 1/2/c3/1:200 create root-leaf-tag leaf-tag 22
MINOR: SVCMGR #7883 evpn configured in service
```

BGP EVPN Control Plane for EVPN E-Tree

No leaf tag needs to be added to frames forwarded to EVPN destinations. Instead, the BGP EVPN control plane for EVPN E-Tree advertises a leaf indication bit and a leaf label in the E-tree extended community, as shown in Figure 141: BGP EVPN Control Plane for EVPN E-Tree.

Figure 141: BGP EVPN Control Plane for EVPN E-Tree


MAC advertisement route with
E-Tree extended community
for unicast ingress filtering
Route Distinguisher(8 bytes)
Ethernet Segment ID (10 bytes)
Ethernet Tag ID (4 bytes)
MAC Address Length (1 byte)
MAC Address (6 bytes)
IP Address Length (1 byte)
IP Address (4 or 16 bytes)
MPLS label1 (3 bytes)
MPLS label2=0
Ethernet A-D per-ES per-PE
route for BUM egress filtering
Route Distinguisher (8 bytes)
ESI = 0
Ethernet Tag ID = MAX-ET
MPLS label = 0
Ethernet A-D per-EVI route for
ES consistency checking
Route Distinguisher (8 bytes)
ESI=non-zero
Ethernet Tag ID = zero
MPLS label= non-zero
0x06
0x05
Flags (1B)
Rsvd
Rsvd
leaf label
EVPN E-Tree extended community
MAC route (Leaf MACs)
AD per-EVI (Leaf MH ES)
AD per-ESI=0 (Leaf Label)
(bit 0) Flags* (bit 7)
Values (*): - - - - - -
L=1—Leaf indication bit
1 means leaf AC
L
Leaf Label:
• Eth AD per-ES route: Non-zero
• MAC/IP route: zero
• Eth AD per-EVI route: zero
27366

The BGP EVPN control plane is extended with the EVPN E-Tree extended community, as per RFC 8317. The low-order bit of the flags field contains the L-bit (L=1 indicates a leaf AC). The leaf label contains a 20-bit MPLS label that is non-zero for Ethernet Auto Discovery (AD) per Ethernet Segment (per-ES) routes (tag MAX-ET), but it equals zero for MAC/IP routes and Ethernet AD per EVPN Instance (per-EVI) routes (tag 0). The following BGP EVPN AD per-ES route contains an EVPN E-Tree extended community with L=0 and leaf label 524282, and is used for egress BUM filtering. RFC 8317 states that the leaf indication bit L must be ignored on reception and should be zero on transmission.

```
On PE-2:
9 2023/07/26 21:52:45.409 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 81
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-AD Len: 25 RD: 192.0.2.2:1 ESI: ESI-0, tag: MAX-ET Label: 0 (Raw Label: 0x0)
 PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64496:1
        etree::L:0/Leaf-Label:524282
        bgp-tunnel-encap:MPLS
"
```

The following BGP EVPN MAC route contains an EVPN E-Tree extended community with L=1 and leaf label 0, and is used for known unicast ingress filtering:

```
On PE-2:
3 2023/07/26 21:51:52.235 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 89
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-MAC Len: 33 RD: 192.0.2.2:1 ESI: ESI-0, tag: 0, mac len: 48 mac:
 ca:fe:09:29:29:29, IP len: 0, IP: NULL, label1: 8388496 (Raw Label: 0x7fff90)
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64496:1
        etree::L:1/Leaf-Label:0
        bgp-tunnel-encap:MPLS
"
```

The following BGP EVPN AD per-EVI route contains an EVPN E-Tree extended community with L=1 and leaf label 0, and is used for ES consistency checking:

```
On PE-4:
80 2023/07/26 22:33:30.588 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 81
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.5
        Type: EVPN-AD Len: 25 RD: 192.0.2.5:2 ESI: 01:00:00:00:00:45:01:00:00:01, tag: 0 Label:
 8388464 (Raw Label: 0x7fff70) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64496:2
        etree::L:1/Leaf-Label:0
        bgp-tunnel-encap:MPLS
"
```

When PE-2 receives a BGP EVPN MAC route with an E-Tree extended community with leaf indication bit L=1, the PE imports the route and installs the MAC address in the forwarding database (FDB) with an EVPN leaf (Lf) flag, as follows:

```
*A:PE-2# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
```

```
-------------------------------------------------------------------------------
1          ca:fe:01:01:01:01 sdp:210:1              L/30     07/26/23 21:55:11
1          ca:fe:06:46:46:46 mpls-1:                Evpn     07/26/23 21:52:46
                             192.0.2.4:524281
           ldp:65538
1          ca:fe:07:47:47:47 mpls-1:                Evpn,Lf  07/26/23 21:52:46
                             192.0.2.4:524281
           ldp:65538
1          ca:fe:08:28:28:28 sap:1/2/c1/1:1         L/0      07/26/23 21:51:58
1          ca:fe:09:29:29:29 sap:1/2/c3/1:1         LT/0     07/26/23 21:51:52
-------------------------------------------------------------------------------
No. of MAC Entries: 5
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

If receiving the same MAC route as root from PE-1 and as leaf from PE-2, the MAC route from PE-1 is selected: root MAC routes have higher priority than leaf MAC routes. Root static MAC routes take precedence over leaf static MAC routes.

EVPN MAC routes with a higher sequence number have a higher priority than root or leaf MAC routes. MAC mobility procedures take precedence to first identify the location of the MAC before associating that MAC with a root or a leaf site. The EVPN MAC route selection criteria in tie-break order are as follows:

1. Conditional static MACs (local protected MACs)
2. Auto-learned protected MACs (locally learned MACs on SAPs or mesh/spoke SDPs because of the configuration of auto-learn-mac-protect)
3. EVPN ES PBR MACs
4. EVPN static MACs (remote protected MACs)
5. Data plane learned MACs (regular MAC learning on SAPs/SDP-bindings)
6. EVPN MACs with a higher sequence number
7. EVPN E-Tree root MACs
8. Lowest IP (next-hop IP of the EVPN NLRI)
9. Lowest Ethernet tag (Ethernet tag is zero for MPLS and non-zero for VXLAN)
10. Lowest RD

Ingress Leaf Filtering for Unicast Traffic

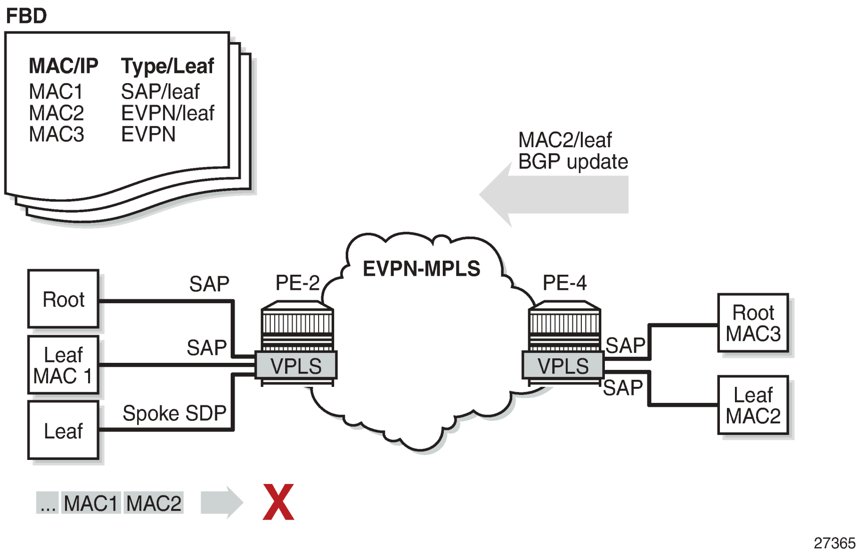
EVPN-MPLS E-Tree is the only E-Tree technology able to do unicast ingress filtering, as opposed to the usual unicast egress filtering that, for example, VPLS does. Remote MAC addresses are learned in the control plane, so EVPN can optimize the forwarding by filtering known unicast traffic at the ingress:

- Unicast frames entering a root AC at the ingress PE are not filtered. The MAC destination address (DA) is looked up in the FDB and the frames are forwarded. The MAC source address (SA) is learned and advertised in BGP EVPN without the E-Tree extended community.
- Unicast frames entering a leaf AC at the ingress PE are filtered. The MAC DA is looked up in the FDB. When the MAC DA is learned from an EVPN leaf (or a leaf AC), the frame is dropped. When the MAC DA is learned from an EVPN root (or root AC), the frame is forwarded. The MAC SA is learned and advertised in BGP EVPN with leaf indication bit L=1.

Figure 142: Ingress Leaf Filtering for Known Unicast Traffic shows that PE-4 advertises MAC2 with leaf indication bit L=1. When a frame is sent with MAC SA MAC1 on a leaf AC of PE-2, PE-2 does a MAC

lookup in the FDB to find out that the DA MAC2 is learned from an EVPN leaf. Therefore, PE-2 does not forward the frame to PE-4, but drops it at the ingress.

Figure 142: Ingress Leaf Filtering for Known Unicast Traffic



The ingress filtering blocks E-Tree leaf-to-leaf traffic and requires the implementation of an extra leaf EVPN-MPLS destination per remote PE containing leaf ACs per E-Tree service. Therefore, a dedicated EVPN-MPLS binding is created per leaf unicast traffic in the service. This additional internal EVPN-MPLS destination is created per remote PE that contains a leaf and that advertises at least one leaf MAC. The MPLS E-Tree leaf destination is created when a MAC route with L=1 is received. Any EVPN E-Tree service could potentially use one additional EVPN-MPLS destination for leaf unicast traffic per remote PE. This additional EVPN-MPLS leaf destination in the E-Tree is only unicast and not part of the flooding list. The EVPN-MPLS leaf destination consumes EVPN resources, as can be verified as follows:

```
*A:PE-2# tools dump service evpn usage | match "Mpls Etree"
Mpls Etree Leaf Dests                            :            1
```

All MAC addresses received with L=1 point to this EVPN-MPLS E-Tree leaf destination, whereas root MAC addresses point to the root destination.

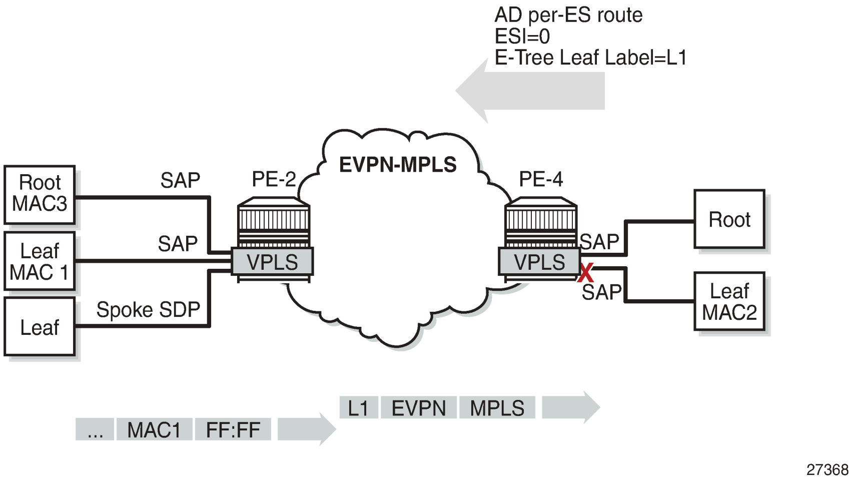
Egress Leaf Filtering for BUM Traffic

Figure 143: Egress Leaf Filtering for BUM Traffic shows that leaf-to-leaf BUM traffic is filtered at the egress, based on the EVPN leaf label advertised in the E-Tree extended community of the zero ESI AD per-ES route (tag=MAX-ET).

- BUM frames that enter a root AC at the ingress PE are not filtered; the BUM frames follow regular EVPN data plane procedures.
- BUM frames that enter a leaf AC at the ingress PE are marked as leaf and forwarded or replicated to the egress IOM. At the egress IOM, the frame is flooded in the default multicast list, subject to the following:

- Leaf entries are skipped when BUM traffic is forwarded, so no BUM traffic is forwarded to local leaf ACs.
- BUM traffic to remote BGP EVPN PEs is encapsulated with the EVPN label stack.
    - If the remote PE has advertised an AD per-ES route with E-Tree leaf label L1, this leaf label L1 is added at the bottom of the stack. At the egress PE, when the leaf label L1 matches the leaf label of the PE, the BUM traffic is only forwarded to the root ACs, not to the leaf ACs.
    - If the egress PE does not have any E-Tree enabled service, it has not advertised any AD per-ES route with E-Tree leaf label. The local PE forwards the BUM traffic with BGP EVPN encapsulation, but without an additional label. Even when the egress PE does not have E-Tree enabled, it can still work with the VPLS E-Tree service available in the ingress PE. No traffic is dropped at the egress PE where no E-Tree is configured.

Figure 143: Egress Leaf Filtering for BUM Traffic



The following command is used to monitor the ESI label entries consumed by the EVPN E-Tree application:

```
*A:PE-2# tools dump service evpn usage | match "BUM"
Evpn Etree Remote BUM Leaf Labels                 :             1
```

Configuration

The initial configuration on the nodes includes the following:

- Cards, MDAs, ports
- Router interfaces
- IS-IS (alternatively, OSPF can be used)
- LDP between the PEs
- BGP for the EVPN address family (between the PEs)

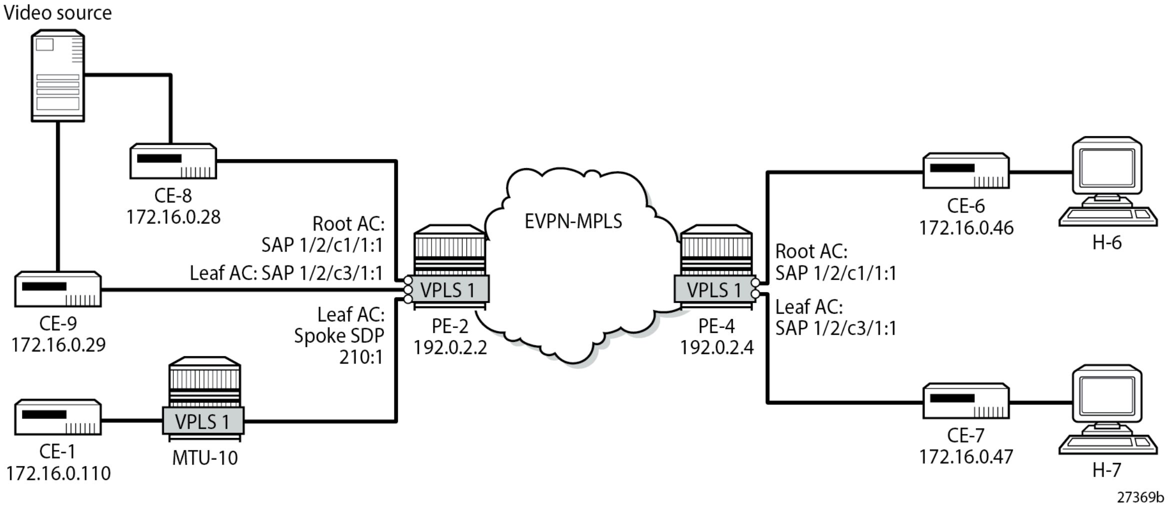
In this section, the following cases are described:

- EVPN-MPLS E-Tree without multi-homing
- EVPN-MPLS E-Tree with all-active and single-active multi-homing

EVPN-MPLS E-Tree without Multi-homing

Figure 144: Example Topology for EVPN-MPLS E-Tree without Multi-homing shows an example topology with two PEs in an EVPN-MPLS network with VPLS 1 configured as E-Tree. CE-6 and CE-8 have root ACs and are able to send and receive traffic to and from all other CEs, whereas CE-7, CE-9, and CE-1 are only able to exchange traffic with CE-6 and CE-8, but not with each other. The video source can be connected to CE-8 (root AC) or CE-9 (leaf AC).

Figure 144: Example Topology for EVPN-MPLS E-Tree without Multi-homing



The service configuration on PE-2 is as follows:

```
On PE-2:
configure
    service
        sdp 210 mpls create
            far-end 192.0.2.10
            ldp
            no shutdown
        exit
        system
            bgp-evpn
                evpn-etree-leaf-label
            exit
        exit
        vpls 1 name "VPLS 1" customer 1 etree create
            bgp
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    ingress-replication-bum-label
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            sap 1/2/c1/1:1 create
            exit
            sap 1/2/c3/1:1 leaf-ac create
            exit
            spoke-sdp 210:1 leaf-ac create
            exit
            no shutdown
        exit
    exit
exit
```

The service configuration on PE-4 is similar, with SAP 1/2/c1/1:1 as root AC and SAP 1/2/c3/1:1 as leaf AC.

The following command on PE-2 shows that SAP 1/2/c1/1:1 is a root AC (default), SAP 1/2/c3/1:1 is a leaf AC (indicated by "L"), and spoke SDP 210:1 is also a leaf AC.

```
*A:PE-2# show service id 1 etree

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : VPLS
---snip---
Etree Mode        : Enabled
Admin State       : Up                  Oper State        : Up
---snip---
-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/2/c1/1:1                           q-tag        8936    8936    Up   Up
sap:1/2/c3/1:1 (L)                       q-tag        8936    8936    Up   Up
sdp:210:1 (L) S(192.0.2.10)              Spok         0       8910    Up   Up
-------------------------------------------------------------------------------
Legend: (L): Leaf-Ac, (RL): Root-Leaf-Tag
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

The following command on PE-2 shows that SAP 1/2/c1/1:1 is not configured as a leaf AC (Leaf-Ac Disabled), while SAP 1/2/c3/1:1 is configured as a leaf AC. Root-leaf tag cannot be configured on objects in an EVPN-MPLS E-Tree, so this is always disabled and no leaf tag is defined.

```
*A:PE-2# show service sap-using etree

===========================================================================
Etree SAP Information
===========================================================================
Svc Id      SAP                               Leaf-Tag    Root-    Leaf-Ac
                                                          leaf-tag
---------------------------------------------------------------------------
1           1/2/c1/1:1                        0           Disabled Disabled
1           1/2/c3/1:1                        0           Disabled Enabled
---snip---
---------------------------------------------------------------------------
Number of etree saps: 5
```

```
===========================================================================
```

Likewise, the following command shows that spoke SDP 210:1 is configured as a leaf AC. Again, root-leaf tag cannot be configured on an object in an EVPN-MPLS E-Tree.

```
*A:PE-2# show service sdp-using etree

===========================================================================
Etree SDP-BIND Information
===========================================================================
Svc Id      SDP-BIND                         Type      Root-     Leaf-Ac
                                                       leaf-tag
---------------------------------------------------------------------------
1           210:1                            Spoke     Disabled  Enabled
---snip---
---------------------------------------------------------------------------
Number of etree sdp-binds: 3
===========================================================================
```

EVPN E-Tree Known Unicast Ingress Filtering

Unicast traffic can be exchanged between CE-8 (root AC) and any other CE. However, unicast traffic from CE-9 on leaf AC can only be exchanged with CE-8 and CE-6 on root ACs, but not with CE-7 (via leaf AC SAP 1/2/c3/1:1) or CE-1 (via leaf AC spoke SDP 210:1), as follows:

```
*A:CE-9# ping 172.16.0.28 rapid     # succeeds - leaf AC can send to root AC
PING 172.16.0.28 56 data bytes
!!!!!
---- 172.16.0.28 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 2.51ms, avg = 3.35ms, max = 6.28ms, stddev = 1.47ms
*A:CE-9# ping 172.16.0.46 rapid     # succeeds - leaf AC can send to root AC
PING 172.16.0.46 56 data bytes
!!!!!
---- 172.16.0.46 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 3.42ms, avg = 3.57ms, max = 3.88ms, stddev = 0.168ms
*A:CE-9# ping 172.16.0.47 rapid     # fails - leaf AC cannot send to leaf AC!
PING 172.16.0.47 56 data bytes
.....
---- 172.16.0.47 PING Statistics ----
5 packets transmitted, 0 packets received, 100% packet loss
*A:CE-9# ping 172.16.0.110 rapid    # fails - leaf AC cannot send to leaf AC!
PING 172.16.0.110 56 data bytes
.....
---- 172.16.0.110 PING Statistics ----
5 packets transmitted, 0 packets received, 100% packet loss
```

The following FDB for VPLS 1 on PE-2 shows that MAC address ca:fe:07:47:47:47 of CE-7 is learned as EVPN leaf, whereas MAC address ca:fe:01:01:01:01 of CE-1 is learned on the local root spoke SDP.

```
*A:PE-2# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
```

```
1          ca:fe:01:01:01:01 sdp:210:1               L/30     07/26/23 21:55:11
1          ca:fe:06:46:46:46 mpls-1:                 Evpn     07/26/23 21:52:46
                             192.0.2.4:524281
           ldp:65538
1          ca:fe:07:47:47:47 mpls-1:                 Evpn,Lf  07/26/23 21:52:46
                             192.0.2.4:524281
           ldp:65538
1          ca:fe:08:28:28:28 sap:1/2/c1/1:1          L/0      07/26/23 21:51:58
1          ca:fe:09:29:29:29 sap:1/2/c3/1:1          LT/0     07/26/23 21:51:52
-------------------------------------------------------------------------------
No. of MAC Entries: 5
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

EVPN E-Tree BUM Egress Filtering

When multicast traffic is sent from a video source via CE-8 (root AC), both CE-6 and CE-7 receive this traffic; for multicast traffic sent via CE-9 (leaf AC), only CE-6 (root AC) receives this traffic. PE-2 received leaf label 524282 in an AD per-ES route from PE-4, as follows:

```
*A:PE-2# show router bgp routes evpn auto-disc rd 192.0.2.4:1 detail
===============================================================================
 BGP Router ID:192.0.2.2        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Original Attributes

Network        : n/a
Nexthop        : 192.0.2.4
Path Id        : None
From           : 192.0.2.4
Res. Nexthop   : 192.168.24.2
Local Pref.    : 100                    Interface Name : int-PE-2-PE-4
---snip---
Community      : target:64496:1 etree::L:0/Leaf-Label:524282
                 bgp-tunnel-encap:MPLS
---snip---
EVPN type      : AUTO-DISC
ESI            : ESI-0
Tag            : MAX-ET
Route Dist.    : 192.0.2.4:1
MPLS Label     : LABEL 0
---snip---
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

Multicast traffic is sent with three labels: MPLS (LDP), EVPN, and leaf label. The EVPN label is 524280 for multicast, as follows:

```
*A:PE-2# show service id 1 evpn-mpls
```

```
===============================================================================
BGP EVPN-MPLS Dest (Instance 1)
===============================================================================
TEP Address                     Transport:Tnl     Egr Label  Oper  Mcast  Num
                                                             State        MACs
-------------------------------------------------------------------------------
192.0.2.4                       ldp:65538         524280     Up    bum    0
192.0.2.4                       ldp:65538         524281     Up    none   2
-------------------------------------------------------------------------------
Number of entries: 2
-------------------------------------------------------------------------------
---snip---
===============================================================================
```

The MPLS transport label is 524287, as follows:

```
*A:PE-2# show router ldp bindings active prefixes prefix 192.0.2.4/32

===============================================================================
---snip---
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                      Op
IngLbl                                      EgrLbl
EgrNextHop                                  EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.4/32                                Push
  --                                        524287
192.168.24.2                                1/1/c1/1

192.0.2.4/32                                Swap
524285                                      524287
192.168.24.2                                1/1/c1/1

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
```

The video source sends the following multicast stream via CE-9 (leaf AC):

```
*A:CE-9# show router pim group detail

===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 232.1.1.1
Source Address     : 192.168.55.2
---snip---
Rpf Neighbor       : 192.168.19.1
Incoming Intf      : int-CE-9-CE-1
Outgoing Intf List : int-CE-9-PE-2

Curr Fwding Rate   : 8239.920 kbps
Forwarded Packets  : 28803              Discarded Packets  : 0
Forwarded Octets   : 42686046           RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

Receiver H-6 has joined the multicast stream and CE-6 (root AC) receives the following multicast group:

```
*A:CE-6# show router pim group detail

===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 232.1.1.1
Source Address     : 192.168.55.2
---snip---
Rpf Neighbor       : 172.16.0.29
Incoming Intf      : int-CE-6-PE-4
Outgoing Intf List : int-CE-6-H-6

Curr Fwding Rate   : 9123.192 kbps
Forwarded Packets  : 24297              Discarded Packets  : 0
Forwarded Octets   : 36008154           RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

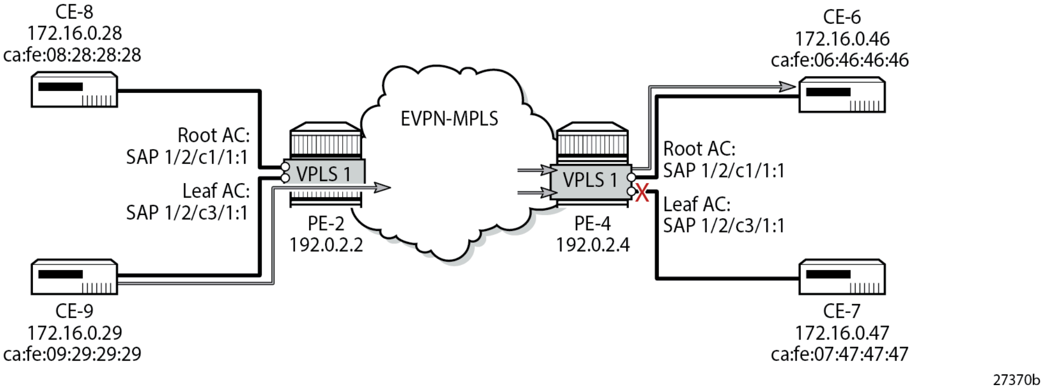
Receiver H-7 has also joined the multicast stream, but CE-7 (leaf AC) cannot receive BUM traffic from a leaf AC, so the forwarding rate is 0 kbps, as follows:

```
*A:CE-7# show router pim group detail

===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 232.1.1.1
Source Address     : 192.168.55.2
---snip---
Rpf Neighbor       :
Incoming Intf      :
Outgoing Intf List : int-CE-7-H-7

Curr Fwding Rate   : 0.000 kbps
Forwarded Packets  : 0                  Discarded Packets  : 0
Forwarded Octets   : 0                  RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

EVPN E-Tree Egress Filtering Based on MAC SA

Egress filtering on MAC SA is required to cover cases when the ingress PE sends traffic received on a leaf AC, but without leaf indication. Figure 145: EVPN E-Tree Egress Filtering Based on MAC SA shows that CE-9 sends traffic with MAC SA ca:fe:09:29:29:29 on a leaf AC.

Figure 145: EVPN E-Tree Egress Filtering Based on MAC SA



When CE-9 sends unicast traffic to CE-6 with root MAC DA ca:fe:06:46:46:46, the ingress PE-2 forwards the frames to this root MAC DA to egress PE-4. However, if PE-4 does not have the MAC DA in its FDB (because of aging or MAC flush and the MAC route has not made it yet to PE-2), it may flood the frame to all the root and leaf ACs, even if the frame originated from a leaf AC. EVPN E-Tree egress filtering based on MAC SA prevents this from happening, so the traffic is only forwarded to the root AC.

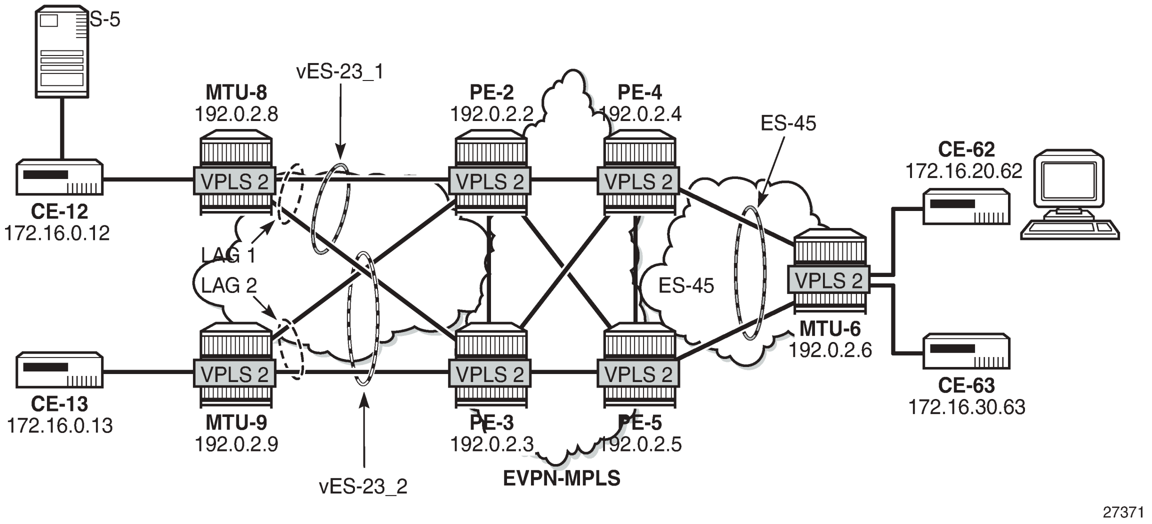
The data path does the egress filtering based on MAC SA as follows:

- First, frames are identified as leaf frames in one of the following cases:
  - Frames arriving on a leaf SAP
  - EVPN traffic arriving with a leaf label
  - Frames arriving with a MAC SA that is flagged as being a leaf SA
- At the egress PE, frames identified as leaf are filtered depending on the type of traffic:
  - For known unicast traffic, the FDB is consulted. If the MAC DA FDB entry is marked as being from a leaf, the frame is dropped to prevent leaf-to-leaf forwarding.
  - For BUM traffic, the leaf frames are filtered at the egress IOM to suppress leaf-to-leaf forwarding.

EVPN-MPLS E-Tree with Multi-homing

Figure 146: Example Topology with All-active ESs and Single-active ES shows the example topology with two all-active multi-homing vESs on PE-2 and PE-3 and one single-active multi-homing ES on PE-4 and PE-5.

Figure 146: Example Topology with All-active ESs and Single-active ES



On PE-2, two all-active multi-homing vESs are configured. VPLS 2 is configured as EVPN-MPLS E-Tree with LAG 1 as root AC and LAG 2 as leaf AC. RD 2.2.2.2 is configured and used in the non-zero AD per-ES routes, while the zero ESI routes (AD per-ES) use the IP address 192.0.2.2. The service configuration on PE-2 is as follows:

```
On PE-2:
configure
    service
        system
            bgp-evpn
                ad-per-es-route-target evi-rt-set route-distinguisher 2.2.2.2
                evpn-etree-leaf-label
                ethernet-segment "vESI-23_1" virtual create
                    esi 01:00:00:00:00:23:01:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 1
                    dot1q
                        q-tag-range 2
                    exit
                    no shutdown
                exit
                ethernet-segment "vESI-23_2" virtual create
                    esi 01:00:00:00:00:23:02:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 2
                    dot1q
                        q-tag-range 2
                    exit
                    no shutdown
                exit
            exit
        exit
        vpls 2 name "VPLS 2" customer 1 etree create
            bgp
            exit
            bgp-evpn
                evi 2
                mpls bgp 1
                    ingress-replication-bum-label
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            sap lag-1:2 create
            exit
            sap lag-2:2 leaf-ac create
            exit
            no shutdown
        exit
    exit
exit
```

The service configuration on PE-3 is identical, but with **evi-rt-set route-distinguisher** 3.3.3.3 instead.

Note:
The command **config service system bgp-evpn ad-per-es-route-target evi-rt-set** is not supported for EVPN E-Tree services. When the command is configured on a router, the AD per-ES routes (with ESI=0) used for EVPN E-Tree services are always advertised with the service route target and route distinguisher, regardless of the **ad-per-es-route-target** configuration. AD per-ES routes for non-zero ESIs (used for regular multi-homing) is usually sent using either **evi-rt-set** or **evi-rt**, based on the router configuration.

It is important that all the ACs in each EVI for an ES must either be root ACs or leaf ACs in both PEs where the ES is defined, not a mix. In this example, SAP lag-1:2 is assigned to vES-23_1 and defined as root AC in both PE-2 and PE-3. Likewise, SAP lag-2:2 is assigned to vES-23_2 and configured as leaf AC in PE-2 and PE-3. However, if the configuration were a mix of root and leaf ACs in different PEs of the same ES, a remote PE (PE-4 or PE-5) would receive the AD per-EVI routes with inconsistent leaf indication and would treat the AC as root AC.

PE-2 sends the following BGP EVPN AD routes: an AD per-ES route with zero ESI and RD 192.0.2.2:2 (for egress filtering of BUM traffic) and an EVPN AD per-EVI route with non-zero ESI and RD 2.2.2.2:1 (to verify the ES consistency).

```
On PE-2:
17 2023/07/26 22:32:48.324 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 81
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-AD Len: 25 RD: 192.0.2.2:2 ESI: ESI-0, tag: MAX-ET Label: 0 (Raw Label: 0x0)
 PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64496:2
        etree::L:0/Leaf-Label:524282
        bgp-tunnel-encap:MPLS
"

20 2023/07/26 22:32:48.328 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 73
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-AD Len: 25 RD: 2.2.2.2:1 ESI: 01:00:00:00:00:23:01:00:00:01, tag: MAX-ET
 Label: 0 (Raw Label: 0x0) PathId:
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64496:2
        esi-label:524276/All-Active
"
```

The following command shows the EVI RT set RD ranging from 2.2.2.2:1 to 2.2.2.2:512. In VPLS 2, the configured EVI is 2 and needs to be divided by 128, the number of EVI RT sets that are advertised. This value is rounded up to 1; therefore, the RD in the preceding AD per-EVI equals 2.2.2.2:1. The minimum EVI RT set RD equals 2.2.2.2:1 and the maximum is 2.2.2.2:512, because the EVI ranges from 1 to 65535 and 65536/128=512.

```
*A:PE-2# show service system bgp-evpn

===============================================================================
System BGP EVPN Information
===============================================================================
Eth Seg Route Dist.               : <none>
Eth Seg Oper Route Dist.          : 192.0.2.2:0
Eth Seg Oper Route Dist Type      : default
Ad Per ES Route Target            : evi-rt-set
EVI RT set Route Dist.            : 2.2.2.2:1 - 2.2.2.2:512
Extended Evi Range                : Disabled
Etree
  Leaf                            : Enabled
  Leaf Label                      : 524282 (dynamic)
---snip---
===============================================================================
```

Remote PE-4 received the following EVPN AD per-ES routes from PE-2: two non-zero ESI routes (for vES-23_1 and vES-23_2) and a zero ESI route.

```
*A:PE-4# show router bgp routes evpn auto-disc tag MAX-ET
===============================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Flag  Route Dist.         ESI                           NextHop
      Tag                                               Label
-------------------------------------------------------------------------------
u*>i  2.2.2.2:1           01:00:00:00:00:23:01:00:00:01 192.0.2.2
      MAX-ET                                            LABEL 0

u*>i  2.2.2.2:1           01:00:00:00:00:23:02:00:00:01 192.0.2.2
      MAX-ET                                            LABEL 0
---snip---
u*>i  192.0.2.2:2         ESI-0                         192.0.2.2
      MAX-ET                                            LABEL 0
---snip---
-------------------------------------------------------------------------------
Routes : 8
===============================================================================
```

On PE-4 and PE-5, ES-45 is configured in single-active mode. The service configuration on PE-4 is as follows:

```
On PE-4:
configure
    service
        system
            bgp-evpn
                ad-per-es-route-target evi-rt-set route-distinguisher 4.4.4.4
                evpn-etree-leaf-label
                ethernet-segment "ES-45" create
                    esi 01:00:00:00:00:45:01:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode manual
                        manual
                            preference create
                                value 10000
                            exit
                        exit
                    exit
                    multi-homing single-active
                    sdp 46
                    no shutdown
                exit
            exit
        exit
        vpls 2 name "VPLS 2" customer 1 etree create
            bgp
            exit
            bgp-evpn
                evi 2
                mpls bgp 1
                    ingress-replication-bum-label
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            spoke-sdp 46:2 leaf-ac create
            exit
            no shutdown
        exit
    exit
exit
```

The service configuration is similar on PE-5, but with a lower preference for the ES, so PE-4 is the DF, as follows.

```
*A:PE-4# show service id 2 ethernet-segment
No sap entries

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
46:2                  ES-45                            DF
===============================================================================
No vxlan instance entries
```

For the all-active multi-homing vESs, PE-2 is the DF, as follows:

```
*A:PE-2# show service id 2 ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
lag-1:2               vESI-23_1                        DF
lag-2:2               vESI-23_2                        DF
===============================================================================
No sdp entries
No vxlan instance entries
```

Ingress Filtering for Unicast Traffic

Traffic can be sent between CE-12 (root AC lag-1:2) and CE-62 (leaf AC spoke SDP 46:2), but traffic between CE-13 (leaf AC lag-2:2) and CE-63 (leaf AC spoke SDP 46:2) is filtered. The following FDB for VPLS 2 on PE-2 shows two EVPN leaf MAC addresses: ca:fe:06:00:20:62 for CE-62 and ca:fe:06:00:30:63 for CE-63.

```
*A:PE-2# show service id 2 fdb detail

===============================================================================
Forwarding Database, Service 2
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
2          ca:fe:01:00:20:12 sap:lag-1:2             L/0      07/26/23 22:37:20
2          ca:fe:01:00:30:13 sap:lag-2:2             Evpn     07/26/23 22:37:21
2          ca:fe:06:00:20:62 eES:                    Evpn,Lf  07/26/23 22:33:24
                             01:00:00:00:00:45:01:00:00:01
2          ca:fe:06:00:30:63 eES:                    Evpn,Lf  07/26/23 22:37:43
                             01:00:00:00:00:45:01:00:00:01
-------------------------------------------------------------------------------
No. of MAC Entries: 4
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

The FDB for VPLS 2 on PE-3 shows the same EVPN leaf MAC addresses. For all PEs in the all-active MH ESs, the MAC addresses ca:fe:06:00:20:12 and ca:fe:06:00:30:13 from the locally attached ACs can be learned on the SAPs or via EVPN from the ES peer where they are learned on the SAPs. In this case, they are learned on the SAPs on PE-2 and PE-3.

```
*A:PE-3# show service id 2 fdb detail

===============================================================================
Forwarding Database, Service 2
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
2          ca:fe:01:00:20:12 sap:lag-1:2             L/0      07/26/23 22:33:03
2          ca:fe:01:00:30:13 sap:lag-2:2             L/0      07/26/23 22:37:21
2          ca:fe:06:00:20:62 eES:                    Evpn,Lf  07/26/23 22:33:24
                             01:00:00:00:00:45:01:00:00:01
2          ca:fe:06:00:30:63 eES:                    Evpn,Lf  07/26/23 22:37:43
                             01:00:00:00:00:45:01:00:00:01
-------------------------------------------------------------------------------
No. of MAC Entries: 4
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

The following FDB for VPLS 2 on DF PE-4 shows one EVPN leaf MAC address: ca:fe:01:00:30:13 for CE-13 on a remote ES.

```
*A:PE-4# show service id 2 fdb detail

===============================================================================
Forwarding Database, Service 2
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
2          ca:fe:01:00:20:12 eES:                    Evpn     07/26/23 22:33:19
                             01:00:00:00:00:23:01:00:00:01
2          ca:fe:01:00:30:13 eES:                    Evpn,Lf  07/26/23 22:37:21
                             01:00:00:00:00:23:02:00:00:01
2          ca:fe:06:00:20:62 sdp:46:2                L/0      07/26/23 22:33:24
2          ca:fe:06:00:30:63 sdp:46:2                L/4      07/26/23 22:37:43
-------------------------------------------------------------------------------
No. of MAC Entries: 4
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

PE-5 is NDF, and the following FDB shows three MAC routes of type EVPN leaf, for CE-13, CE-62, and CE-63.

```
*A:PE-5# show service id 2 fdb detail

===============================================================================
Forwarding Database, Service 2
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
2          ca:fe:01:00:20:12 eES:                    Evpn     07/26/23 22:33:33
                             01:00:00:00:00:23:01:00:00:01
```

```
2          ca:fe:01:00:30:13 eES:                       Evpn,Lf  07/26/23 22:37:21
                             01:00:00:00:00:23:02:00:00:01
2          ca:fe:06:00:20:62 eES:                       Evpn,Lf  07/26/23 22:33:33
                             01:00:00:00:00:45:01:00:00:01
2          ca:fe:06:00:30:63 eES:                       Evpn,Lf  07/26/23 22:37:43
                             01:00:00:00:00:45:01:00:00:01
-------------------------------------------------------------------------------
No. of MAC Entries: 4
-------------------------------------------------------------------------------
Legend:L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf T=Trusted
===============================================================================
```

Egress Filtering for BUM Traffic

Each PE advertises zero ESI AD per-ES routes (with tag MAX-ET) that are needed for egress BUM filtering.

BUM frames received on an ES root AC are flooded to the EVPN, based on regular EVPN procedures. The regular ESI label is sent for split horizon when frames are sent to the DF or NDF PEs in the same ES.

BUM frames received on an ES leaf AC are flooded in the default multicast list. The egress PE does not forward BUM traffic to any leaf ACs, including the ES leaf ACs. However, in the unlikely event that some ACs in a specific ES for an EVI have an inconsistent E-Tree configuration, these ACs are treated as root ACs, and the traffic is forwarded.

The remote PE-4 receives the following EVPN AD routes from DF PE-2: a zero ESI AD per-ES (tag MAX-ET), two AD per-EVI (tag 0) routes with a non-zero label, and two AD per-ES routes (tag MAX-ET).

```
*A:PE-4# show router bgp routes evpn auto-disc next-hop 192.0.2.2
===============================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Flag  Route Dist.         ESI                           NextHop
      Tag                                               Label
-------------------------------------------------------------------------------
u*>i  2.2.2.2:1           01:00:00:00:00:23:01:00:00:01 192.0.2.2
      MAX-ET                                            LABEL 0

u*>i  2.2.2.2:1           01:00:00:00:00:23:02:00:00:01 192.0.2.2
      MAX-ET                                            LABEL 0

u*>i  192.0.2.2:2         ESI-0                         192.0.2.2
      MAX-ET                                            LABEL 0

u*>i  192.0.2.2:2         01:00:00:00:00:23:01:00:00:01 192.0.2.2
      0                                                 LABEL 524274

u*>i  192.0.2.2:2         01:00:00:00:00:23:02:00:00:01 192.0.2.2
      0                                                 LABEL 524274

-------------------------------------------------------------------------------
Routes : 5
```

```
===============================================================================
```

The same remote PE-4 receives similar EVPN AD routes from NDF PE-3: a zero ESI AD per-ES (tag MAX-ET), two AD per-EVI (tag 0) routes with a non-zero label, and two AD per-ES routes (tag MAX-ET).

```
*A:PE-4# show router bgp routes evpn auto-disc next-hop 192.0.2.3
===============================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Flag  Route Dist.         ESI                           NextHop
      Tag                                               Label
-------------------------------------------------------------------------------
u*>i  3.3.3.3:1           01:00:00:00:00:23:01:00:00:01 192.0.2.3
      MAX-ET                                            LABEL 0

u*>i  3.3.3.3:1           01:00:00:00:00:23:02:00:00:01 192.0.2.3
      MAX-ET                                            LABEL 0

u*>i  192.0.2.3:2         ESI-0                         192.0.2.3
      MAX-ET                                            LABEL 0

u*>i  192.0.2.3:2         01:00:00:00:00:23:01:00:00:01 192.0.2.3
      0                                                 LABEL 524278

u*>i  192.0.2.3:2         01:00:00:00:00:23:02:00:00:01 192.0.2.3
      0                                                 LABEL 524278

-------------------------------------------------------------------------------
Routes : 5
===============================================================================
```

The following detailed information about the AD per-ES route (tag MAX-ET) for mass withdraw on PE-4 shows that no E-Tree extended community is sent by PE-2; only the ESI-label extended community is sent.

```
*A:PE-4# show router bgp routes evpn auto-disc rd 2.2.2.2:1 tag MAX-ET esi
 01:00:00:00:00:23:01:00:00:01 detail
===============================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Original Attributes

Network        : n/a
Nexthop        : 192.0.2.2
Path Id        : None
From           : 192.0.2.2
Res. Nexthop   : 192.168.24.1
```

```
---snip---
Community      : target:64496:2 esi-label:524276/All-Active
---snip---
EVPN type      : AUTO-DISC
ESI            : 01:00:00:00:00:23:01:00:00:01
Tag            : MAX-ET
Route Dist.    : 2.2.2.2:1
MPLS Label     : LABEL 0
---snip---
--------------------------------------------------------------------------------
Routes : 1
================================================================================
```

A similar result is seen for the other vES:

```
*A:PE-4# show router bgp routes evpn auto-disc rd 2.2.2.2:1 tag MAX-ET esi
 01:00:00:00:00:23:02:00:00:01 detail
================================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
================================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

================================================================================
BGP EVPN Auto-Disc Routes
================================================================================
Original Attributes

Network        : n/a
Nexthop        : 192.0.2.2
Path Id        : None
From           : 192.0.2.2
Res. Nexthop   : 192.168.24.1
---snip---
Community      : target:64496:2 esi-label:524275/All-Active
---snip---
EVPN type      : AUTO-DISC
ESI            : 01:00:00:00:00:23:02:00:00:01
Tag            : MAX-ET
Route Dist.    : 2.2.2.2:1
MPLS Label     : LABEL 0
---snip---
--------------------------------------------------------------------------------
Routes : 1
================================================================================
```

The following detailed information about the AD per-EVI (tag 0) on PE-4 shows that if the ES is root (as for vES-23_1), the regular extended community is sent, not the E-Tree extended community.

```
*A:PE-4# show router bgp routes evpn auto-disc rd 192.0.2.2:2 tag 0 esi
 01:00:00:00:00:23:01:00:00:01 detail
================================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
================================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

================================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Original Attributes

Network        : n/a
Nexthop        : 192.0.2.2
Path Id        : None
From           : 192.0.2.2
Res. Nexthop   : 192.168.24.1
---snip---
Community      : target:64496:2 bgp-tunnel-encap:MPLS
---snip---
EVPN type      : AUTO-DISC
ESI            : 01:00:00:00:00:23:01:00:00:01
Tag            : 0
Route Dist.    : 192.0.2.2:2
MPLS Label     : LABEL 524274
---snip---
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

The following detailed information about the AD per-EVI (tag 0) on PE-4 shows that if the ES is leaf (as for vES-23_2), the E-Tree extended community is sent, along with the regular extended community.

```
*A:PE-4# show router bgp routes evpn auto-disc rd 192.0.2.2:2 tag 0 esi
 01:00:00:00:00:23:02:00:00:01 detail
===============================================================================
 BGP Router ID:192.0.2.4        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Original Attributes

Network        : n/a
Nexthop        : 192.0.2.2
Path Id        : None
From           : 192.0.2.2
Res. Nexthop   : 192.168.24.1
---snip---
Community      : target:64496:2 etree::L:1/Leaf-Label:0
                 bgp-tunnel-encap:MPLS
---snip---
EVPN type      : AUTO-DISC
ESI            : 01:00:00:00:00:23:02:00:00:01
Tag            : 0
Route Dist.    : 192.0.2.2:2
MPLS Label     : LABEL 524274
---snip---
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

The **tools dump service evpn usage** command shows that there are three EVPN E-Tree remote BUM leaf labels:

```
*A:PE-2# tools dump service evpn usage | match "BUM"
Evpn Etree Remote BUM Leaf Labels               :             3
```

This corresponds to the following three ESI-0 AD per-ES routes (tag MAX-ET) on PE-2:

```
*A:PE-2# show router bgp routes evpn auto-disc esi ESI-0
===============================================================================
 BGP Router ID:192.0.2.2        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Flag  Route Dist.         ESI                           NextHop
      Tag                                               Label
-------------------------------------------------------------------------------
u*>i  192.0.2.3:2         ESI-0                         192.0.2.3
      MAX-ET                                            LABEL 0

u*>i  192.0.2.4:2         ESI-0                         192.0.2.4
      MAX-ET                                            LABEL 0

u*>i  192.0.2.5:2         ESI-0                         192.0.2.5
      MAX-ET                                            LABEL 0

-------------------------------------------------------------------------------
Routes : 3
===============================================================================
```

Conclusion

E-Trees can be used for enterprise business services, for the distribution of IPTV multicast content, for centralized backup BNGs, and so on. In a VPLS E-Tree, leaf SAPs or leaf SDP bindings cannot exchange traffic with each other, similar to split horizon group behavior. The E-Tree restrictions apply to all remote PEs that are part of the same service. E-Trees can be applied in an EVPN-MPLS VPLS as well as in a regular VPLS.

EVPN-MPLS Interconnect for EVPN-VXLAN VPLS Services

This chapter provides information about EVPN-MPLS Interconnect for EVPN-VXLAN VPLS Services.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS Release 14.0.R5, but the CLI in the current edition is based on SR OS Release 21.2.R1.

Chapters EVPN for MPLS Tunnels and EVPN for VXLAN Tunnels (Layer 2) are prerequisite reading.

Overview

When EVPN-MPLS is deployed in the WAN, many service providers are looking for a way to integrate existing Layer 2 EVPN-VXLAN based data center services into the WAN, while keeping the end-to-end advantages of EVPN. The IETF draft-ietf-bess-dci-evpn-overlay describes how to provide Layer 2 connectivity for EVPN-overlay data centers in different ways. This chapter follows section 4.4 of that document, in which EVPN-MPLS is used in the same VPLS service that terminates overlay (VXLAN) tunnels.

To provide EVPN-MPLS connectivity to VPLS services terminating EVPN-VXLAN, SR OS supports the configuration of BGP-EVPN MPLS and BGP-EVPN VXLAN at the same time by adding two BGP instances to the service. Two BGP instances are supported in the same VPLS at most. BGP-EVPN MPLS and BGP-EVPN VXLAN can both use BGP instance 1 or 2, but they must use different instances.

In a service with EVPN-VXLAN and EVPN-MPLS, the **config service vpls bgp-evpn mpls bgp 2** command allows the user to associate BGP-EVPN MPLS to a different instance than BGP-EVPN VXLAN, and therefore, have both encapsulations simultaneously enabled in the same service. When the two BGP instances are successfully added to the same VPLS service, the service behaves as follows:

- MAC/IP routes received on one instance will be "consumed" (accepted, imported, and installed in FDB) and re-advertised in the other instance, as long as the route is the best route for a specific MAC or MAC/IP.
- Inclusive multicast routes are independently generated for each BGP instance.
- From a data plane perspective, EVPN-MPLS and EVPN-VXLAN destinations are instantiated in different implicit Split-Horizon Groups (SHGs) so that traffic can be forwarded between the two SHGs,

but not between destinations of the same kind. For example, traffic coming from EVPN-MPLS cannot be forwarded to other destinations in the EVPN-MPLS SHG.

The following example shows a VPLS service configured on PE-2 with two BGP instances and both encapsulations, VXLAN and MPLS, configured at the same time:

```
# on PE-2:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            description "evpn-mpls and evpn-vxlan in the same service"
            vxlan instance 1 vni 1 create
            exit
            bgp
                route-distinguisher 10:1
                route-target export target:64500:1 import target:64500:1
            exit
            bgp 2
                route-distinguisher 10:2
                route-target export target:64500:1 import target:64500:1
            exit
            bgp-evpn
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
                mpls bgp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
```

In the preceding example

- **bgp 1** or simply **bgp** is the default BGP instance.
- **bgp 2** is the additional instance that is required when both BGP-EVPN VXLAN and BGP-EVPN MPLS are enabled in the service.
- The same commands supported under BGP instance 1 exist for this second BGP instance, with the following considerations:
  - **pw-template-binding** – the pseudowire (PW) template binding can only exist in BGP instance 1; it is not supported in BGP instance 2. Because no SDP-bindings can exist in a VPLS service with two BGP instances, the **pw-template-binding** command is ineffective in this configuration.
  - **route-distinguisher** – the route distinguisher in both BGP instances must be different.
  - **route-target** – the route target in both instances can be the same or different.
  - **vsi-import** and **vsi-export** – import and export policies can also be defined for either BGP instance.
- The **mpls bgp 2** command will assign BGP instance 2 to MPLS. The **bgp-evpn vxlan bgp 1 vxlan-instance 1 no shutdown** command will only be allowed if **bgp-evpn** is **shutdown** or if the BGP instance associated with MPLS has a different route distinguisher than the VXLAN instance (and vice versa).

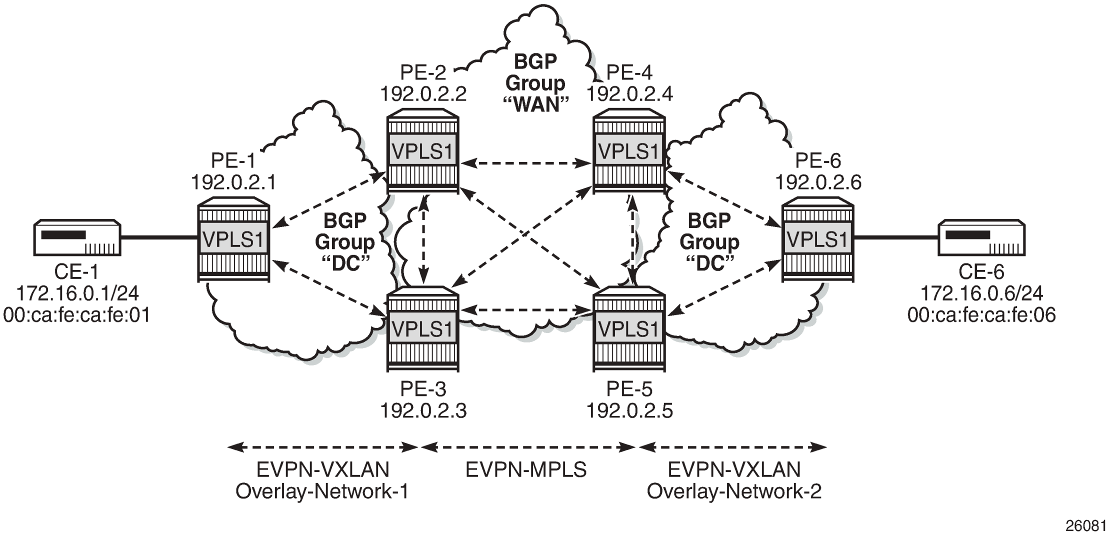
- The **evi** can still be used for auto-derivation of RD/RT on BGP instance 1 and auto-derivation of RT (not RD) on BGP instance 2. Auto-RD or an explicitly configured RD is needed in BGP instance 2.

Configuration

Figure 147: EVPN-MPLS interconnect for EVPN-VXLAN - example topology shows the example topology that will be used throughout this chapter, as well as the BGP peering topology. PE-1, PE-2, and PE-3 simulate a data center, shown as Overlay-Network-1, where PE-2 and PE-3 are DC GWs. In the same way, PE-4, PE-5, and PE-6 simulate a remote data center, Overlay-Network-2. Inside each DC, EVPN-VXLAN is used.

The two DC GW pairs are connected by EVPN-MPLS; therefore, CE-1 and CE-6 are end-to-end connected by EVPN without any VLAN or PW hand-off, maintaining all the EVPN advantages across the DC Interconnect (DCI) network.

Figure 147: EVPN-MPLS interconnect for EVPN-VXLAN - example topology



The example topology consists of six 7750 SR routers with the following initial configuration:

- Hybrid ports (they could have been network type too) are interconnecting the six PEs with configured router interfaces.
- The six PEs are running IS-IS and creating point-to-point adjacencies.
- Link LDP is configured in the core, among PE-2, PE-3, PE-4, and PE-5, while PE-1 and PE-6 are only running VXLAN.
- EVPN uses MP-BGP for exchanging reachability at service level. Therefore, BGP peering sessions must be established among the PEs for the EVPN family. Figure 147: EVPN-MPLS interconnect for EVPN-VXLAN - example topology shows the peering sessions established among the six PEs. Although usually a Route-Reflector (RR) is used in each DC and another RR in the WAN, in this example, there are direct peering sessions in each DC and in the WAN.

The following output shows the BGP configuration of PE-2. The BGP configuration on the rest of the DC GWs (PE-3, PE-4, and PE-5) is similar:

```
# on PE-2:
configure
    router Base
        bgp
            family evpn
            vpn-apply-import
            vpn-apply-export
            rapid-withdrawal
            rapid-update evpn
            group "DC"
                type internal
                import "drop SOO-DCGW-23"
                export "allow only vxlan and add SOO"
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.3
                exit
            exit
            group "WAN"
                type internal
                import "drop SOO-DCGW-23"
                export "allow only mpls and add SOO"
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
            no shutdown
```

Two different BGP groups are configured: DC and WAN. The DC group contains the DC neighbors (including the peer DC GW) and the WAN group contains the WAN neighbors. This grouping makes the use of policies easier. These policies will be explained in the section The mandatory use of BGP policies in the multi-homed anycast solution.

The following output shows the BGP configuration of PE-1. PE-6 has a similar BGP configuration.

```
# on PE-1:
configure
    router Base
        bgp
            family evpn
            rapid-withdrawal
            rapid-update evpn
            group "DC"
                type internal
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
            no shutdown
```

VPLS service configuration

After the base infrastructure (interfaces, IGP, LDP in the core, and BGP) is configured, the services can be added. The configuration example in this section will use VPLS 1 as the service to be interconnected across the two DCs.

PE-1 and PE-6 have a regular EVPN-VXLAN configuration; DCI connectivity provided by EVPN-MPLS is completely transparent to them. The configuration of VPLS 1 in PE-1 is as follows:

```
# on PE-1:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            vxlan instance 1 vni 1 create
            exit
            bgp
            exit
            bgp-evpn
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/2/1:1 create
                no shutdown
            exit
            no shutdown
        exit
```

See the EVPN for VXLAN Tunnels (Layer 2) chapter for a complete description of the EVPN-VXLAN commands.

The configuration on PE-2, PE-3, PE-4, and PE-5 (see Figure 147: EVPN-MPLS interconnect for EVPN-VXLAN - example topology) enables EVPN-VXLAN and EVPN-MPLS in the same VPLS service. As an example, the VPLS 1 configuration on PE-2 is as follows:

```
# on PE-2:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            vxlan instance 1 vni 1 create
            exit
            bgp
                route-distinguisher 64500:1
            exit
            bgp 2
                route-distinguisher 64500:2
            exit
            bgp-evpn
                incl-mcast-orig-ip 23.23.23.23
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
                mpls bgp 2
                    ingress-replication-bum-label
                    auto-bind-tunnel
```

```
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
```

As described in the Overview section, the preceding configuration enables the router to create EVPN-VXLAN and EVPN-MPLS destinations in the same VPLS service, but in different SHGs. In addition to the **bgp 2** commands already described in the Overview section, the **incl-mcast-orig-ip command** is added in the configuration. If configured, this command will change the originating IP address in the inclusive multicast routes (from the default system IP) for both BGP instances. The section Multi-homed anycast configuration for dual BGP-instance VPLS services describes why this command is added.

The following section provides a detailed description of the expected behavior for EVPN routes that are imported and exported on dual BGP instance VPLS services.

EVPN route handling in dual BGP-instance VPLS services

This section describes how the BGP-EVPN routes are processed in dual BGP instance services.

Usually, the router validates the received tunnel encapsulation (from the RFC 5512 Extended Community) with the configured encapsulation of the service/BGP-instance. Therefore, an EVPN-VXLAN route will not get imported into the BGP-EVPN MPLS instance and vice-versa. This is also how the different EVPN route types are handled in dual BGP instance services:

- *Route type 1 - auto-discovery routes*

  AD per-EVI routes are never generated by services with two BGP instances (because no Ethernet Segment (ES) can be associated with the dual BGP instance service). However, AD per-EVI routes can still be received from the EVPN-MPLS peers and are processed as usual. Therefore, a VPLS service with two BGP instances will still support aliasing/backup and AD per-ES checking procedures for a remote multi-homed ES, as described in the EVPN for MPLS Tunnels chapter. However, in the example in Figure 147: EVPN-MPLS interconnect for EVPN-VXLAN - example topology, PE-6 does not have any local multi-homed ES configured; therefore, no AD per-EVI routes are present in this example.

- *Route type 2 - MAC/IP routes*

  MAC/IP routes received on one of the two BGP instances will be imported and the MAC addresses added to the FDB according to the existing selection rules. If the MAC address is active (therefore installed in the FDB), it will be re-advertised in the other BGP instance with the BGP attributes of the other BGP instance (new route target if different, new route distinguisher, and so on). The **mac-advertisement** command will govern the advertisement of MAC addresses in either BGP instance.

  The MAC/IP route redistribution across BGP instances is performed according to the following rules:

  – A MAC route is redistributed only if it is the best route according to the EVPN selection rules in the EVPN for MPLS Tunnels chapter.

  – Assuming a specific MAC route is the best one and has to be redistributed, the MAC/IP information along with the sticky bit is propagated in the redistribution.

  – A change in the MAC/IP route sequence number or sticky bit in one instance is updated in the other instance, as long as that route is the best MAC route for the route key.

– When a MAC address moves within the EVPN-VXLAN (or the EVPN-MPLS) network, the MAC route is received on the same BGP instance where it was previously received, but now with a higher sequence number. In this case, the MAC route will be redistributed with the new sequence number. However, a router with two BGP instances in the same service will not detect any duplicate MAC on the EVPN-VXLAN and EVPN-MPLS networks.

As an example, the following output shows the debug of a MAC/IP route received on PE-2, on the BGP instance for EVPN-VXLAN on VPLS 1, and how the route is re-advertised to the BGP instance used for MPLS (with a different next-hop, route distinguisher, label, and BGP tunnel encapsulation):

```
# on PE-2:
18 2021/03/15 16:39:03.570 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 81
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.1
        Type: EVPN-MAC Len: 33 RD: 192.0.2.1:1 ESI: ESI-0, tag: 0, mac len: 48
                       mac: 00:ca:fe:ca:fe:01, IP len: 0, IP: NULL, label1: 1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:1
        bgp-tunnel-encap:VXLAN
"
```

```
19 2021/03/15 16:39:03.570 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 89
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-MAC Len: 33 RD: 64500:2 ESI: ESI-0, tag: 0, mac len: 48
                       mac: 00:ca:fe:ca:fe:01, IP len: 0, IP: NULL, label1: 8388496
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        origin:64500:23
        target:64500:1
        bgp-tunnel-encap:MPLS
"
```

```
20 2021/03/15 16:39:03.571 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.5
"Peer 1: 192.0.2.5: UPDATE
Peer 1: 192.0.2.5 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 89
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-MAC Len: 33 RD: 64500:2 ESI: ESI-0, tag: 0, mac len: 48
                       mac: 00:ca:fe:ca:fe:01, IP len: 0, IP: NULL, label1: 8388496
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
```

```
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        origin:64500:23
        target:64500:1
        bgp-tunnel-encap:MPLS
"
```

- *Route type 3 - inclusive multicast routes*

  EVPN Inclusive Multicast Ethernet Tag (IMET) routes are generated independently for each BGP instance with the correct BGP tunnel encapsulation extended community and the tunnel type associated to the BGP instance; for example, Ingress Replication (IR), P2MP mLDP, or Assisted Replication (AR):

  – On the EVPN-VXLAN BGP instance, IR or AR IMET routes are supported.

    - When **assisted-replication replicator** is enabled and the received VXLAN broadcast and multicast packets contain an IP DA = AR-IP, the DC GW will send the packets back to VXLAN (but not to the VXLAN termination end-point (VTEP) from where the packet is received) in addition to the EVPN-MPLS destinations.
    - If **assisted-replication replicator** is used on the DC GWs, the AR-IP (**configure>service>system>vxlan>assisted-replication-ip**) must be a loopback different from the router's system IP and the configured **bgp-evpn>incl-mcast-orig-ip**. The two AR-IP addresses in the DC GW pair do not need to be the same IP address.

  – On the EVPN-MPLS BGP instance, IR, P2MP mLDP, or composite IMET routes are supported.

  – Following is the behavior when the **incl-mcast-orig-ip** command is used:

    - The configured IP in the **incl-mcast-orig-ip** command is encoded in the originating IP field of the IMET routes for IR, P2MP, and composite routes for both BGP instances.
    - The originating IP field of the IMET AR routes is still derived from the configured **service>system>vxlan>assisted-replication-ip** value.

  – The received IMET routes will be processed in the following way depending on their type:

    - IMET-IR routes: the EVPN destination (MPLS or VXLAN) is set up based on the NLRI next-hop.
    - IMET-P2MP routes: the Provider Multicast Service Interface (PMSI) Tunnel Attribute (PTA) tunnel ID will be used to join the mLDP tree (as mLDP FEC in the LDP mapping messages).
    - IMET-P2MP-IR (composite) routes: the PTA tunnel ID is used to join the mLDP tree. The NLRI next-hop is used to build the EVPN destination.
    - IMET-AR routes: the NLRI next-hop is used to build the EVPN-VXLAN destination.

  – Upon reception of two IMET routes with similar information, the router behaves as follows:

    - If the router receives two IMET routes with the same originating IP, different RDs, and different NLRI next-hops, it will set up two EVPN destinations, one to each next-hop.
    - If the router gets two IMET routes with the same originating IP, different RDs, but the same next-hop, it will set up only one EVPN destination.
    - The router will not set up an EVPN destination to its DC GW peer if the received originating IP matches its own originating IP, regardless of whether the local RD and the remote RD are the same or different. This enables the use of the redundant anycast solution that is described in the following section: Multi-homed anycast configuration for dual BGP-instance VPLS services.

- *Route type 4 - ES routes*

ESs are supported in routers where dual BGP-instance services exist. However, because dual BGP-instance VPLS services do not support SDP-bindings, ESs and ES routes are not relevant to these types of services.

- *Route type 5 - IP-prefix routes*

  R-VPLS services are not supported along with dual BGP instances; therefore, IP-prefix routes are neither generated nor processed by the service.

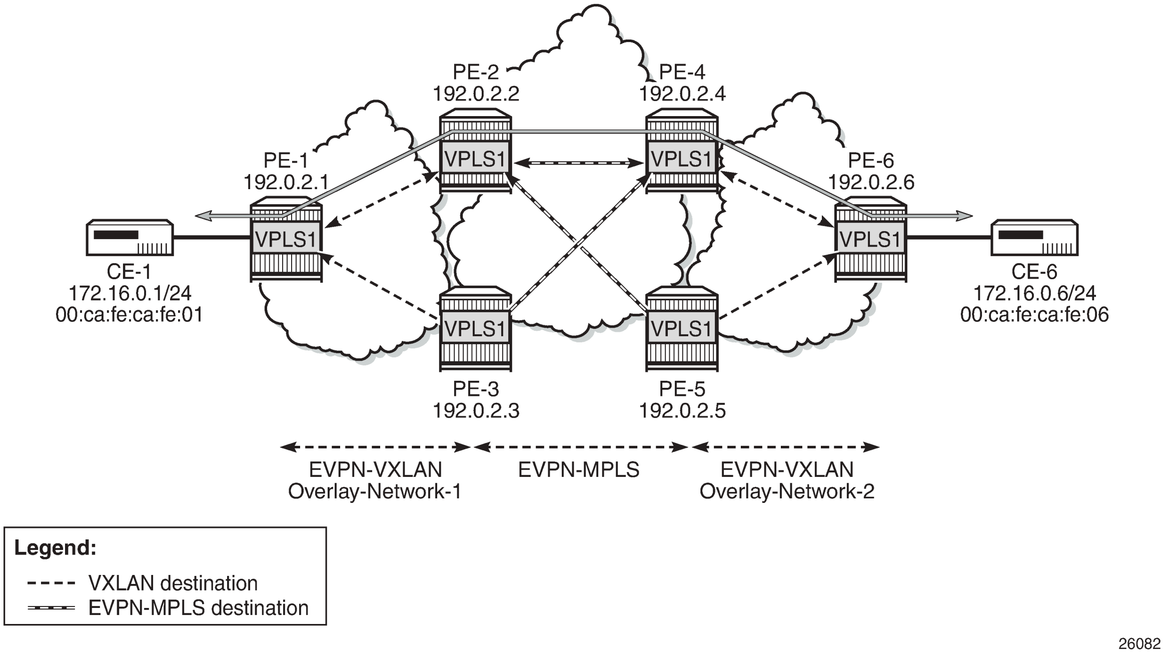
Multi-homed anycast configuration for dual BGP-instance VPLS services

Services with EVPN-MPLS and EVPN-VXLAN SHGs are specified in *draft-ietf-bess-dci-evpn-overlay* and the associated multi-homing solution is also described in the same draft. That multi-homing solution is based on an interconnect ES that allows all-active and single-active multi-homed EVPN networks as well as local attachment circuits in the DC GWs (SAP/SDP-bindings).

This chapter was initially written for SR OS Release 14.0.R5 and interconnect ESs were not supported in that release. Therefore, an anycast solution is used to provide redundancy. This anycast solution is based on the two PE DC GWs in the redundant pair being configured to advertised MAC/IP and IMET routes with the same route key, so that the remote PEs will only pick up one of the two anycast DC GWs when sending unicast or BUM traffic, and no loop or packet duplication is created.

Figure 148: EVPN destinations created on multi-homed anycast DC GWs is an example of how multi-homing can be achieved for dual BGP-instance VPLS services. The figure also shows the EVPN destinations created and their direction (see the arrows). For instance, only one EVPN multicast destination is created for PE-1, PE-2, or PE-4. Therefore, BUM traffic sent by CE-1 will be sent via PE-2, PE-4, and PE-6 only, and no duplication or loops occur.

Figure 148: EVPN destinations created on multi-homed anycast DC GWs



The following output shows the VPLS 1 configuration on PE-2 and PE-3 so that this anycast redundancy can be realized. The route distinguishers as well as the **incl-mcast-orig-ip** addresses must match between the two PEs in the redundant pair. VPLS 1 is configured on PE-2 as follows:

```
# on PE-2:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            vxlan instance 1 vni 1 create
            exit
            bgp
                route-distinguisher 64500:1
            exit
            bgp 2
                route-distinguisher 64500:2
            exit
            bgp-evpn
                incl-mcast-orig-ip 23.23.23.23
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
                mpls bgp 2
                    ingress-replication-bum-label
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
```

The VPLS 1 configuration on PE-3 is as follows:

```
# on PE-3:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            vxlan instance 1 vni 1 create
            exit
            bgp
                route-distinguisher 64500:1
            exit
            bgp 2
                route-distinguisher 64500:2
            exit
            bgp-evpn
                incl-mcast-orig-ip 23.23.23.23
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
                mpls bgp 2
                    ingress-replication-bum-label
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
```

The VPLS 1 configuration on PE-4 is as follows:

```
# on PE-4:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            vxlan instance 1 vni 1 create
            exit
            bgp
                route-distinguisher 64501:1
            exit
            bgp 2
                route-distinguisher 64501:2
            exit
            bgp-evpn
                incl-mcast-orig-ip 45.45.45.45
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
                mpls bgp 2
                    ingress-replication-bum-label
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
```

The VPLS 1 configuration on PE-5 is as follows:

```
# on PE-5:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            vxlan instance 1 vni 1 create
            exit
            bgp
                route-distinguisher 64501:1
            exit
            bgp 2
                route-distinguisher 64501:2
            exit
            bgp-evpn
                incl-mcast-orig-ip 45.45.45.45
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
                mpls bgp 2
                    ingress-replication-bum-label
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
```

Based on the preceding configuration example, the DC GWs behavior in this scenario is as follows:

- PE-2 and PE-3 both send IMET IR routes to the other PEs with the same route key but a different next-hop. The route key in IMET routes comprises [RD, Ethernet tag, originator-IP/length], which in this case will be [64500:1, 0, 23.23.23.23/32] for the EVPN-VXLAN IMET routes and [64500:2, 0, 23.23.23.23/32] for the EVPN-MPLS IMET routes.
- In the same way, PE-2 and PE-3 both send MAC/IP routes to the other PEs with the same route key but a different next-hop. The route key comprises [RD, Ethernet tag, MAC/MAC-length, IP/IP-length].

The configuration of the same **incl-mcast-orig-ip** address and RDs in both DC GWs enables the anycast solution due to the following:

- The configured originating IP (for example, 23.23.23.23 in PE-2 and PE-3) is not required to be a reachable IP address, which forces the remote PEs (or RRs if they exist) to select only one of the two DC GWs for BUM traffic (based on regular BGP selection). In this example, the remote PEs will select the PE-2 IMET route and create only one destination. The following output shows the IMET routes received by PE-1 (only the PE-2 route is used) and the created EVPN-VXLAN destination to PE-2. The same behavior could have been shown in the rest of the PEs.

```
*A:PE-1# show router bgp routes evpn incl-mcast
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Flag  Route Dist.         OrigAddr
      Tag                 NextHop
-------------------------------------------------------------------------------
u*>i  64500:1             23.23.23.23
      0                   192.0.2.2

*>i   64500:1             23.23.23.23
      0                   192.0.2.3

-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

```
*A:PE-1# show service id 1 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
Instance    VTEP Address                            Egress VNI  EvpnStatic Num
 Mcast       Oper State                              L2 PBR     SupBcasDom MACs
```

```
-------------------------------------------------------------------------------
1           192.0.2.2                               1           evpn       1
 BUM         Up                                      No          No
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

- Due to the same RD and originating IP configured on PE-2 and PE3 (similarly in PE-4 and PE-5), the DC GW redundant PEs will never establish an EVPN destination between each other. PE-2 only sets up EVPN multicast destinations to PE-1 and PE-4, as follows:

```
*A:PE-2# show service id 1 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
Instance    VTEP Address                            Egress VNI  EvpnStatic Num
 Mcast       Oper State                              L2 PBR     SupBcasDom MACs
-------------------------------------------------------------------------------
1           192.0.2.1                               1           evpn       1
 BUM         Up                                      No          No
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

```
*A:PE-2# show service id 1 evpn-mpls

===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address     Egr Label     Num. MACs   Mcast           Last Change
                Transport:Tnl                             Sup BCast Domain
-------------------------------------------------------------------------------
192.0.2.4       524282        0           bum             03/15/2021 16:38:34
                ldp:65538                                 No
192.0.2.4       524283        1           none            03/15/2021 16:39:13
                ldp:65538                                 No
-------------------------------------------------------------------------------
Number of entries : 2
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

- Likewise, when the two redundant PEs receive the same MAC/IP route, they will both re-advertise it with the same route key, forcing the remote PEs to pick up only one of the two (based on regular BGP selection) and create only one EVPN destination (if different from the multicast destination). In the following example, PE-6 advertised the CE-6 MAC address, that is, re-advertised by PE-4/PE-5 and then by PE-2/PE-3, but only one of the routes is selected at each hop. The following output shows that PE-1 selects the PE-2 MAC/IP route (see the "used" flag) and uses the existing EVPN destination to PE-2:

```
*A:PE-1# show router bgp routes evpn mac
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  64500:1             00:ca:fe:ca:fe:06 ESI-0
      0                   Seq:0             VNI 1
                          n/a
                          192.0.2.2

*>i   64500:1             00:ca:fe:ca:fe:06 ESI-0
      0                   Seq:0             VNI 1
                          n/a
                          192.0.2.3

-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

```
*A:PE-1# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:ca:fe:ca:fe:01 sap:1/2/1:1             L/30     03/15/21 16:39:04
1          00:ca:fe:ca:fe:06 vxlan-1:                Evpn     03/15/21 16:39:13
                             192.0.2.2:1
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

```
*A:PE-1# show service id 1 vxlan destinations

===============================================================================
Egress VTEP, VNI
```

```
===============================================================================
Instance    VTEP Address                            Egress VNI  EvpnStatic Num
 Mcast       Oper State                              L2 PBR     SupBcasDom MACs
-------------------------------------------------------------------------------
1           192.0.2.2                               1           evpn       1
 BUM         Up                                      No          No
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

- As shown in the preceding outputs, the EVPN destinations are always created to the IMET or MAC/IP route's BGP next-hops, which are still the system IP address of the routers (they could have also been a loopback address). The BGP next-hops need to be reachable in their respective network: DC or WAN.

The mandatory use of BGP policies in the multi-homed anycast solution

BGP policies must be configured in a multi-homed anycast solution, such as the one described in the previous section. Without policies, the following undesired behavior would happen:

- IMET routes with VXLAN encapsulation would be sent to the BGP peers in the MPLS network and IMET routes with MPLS encapsulation sent to BGP peers in the DC. The configured BGP policies will avoid that and make sure that the VXLAN routes are only sent to the DC and MPLS routes only to the WAN.
- MAC/IP routes received in the VXLAN BGP instance of a DC GW would be re-advertised to the redundant DC GW in the MPLS BGP instance and the redundant DC GW would re-advertise the same MAC again into the VXLAN instance, creating a control plane loop. The same thing would happen for MAC/IP routes received in an MPLS BGP instance. The configured BGP policies will prevent a DC GW from re-advertising MAC/IP routes received from the redundant DC GW.

While service-level BGP policies (**config>service>vpls>bgp>vsi-import/export**) may have been configured to prevent these loops and misbehavior, the use of BGP peer-level policies (**config>router>bgp>group>import/export**) is recommended due to the following reasons:

- Simplicity - BGP peer-level policies do not require any extra configuration at the service level, only at the BGP level.
- Scalability - BGP peer-level policies scale better than VSI-level policies, because the number of services where the VSI policies should be configured may be significant.

The following policies are configured in the example used in this chapter. No policies are needed in PE-1 and PE-6; only the DC GWs must be configured.

Following are the policies on PE-2 and PE-3:

```
# on PE-2, PE-3:
configure
    router Base
        policy-options
            begin
            community "mpls"
                members "bgp-tunnel-encap:MPLS"
            exit
            community "vxlan"
                members "bgp-tunnel-encap:VXLAN"
            exit
            community "SOO-DCGW-23"
                members "origin:64500:23"
            exit

/* "drop SOO-DCGW-23" will drop any EVPN route that is received from PE-3,
the other DC GW in the pair. */

            policy-statement "drop SOO-DCGW-23"
                entry 10
                    from
                        community "SOO-DCGW-23"
                        family evpn
                    exit
                    action drop
                    exit
                exit
            exit

/* "allow only mpls and add SOO" has a twofold objective: avoids sending EVPN-VXLAN
routes to the MPLS network and marks the advertised EVPN routes with a Site-Of-Origin extended
 community that identifies the DC GW pair. */

            policy-statement "allow only mpls and add SOO"
                entry 10
                    from
                        community "vxlan"
                        family evpn
                    exit
                    action drop
                    exit
                exit
                entry 20
                    from
                        family evpn
                    exit
                    action accept
                        community add "SOO-DCGW-23"
                    exit
                exit
            exit

/* In the same way, "allow only vxlan and add SOO" avoids sending EVPN-MPLS routes
to the VXLAN network and marks the EVPN routes with a Site-Of-Origin extended
community that identifies the DC GW pair. */

            policy-statement "allow only vxlan and add SOO"
                entry 10
                    from
                        community "mpls"
                        family evpn
                    exit
                    action drop
                    exit
                exit
                entry 20
                    from
                        family evpn
                    exit
                    action accept
                        community add "SOO-DCGW-23"
                    exit
                exit
            exit
            commit
        exit
```

The policies are properly applied at BGP group level.

```
# on PE-2:
configure
    router Base
        bgp
            family evpn
            vpn-apply-import
            vpn-apply-export
            rapid-withdrawal
            rapid-update evpn
            group "DC"
                type internal
                import "drop SOO-DCGW-23"
                export "allow only vxlan and add SOO"
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.3
                exit
            exit
            group "WAN"
                type internal
                import "drop SOO-DCGW-23"
                export "allow only mpls and add SOO"
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
            no shutdown
        exit
```

The same policies are configured and applied on PE-3 (including the addition and filtering of the same Site-Of-Origin because PE-3 is part of the same DC GW pair).

PE-4 and PE-5 use the same BGP peer policies, but using a Site Of Origin extended community identifying the PE-4/PE-5 pair instead of the PE-2/PE-3 pair:

```
# on PE-4, PE-5:
configure
    router
        policy-options
            begin
            community "mpls"
                members "bgp-tunnel-encap:MPLS"
            exit
            community "vxlan"
                members "bgp-tunnel-encap:VXLAN"
            exit
            community "SOO-DCGW-45"
                members "origin:64500:45"
            exit
---snip---
```

Dual BGP instance VPLS service caveats

When two BGP instances are enabled on the same VPLS service, the following considerations apply:

- SDP-bindings are not supported (therefore, no pw-template-binding is needed in the service). Any attempt to add an SDP-binding to a service with two BGP instances will be blocked by the CLI, as follows:

```
*A:PE-2>config>service>vpls# spoke-sdp 21:1 create
MINOR: SVCMGR #7888 Cannot be configured/enabled with EVPN - sdp-binds not allowed when both
 vxlan and evpn-mpls are enabled
```

- Services that are not supported: R-VPLS, M-VPLS, I-VPLS, B-VPLS, or E-Tree VPLS
  - A consequence of not supporting R-VPLS is that no routes type 5 (IP-Prefix routes) are supported on dual BGP-instance services.
- Proxy-ARP/ND is not supported.
- BGP multi-homing is not supported.
- Although the Assisted-Replication feature is supported on dual BGP-instance VPLS services, the Assisted-Replication configuration is only relevant to the VXLAN destinations. See section EVPN route handling in dual BGP-instance VPLS services for some considerations about how EVPN handles IMET AR routes.

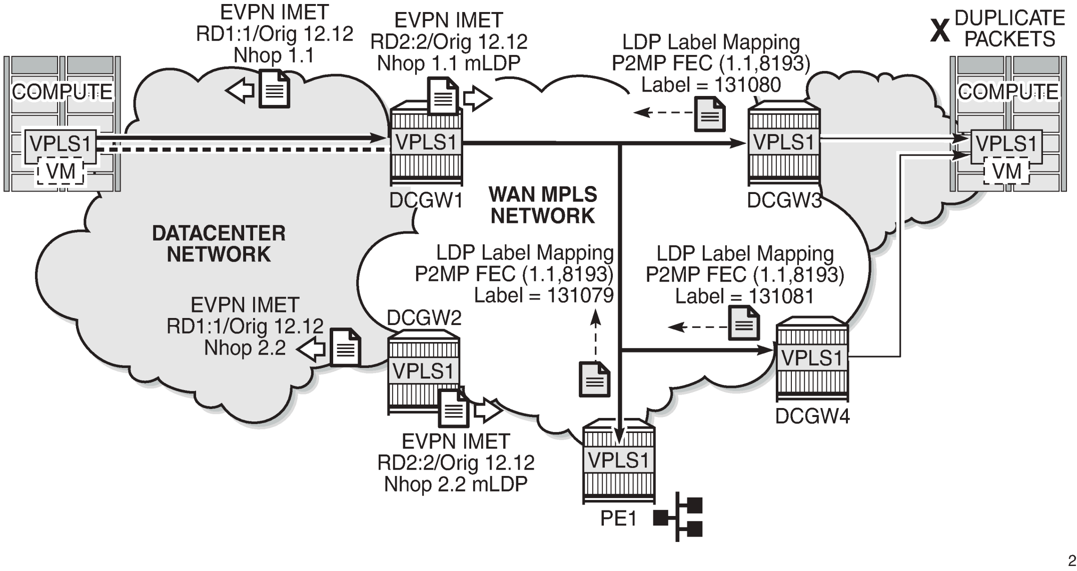
In addition to the preceding restrictions, some commands have a specific behavior when two BGP instances are configured:

- **config>service>vpls>bgp-evpn>[no] mac-advertisement** enables/disables the re-advertisement of MAC/IP routes in a BGP instance for MAC addresses that have been learned in the other BGP instance in the service.
- **config>service>vpls>bgp-evpn>[no] unknown-mac-route** enables/disables the advertisement of the unknown MAC route (MAC 00:..:00) on the BGP-EVPN VXLAN instance. The unknown MAC route is never sent to the BGP-EVPN MPLS instance.

The use of provider tunnels on multi-homed anycast solutions

The use of provider tunnels in dual BGP-instance VPLS services connecting multiple DCs is not recommended. Figure 149: Use of provider-tunnels between anycast DC GWs create packet duplication shows the case where the same BGP-EVPN service is configured in redundant anycast DC GWs and mLDP is used in the MPLS instance. In this case, packet duplication may occur if the configuration is not done carefully.

Figure 149: Use of provider-tunnels between anycast DC GWs create packet duplication



When mLDP is used along with multiple anycast multi-homing DC GWs to send BUM traffic to remote PEs, but no BUM traffic between DCs is needed, the same originating IP must be used on all the DC GWs; otherwise, packet duplication may happen. In the example in Figure 149: Use of provider-tunnels between anycast DC GWs create packet duplication, each pair of DC GWs, DCGW1/DCGW2 and DCGW3/DCGW4, is configured with a different originating IP (**config>service>vpls>bgp-evpn> incl-mcast-orig-ip**):

- DCGW3 and DCGW4 will receive the IMET route with the same route key from DCGW1 and DCGW2.
- DCGW3 and DCGW4 will select only one route, which will usually be the same; for example, the DCGW1 IMET route.
- Because of that, both DCGW3 and DCGW4 will join the mLDP tree with root in DCGW1, creating packet duplication when DCGW1 sends BUM traffic.
- Remote PE nodes with a single MPLS instance will join the mLDP tree without any issue.

To avoid the packet duplication shown by the example of Figure 149: Use of provider-tunnels between anycast DC GWs create packet duplication, the same originating IP may be configured in the four DCGWs, while the RD is still different per pair. By doing that:

- In the example of Figure 149: Use of provider-tunnels between anycast DC GWs create packet duplication, DCGW3 and DCGW4 will never join any mLDP tree sourced from DCGW1 or DCGW2. This will prevent any packet duplication because a router will ignore IMET routes received with its own originating IP, regardless of the RD.
- PE-1 (a remote EVPN-MPLS PE) will still join the mLDP trees from the two DCs.
- The preceding configuration allows the use of mLDP as long as no BUM traffic is required between the two DCs. If BUM traffic is required between DCs, IR must be used.

Troubleshooting and debugging

The following show and debug commands can be used in dual BGP-instance VPLS services:

- show router bgp routes evpn (and filters)
- show service evpn-mpls [<TEP ip-address>]
- show service vxlan [<TEP ip-address>]
- show service id bgp-evpn
- show service id evpn-mpls (and modifiers)
- show service id vxlan destinations
- debug router bgp update
- show log log-id "99"

See chapter EVPN for MPLS Tunnels and EVPN for VXLAN Tunnels (Layer 2) for a detailed description of these commands.

Also, in dual BGP-instance VPLS services, the **show service id** <*service-id*> **bgp** <*bgp-instance*> command may help see the BGP parameters of each individual BGP instance:

```
*A:PE-2# show service id 1 bgp ?
  - bgp [<bgp-instance>]

 <bgp-instance>       : [1..2]
```

```
*A:PE-2# show service id 1 bgp 1

===============================================================================
BGP Information
===============================================================================
Vsi-Import           : None
Vsi-Export           : None
Route Dist           : 64500:1
Oper Route Dist      : 64500:1
Oper RD Type         : configured
Rte-Target Import    : None                 Rte-Target Export: None
Oper RT Imp Origin   : derivedEvi           Oper RT Import   : 64500:1
Oper RT Exp Origin   : derivedEvi           Oper RT Export   : 64500:1
PW-Template Id       : None
-------------------------------------------------------------------------------
===============================================================================
```

```
*A:PE-2# show service id 1 bgp 2

===============================================================================
BGP Information
===============================================================================
Vsi-Import           : None
Vsi-Export           : None
Route Dist           : 64500:2
Oper Route Dist      : 64500:2
Oper RD Type         : configured
Rte-Target Import    : None                 Rte-Target Export: None
Oper RT Imp Origin   : derivedEvi           Oper RT Import   : 64500:1
Oper RT Exp Origin   : derivedEvi           Oper RT Export   : 64500:1
-------------------------------------------------------------------------------
```

```
===============================================================================
```

Conclusion

As service providers deploy EVPN-MPLS in the network for Ethernet local area network (E-LAN) and Ethernet point-to-point (E-Line) services, the use of EVPN-MPLS to interconnect data centers is becoming a popular option. Based on *draft-ietf-bess-dci-evpn-overlay*, SR OS supports the connectivity of Layer 2 EVPN-VXLAN services to an EVPN-MPLS network. To implement that EVPN-MPLS Data Center Interconnect (DCI) solution, VPLS services support dual BGP instances, where EVPN-VXLAN and EVPN-MPLS can coexist simultaneously in the same VPLS service. This chapter describes the configuration of such dual BGP-instance VPLS services and how to deploy them in a redundant anycast DC GW configuration.

EVPN-VXLAN VPWS

This chapter provides information about EVPN-VXLAN VPWS.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS Release 16.0.R7, but the CLI in the current edition is based on SR OS Release 21.5.R2.

Overview

Some service providers use VXLAN as a next-generation access technology between Multi-Service Access Node (MSAN) PE and core PE routers. VXLAN allows any IP router in the aggregation core and provides a simple alternative to MPLS. Static VXLAN bindings can be used when the MSAN PEs do not support any control plane. However, EVPN offers a control plane protocol for the VXLAN bindings for faster convergence and fault propagation. In this chapter, the focus is on EVPN-VPWS, which provides a lighter control plane compared to full-blown EVPN when point-to-point services need to be extended to the Data Center (DC).

EVPN-VXLAN VPWS is similar to EVPN-MPLS VPWS, including support of Equal Cost Multi-Path (ECMP), and EVPN All-Active (AA) and Single-Active (SA) Multi-Homing (MH). The configuration resembles the EVPN-MPLS Epipe configuration, as described in the EVPN for MPLS Tunnels in Epipe Services (EVPN-VPWS) chapter. As an example, the following configures EVPN-VXLAN Epipe 4 with SA MH.

```
# on PE-4:
configure
    service
        sdp 460 create
            description "GRE SDP for SA MH"
            far-end 192.0.2.6
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        system
            bgp-evpn
                ethernet-segment "ES45" create
                    esi 01:00:00:00:00:45:00:00:00:04
                    es-activation-timer 3
                    service-carving
```

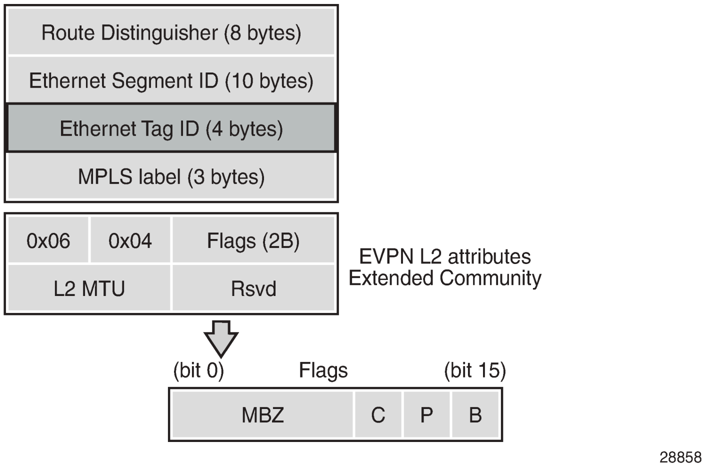
```
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 460
                    no shutdown
                exit
            exit
        exit
        epipe 4 name "Epipe-4" customer 1 create
            vxlan instance 1 vni 4 create
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-45 create
                    eth-tag 145
                exit
                remote-attachment-circuit AC-23 create
                    eth-tag 123
                exit
                evi 4
                vxlan bgp 1 vxlan-instance 1
                    ecmp 2
                    send-tunnel-encap     # default
                    no shutdown
                exit
            exit
            spoke-sdp 460:4 create
                no shutdown
            exit
            no shutdown
        exit
```

The SDP is a GRE SDP, because no MPLS is configured in the network. The VNI is 4, and the local Attachment Circuit (AC) name is "AC-45" with Ethernet tag 145, whereas the remote AC name is "AC-23" with Ethernet tag 123. An ES can contain up to four nodes. Each of these nodes will have the same local Ethernet tag.

On Epipe services, the BGP instance is 1 and the VXLAN instance is 1. ECMP is configured with a value of 2, so the traffic flows can be sprayed over two paths with equal cost (a value greater than 2 can be configured if aliasing to more than two nodes is needed). By default, **send-tunnel-encap** is enabled, which determines whether the RFC 5512 encapsulation extended community is sent with VXLAN value (if enabled) or not sent.

EVPN-VPWS uses BGP-EVPN route type 1 (autodiscovery (AD) per-EVI routes and AD per-ES routes) and route type 4 (Ethernet Segment (ES) routes); it does not use route types 2 (MAC/IP routes), 3 (Inclusive Multicast routes), or 5 (IP Prefix routes). Figure 150: BGP-EVPN AD per-EVI route shows the fields in a BGP-EVPN AD per-EVI route.

Figure 150: BGP-EVPN AD per-EVI route



The Route Distinguisher (RD) is encoded as specified in RFC 7432; in this example, the system IP address is followed by the service ID, such as 192.0.2.2:1 for Epipe 1 on PE-2. The MPLS label field is encoded as the VXLAN Network Identifier (VNI) and the Ethernet tag field defines the local Attachment Circuit (AC) ID. The ES ID (ESI) is the 10 bytes configured ESI for MH and equals zero for single-homed services.

The EVPN L2 attributes extended community has type 0x06 (EVPN) and subtype 0x04 (EVPN L2 attributes). The flags are defined as follows:

- Flag C (control word) is set if control word is configured in the service. For EVPN-MPLS VPWS, the control word can be configured in the **bgp-evpn>mpls** context, but for EVPN-VXLAN VPWS, the control word cannot be configured in the **bgp-evpn>vxlan** context, so flag C is always zero (C=0).
- Flag P (primary) is set in MH scenarios: all nodes in an AA MH ES send P=1, but in an SA MH ES, only the Designated Forwarder (DF) sends P=1, while the NDFs send P=0. In single-homed scenarios, all nodes send P=0.
- Flag B (backup) is set in SA MH scenarios: the NDF that will take the primary role after the original primary node has failed is the backup, so it sends B=1. All other NDFs have B=0. In AA MH scenarios, all nodes send B=0. Also, in single-homed scenarios, all nodes except for the backup DF send B=0.

If the received L2 MTU does not match the configured service MTU, the EVPN binding is not set up. However, if the received L2 MTU is zero, the MTU is ignored.

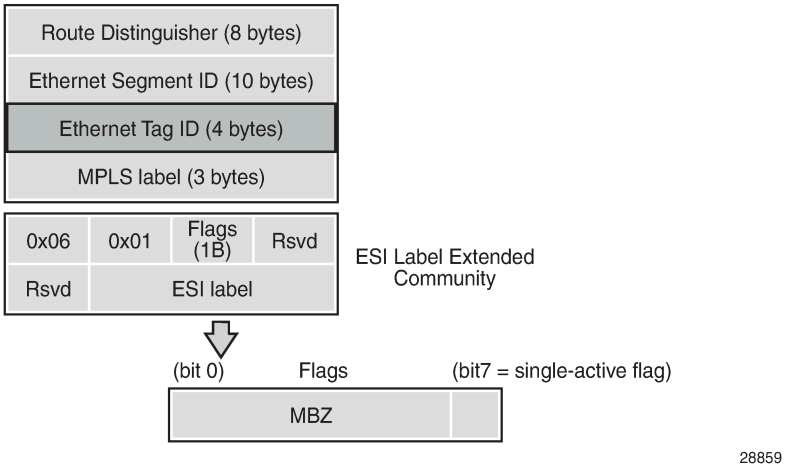
AD per-EVI routes are responsible for aliasing. The following BGP update shows an AD per-EVI route received from DF 192.0.2.4 (PE-4) in an SA MH ES with ESI 01:00:00:00:00:45:00:00:00:04, Ethernet tag 145 for the local AC on PE-4, and MPLS label 4 for Epipe 4. The primary flag is set: P=1.

```
50 2021/06/29 12:03:54.278 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 81
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.4
        Type: EVPN-AD Len: 25 RD: 192.0.2.4:4 ESI: 01:00:00:00:00:45:00:00:00:04,
                      tag: 145 Label: 4
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
```

```
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:4
        l2-attribute:MTU: 1514 C: 0 P: 1 B: 0
        bgp-tunnel-encap:VXLAN
"
```

As per RFC 8214, in an AD per-ES route, the Ethernet tag is MAX-ET (all bits are set), the MPLS label is zero, and the BGP extended community contains the single-active flag (1 for SA and 0 for AA) and ESI label. Figure 151: BGP-EVPN AD per-ES route shows the fields in a BGP-EVPN AD per-ES route.

Figure 151: BGP-EVPN AD per-ES route

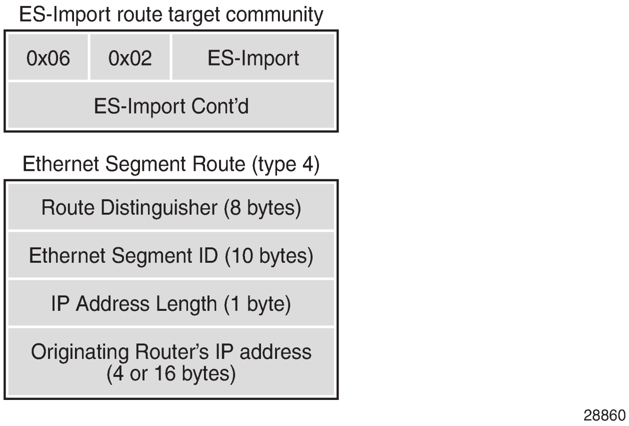

Route Distinguisher (8 bytes)
Ethernet Segment ID (10 bytes)
Ethernet Tag ID (4 bytes)
MPLS label (3 bytes)
0x06
0x01
Flags (1B)
Rsvd
Rsvd
ESI label
ESI Label Extended Community
(bit 0)
Flags
(bit7 = single-active flag)
MBZ
28859

The following AD per-ES route is received by PE-2 from PE-4, which is in an SA MH ES with ESI 01:00:00:00:00:45:00:00:00:04.

```
52 2021/06/29 12:03:18.185 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 73
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.4
        Type: EVPN-AD Len: 25 RD: 192.0.2.4:4 ESI: 01:00:00:00:00:45:00:00:00:04,
                      tag: MAX-ET Label: 0
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:4
        esi-label:524284/Single-Active
"
```

Figure 152: BGP-EVPN ES route shows a BGP-EVPN route type 4 (ES route) that is used for MH ES discovery and DF election.

Figure 152: BGP-EVPN ES route


ES-Import route target community
0x06
0x02
ES-Import
ES-Import Cont'd
Ethernet Segment Route (type 4)
Route Distinguisher (8 bytes)
Ethernet Segment ID (10 bytes)
IP Address Length (1 byte)
Originating Router's IP address
(4 or 16 bytes)
28860

The RD is taken from the system level RD; by default, the RD is derived as system-IP:0, such as 192.0.2.4:0 for PE-4. The ESI contains the 10-byte identifier as configured in the ES. The ES import route target community has type 0x06 (EVPN) and subtype 0x02 (ES import route target), and is derived from the MAC address portion of the ESI. This extended community is treated as a route target, such as: target:00:00:00:00:45:00. Only the PEs attached to the ES will import the ES route.

The following BGP update shows a BGP-EVPN ES route sent by PE-4. The RD is defined as 192.0.2.4:0, the ESI is 01:00:00:00:00:45:00:00:00:04, and the originating IP address is 192.0.2.4 for PE-4. The ES import route target is target:00:00:00:00:45:00.

```
45 2021/06/29 12:07:09.822 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 71
    Flag: 0x90 Type: 14 Len: 34 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.4
        Type: EVPN-ETH-SEG Len: 23 RD: 192.0.2.4:0 ESI: 01:00:00:00:00:45:00:00:00:04,
                          IP-Len: 4 Orig-IP-Addr: 192.0.2.4
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        df-election::DF-Type:Auto/DP:0/DF-Preference:0/AC:1
        target:00:00:00:00:45:00
"
```

By default, the system IP addresses are used for the VXLAN tunnel termination. However, it is possible to use non-system IPv4 or IPv6 termination for EVPN-VXLAN VPWS, both for single-homed and multi-homed services. In that case, Forwarding Path Extension (FPE) needs to be defined with VXLAN termination, as described in chapter Static VXLAN Termination in Epipe Services.

The following shows the configuration of the single-homed Epipe 2 using non-system IPv4 source VXLAN Tunnel Endpoint (VTEP) 10.0.3.1 on PE-3. Likewise, it is possible to use a non-system IPv6 source VTEP, such as **vxlan-src-vtep 2001::3:1**. Unlike the source VTEP, the egress VTEP cannot be configured when BGP-EVPN is enabled. The egress VTEP is dynamically learned via BGP instead.

```
# on PE-3:
```

```
configure
    service
        epipe 2 name "Epipe-2" customer 1 create
            vxlan-src-vtep 10.0.3.1
            vxlan instance 1 vni 2 create
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-3 create
                    eth-tag 103
                exit
                remote-attachment-circuit AC-5 create
                    eth-tag 105
                exit
                evi 2
                vxlan bgp 1 vxlan-instance 1
                    send-tunnel-encap     # default
                    no shutdown
                exit
            exit
            sap 1/1/1:2 create
                no shutdown
            exit
            no shutdown
```

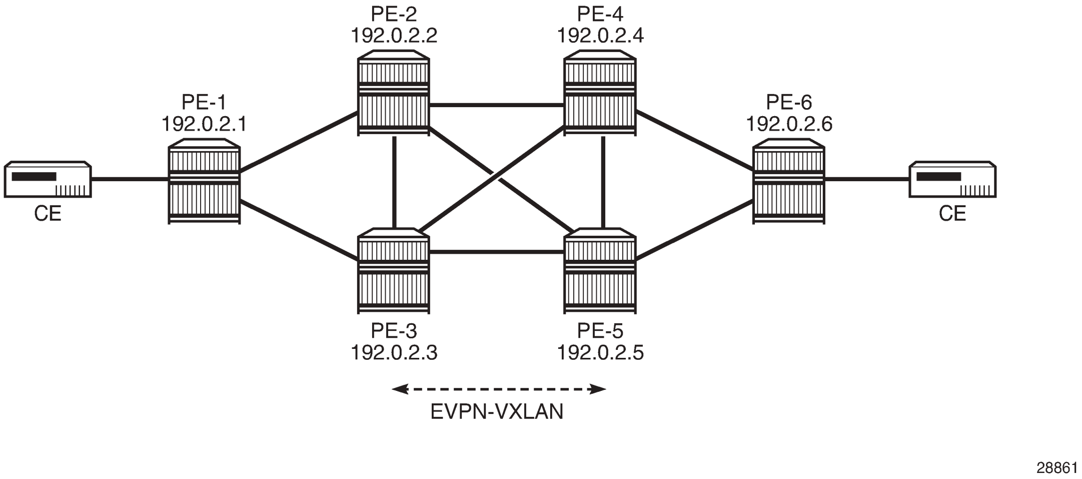
Configuration

The following use cases are included in the configuration section:

- Single-homed EVPN-VXLAN Epipe using IPv4 system addresses
- Single-homed EVPN-VXLAN Epipe using non-system IPv4 addresses
- Single-homed EVPN-VXLAN Epipe using non-system IPv6 addresses
- AA and SA multi-homed EVPN-VXLAN Epipe using IPv4 system addresses
- AA and SA multi-homed EVPN-VXLAN Epipe using non-system IPv4 addresses
- AA and SA multi-homed EVPN-VXLAN Epipe using non-system IPv6 addresses

Figure 153: Example topology shows the example topology with six PEs. EVPN-VXLAN Epipe services will be configured on the core PEs PE-2, PE-3, PE-4, and PE-5. On the access nodes PE-1 and PE-6, ordinary Epipe services will be configured, without EVPN-VXLAN. The CEs are emulated by VPRN services configured on PE-1 or PE-6.

Figure 153: Example topology



The initial configuration includes:

- Cards, MDAs, ports
- Router interfaces
- IS-IS on all router interfaces: level 2 between the core PEs and level 1 in the access networks

No MPLS protocol is configured.

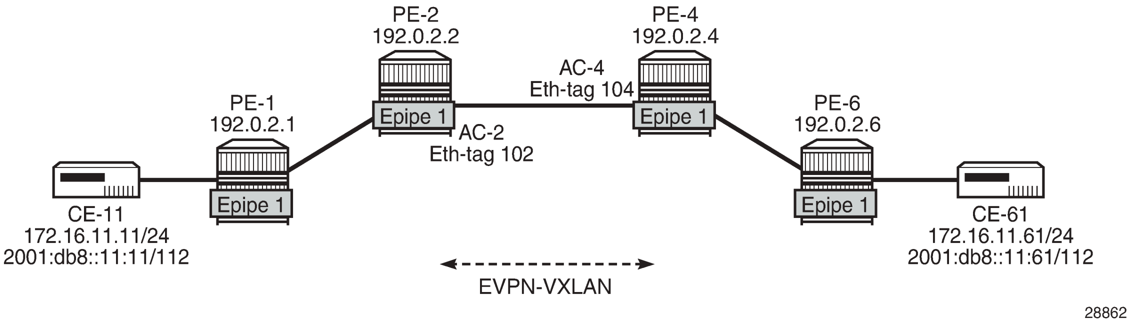
BGP is configured on the core PEs for the EVPN address family with RR PE-2. The BGP configuration on RR PE-2 is as follows:

```
# on PE-2:
configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            rapid-update evpn
            group "iBGP"
                family evpn
                type internal
                cluster 192.0.2.2
                split-horizon
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
            no shutdown
```

Single-homed EVPN-VXLAN Epipe using system IPv4 addresses

Figure 154: Single-homed EVPN-VXLAN Epipe 1 using system IP addresses shows the routers PE-1, PE-2, PE-4, and PE-6 configured with Epipe 1. VXLAN-EVPN is only configured on the core PEs PE-2 and PE-4.

Figure 154: Single-homed EVPN-VXLAN Epipe 1 using system IP addresses



Configuration of Epipe 1

On PE-1, Epipe 1 is configured without EVPN-VXLAN, as follows.

```
# on PE-1:
configure
    service
        epipe 1 name "Epipe-1" customer 1 create
            sap 1/1/1:1 create
                no shutdown
            exit
            sap 1/2/1:1 create
                no shutdown
            exit
            no shutdown
        exit
```

On PE-2, Epipe 1 is configured with EVPN-VXLAN. The local AC "AC-2" has Ethernet tag 102 and the remote AC is "AC-4" with Ethernet tag 104, as follows:

```
# on PE-2:
configure
    service
        epipe 1 name "Epipe-1" customer 1 create
            vxlan instance 1 vni 1 create
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-2 create
                    eth-tag 102
                exit
                remote-attachment-circuit AC-4 create
                    eth-tag 104
                exit
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            sap 1/1/2:1 create
                no shutdown
            exit
            no shutdown
        exit
```

The Epipe configuration on PE-4 is similar, but the local AC and remote AC are swapped, as follows. Instead of a SAP, a spoke-SDP is configured toward PE-6. The SDP itself is GRE-based.

```
# on PE-4:
configure
    service
        sdp 46 create
            description "GRE SDP for single-homing"
            far-end 192.0.2.6
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        epipe 1 name "Epipe-1" customer 1 create
            vxlan instance 1 vni 1 create
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-4 create
                    eth-tag 104
                exit
                remote-attachment-circuit AC-2 create
                    eth-tag 102
                exit
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    send-tunnel-encap     # default
                    no shutdown
                exit
            exit
            spoke-sdp 46:1 create
                no shutdown
            exit
            no shutdown
        exit
```

On PE-6, Epipe 1 is an ordinary Epipe with spoke-SDP 64:1 toward PE-4 and SAP 1/2/1:1 toward a CE, as follows:

```
# on PE-6:
configure
    service
        sdp 64 create
            description "GRE SDP for single-homing"
            far-end 192.0.2.4
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        epipe 1 name "Epipe-1" customer 1 create
            sap 1/2/1:1 create
                no shutdown
            exit
            spoke-sdp 64:1 create
                no shutdown
            exit
            no shutdown
        exit
```

Verification

VPRN 11 on PE-1 and PE-6 simulates the CEs CE-11 and CE-61. The connectivity between the CEs can be verified as follows:

```
*A:PE-1# ping router 11 172.16.11.61 rapid
PING 172.16.11.61 56 data bytes
!!!!!
---- 172.16.11.61 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 4.26ms, avg = 4.40ms, max = 4.54ms, stddev = 0.099ms
```

```
*A:PE-1# ping router 11 2001:db8::11:61 rapid
PING 2001:db8::11:61 56 data bytes
!!!!!
---- 2001:db8::11:61 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 3.80ms, avg = 3.96ms, max = 4.22ms, stddev = 0.153ms
```

On PE-2, the VXLAN destination for Epipe 1 is the system address of PE-4: 192.0.2.4, as follows. There are no VXLAN ES destinations for Epipe 1, because the service is single-homed.

```
*A:PE-2# show service id 1 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address                            Egress VNI          Oper      Vxlan
                                                            State     Type
-------------------------------------------------------------------------------
192.0.2.4                               1                   Up        evpn
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN VXLAN ES Dest
===============================================================================
I Eth Seg Id                    TEP Address       VNI       Last Changed
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

The following BGP-EVPN information for Epipe 1 on PE-2 includes the EVI and the AC names and Ethernet tags. For Epipes, the BGP instance ID and VXLAN instance ID always equal 1.

```
*A:PE-2# show service id 1 bgp-evpn

===============================================================================
BGP EVPN Table
===============================================================================
EVI                : 1                  Creation Origin    : manual
Local AC Name      : AC-2
Eth Tag            : 102
Endpoint           : (Not Specified)
Ingress Label      : 0
Remote AC Name     : AC-4
Eth Tag            : 104
Endpoint           : (Not Specified)

===============================================================================
BGP EVPN VXLAN Information
===============================================================================
Admin Status       : Enabled            Bgp Instance       : 1
Vxlan Instance     : 1
Max Ecmp Routes    : 1
Default Route Tag  : none
Send EVPN Encap    : Enabled
===============================================================================
```

PE-2 has received the following BGP-EVPN AD per-EVI route with RD 192.0.2.4:1 and Ethernet tag 104 from PE-4. Epipe 1 is single-homed, so ESI=0 and there is no primary or backup node (P=B=0). Also, no control word is used, so C=0.

```
5 2021/06/29 09:12:57.131 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 81
    Flag: 0x90 Type: 14 Len: 36 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.4
        Type: EVPN-AD Len: 25 RD: 192.0.2.4:1 ESI: ESI-0, tag: 104 Label: 1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:1
        l2-attribute:MTU: 1514 C: 0 P: 0 B: 0
        bgp-tunnel-encap:VXLAN
"
```

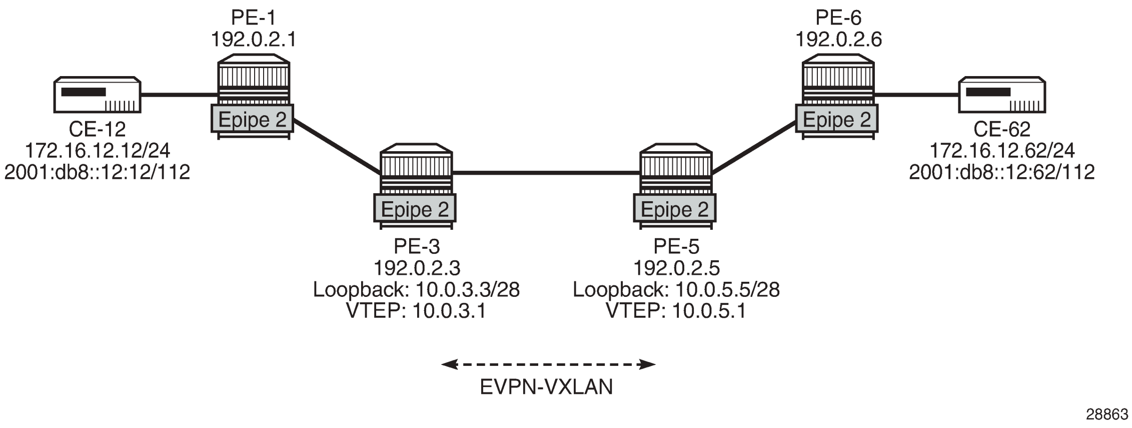
The following shows the received BGP-EVPN AD per-EVI routes with RD 192.0.2.4:1 on PE-2.

```
*A:PE-2# show router bgp routes evpn auto-disc rd 192.0.2.4:1
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Flag  Route Dist.         ESI                           NextHop
      Tag                                               Label
-------------------------------------------------------------------------------
u*>i  192.0.2.4:1         ESI-0                         192.0.2.4
      104                                               VNI 1

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

Single-homed EVPN-VXLAN Epipe using non-system IPv4 addresses

Figure 155: Single-homed EVPN-VXLAN Epipe 2 using non-system IP addresses shows the single-homed service Epipe 2 configured on PE-1, PE-3, PE-5, and PE-6. On PE-3, a loopback interface is created in the base router with IPv4 address 10.0.3.3/28. Epipe 2 uses VXLAN termination 10.0.3.1 from the same subnet.

Figure 155: Single-homed EVPN-VXLAN Epipe 2 using non-system IP addresses



Configuration of Epipe 2

On PE-1 and PE-6, the configuration of Epipe 2 is similar to the configuration of Epipe 1.

On PE-3, FPE needs to be configured using PXC, as described in chapter Static VXLAN Termination in Epipe Services. The following configuration is included without further explanation about FPE or PXC. The same configuration is required on PE-5.

```
# on PE-3:
configure
    port-xc
        pxc 1 create
            port 1/2/5
            no shutdown
        exit
    exit
    port 1/2/5
        ethernet
```

```
            dot1x
                tunneling
            exit
            mode hybrid
            encap-type dot1q
        exit
        no shutdown
    exit
    port pxc-1.a
        ethernet
        exit
        no shutdown
    exit
    port pxc-1.b
        ethernet
        exit
        no shutdown
    exit
    fwd-path-ext
        sdp-id-range from 10000 to 10127
        fpe 1 create
            path pxc 1
            vxlan-termination
        exit
```

On PE-3, the following loopback interface is created and IS-IS is enabled on it. The subnet must allow multiple IP addresses; one other IP address from the subnet will be defined as VXLAN tunnel termination. The IPv6 address is only required in the next use-case, but this configuration will not be repeated in that section.

```
# on PE-3:
configure
    router Base
        interface "lo1"
            address 10.0.3.3/28
            loopback
            ipv6
                address 2001::3:3/124
            exit
        exit
        isis 0
            interface "lo1"
                passive
            exit
        exit
```

Up to three VXLAN tunnel terminations can be defined per system. On PE-3, the following two VXLAN tunnel terminations are configured. For Epipe 2, only the first VXLAN tunnel termination is required; the second (IPv6) VXLAN tunnel termination is used in Epipe 3. The VXLAN tunnel termination is used as VXLAN source VTEP in Epipe 2. No egress VTEP can be defined when BGP-EVPN is configured in the service; egress VTEPs are configured in static VXLAN tunnels instead.

```
# on PE-3:
configure
    service
        system
            vxlan
                tunnel-termination 10.0.3.1 fpe 1 create
                tunnel-termination 2001::3:1 fpe 1 create
            exit
        exit
        epipe 2 name "Epipe-2" customer 1 create
            vxlan-src-vtep 10.0.3.1
            vxlan instance 1 vni 2 create
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-3 create
                    eth-tag 103
                exit
                remote-attachment-circuit AC-5 create
                    eth-tag 105
                exit
                evi 2
                vxlan bgp 1 vxlan-instance 1
                    send-tunnel-encap     # default
                    ecmp 2
                    no shutdown
                exit
            exit
            sap 1/1/1:2 create
                no shutdown
            exit
            no shutdown
        exit
```

The configuration on PE-5 is similar. The following is the service configuration on PE-5.

```
# on PE-5:
configure
    service
        sdp 56 create
            description "GRE SDP for single-homing"
            far-end 192.0.2.6
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        system
            vxlan
                tunnel-termination 10.0.5.1 fpe 1 create
                tunnel-termination 2001::5:1 fpe 1 create
            exit
        exit
        epipe 2 name "Epipe-2" customer 1 create
            vxlan-src-vtep 10.0.5.1
            vxlan instance 1 vni 2 create
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-5 create
                    eth-tag 105
                exit
                remote-attachment-circuit AC-3 create
                    eth-tag 103
                exit
                evi 2
                vxlan bgp 1 vxlan-instance 1
                    send-tunnel-encap     # default
                    ecmp 2
                    no shutdown
                exit
            exit
            spoke-sdp 56:2 create
                no shutdown
            exit
            no shutdown
        exit
```

It is possible to use a system IPv4 address as a VXLAN tunnel termination on one of the nodes and a non-system IPv4 address on another, but that is not configured here.

Verification

The connectivity between the CEs that are emulated by VPRN 12 can be verified as follows:

```
*A:PE-1# ping router 12 172.16.12.62 rapid
PING 172.16.12.62 56 data bytes
!!!!!
---- 172.16.12.62 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 4.27ms, avg = 4.77ms, max = 5.72ms, stddev = 0.509ms
```

```
*A:PE-1# ping router 12 2001:db8::12:62 rapid
PING 2001:db8::12:62 56 data bytes
!!!!!
---- 2001:db8::12:62 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 4.58ms, avg = 4.87ms, max = 5.59ms, stddev = 0.367ms
```

On PE-3, the VXLAN destination for Epipe 2 is the non-system address 10.0.5.1 on PE-5, as follows:

```
*A:PE-3# show service id 2 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address                            Egress VNI          Oper      Vxlan
                                                            State     Type
-------------------------------------------------------------------------------
10.0.5.1                                2                   Up        evpn
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN VXLAN ES Dest
===============================================================================
I Eth Seg Id                    TEP Address       VNI       Last Changed
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

The following BGP-EVPN information for Epipe 2 on PE-3 includes the EVI, AC names, and Ethernet tags.

```
*A:PE-3# show service id 2 bgp-evpn

===============================================================================
BGP EVPN Table
===============================================================================
EVI                 : 2                 Creation Origin    : manual
Local AC Name       : AC-3
Eth Tag             : 103
Endpoint            : (Not Specified)
Ingress Label       : 0
Remote AC Name      : AC-5
Eth Tag             : 105
Endpoint            : (Not Specified)

===============================================================================
BGP EVPN VXLAN Information
===============================================================================
Admin Status        : Enabled           Bgp Instance       : 1
Vxlan Instance      : 1
Max Ecmp Routes     : 1
Default Route Tag   : none
Send EVPN Encap     : Enabled
===============================================================================
```

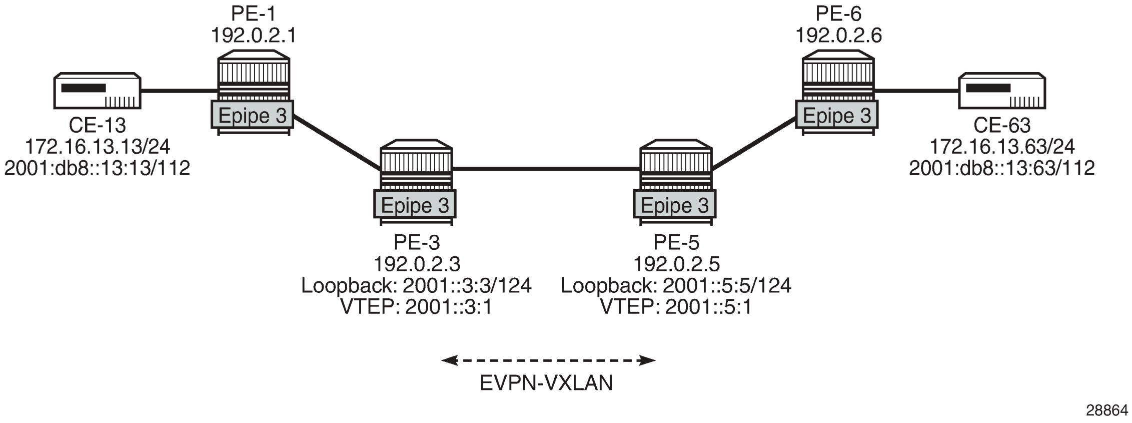
PE-3 received the following BGP-EVPN AD per-EVI route with RD 192.0.2.5:2 from PE-5. The Ethernet tag is 105 and the next-hop is the non-system address 10.0.5.1. ESI=0 for single-homed services.

```
*A:PE-3# show router bgp routes evpn auto-disc rd 192.0.2.5:2
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Flag  Route Dist.         ESI                            NextHop
      Tag                                                Label
-------------------------------------------------------------------------------
u*>i  192.0.2.5:2         ESI-0                          10.0.5.1
      105                                                VNI 2

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

Single-homed EVPN-VXLAN Epipe using non-system IPv6 addresses

Figure 156: Single-homed EVPN-VXLAN Epipe 3 using non-system IPv6 addresses shows the example topology for single-homed EVPN-VXLAN Epipe 3 using non-system IPv6 addresses for VXLAN tunnel termination.

Figure 156: Single-homed EVPN-VXLAN Epipe 3 using non-system IPv6 addresses



Configuration of Epipe 3

The following single-homed Epipe 3 using non-system IPv6 addresses for the VXLAN tunnel terminations is configured on PE-3.

```
# on PE-3:
configure
    service
        system
            vxlan
                tunnel-termination 10.0.3.1 fpe 1 create
                tunnel-termination 2001::3:1 fpe 1 create
            exit
        exit
        epipe 3 name "Epipe-3" customer 1 create
            vxlan-src-vtep 2001::3:1
            vxlan instance 1 vni 3 create
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-3_v6 create
                    eth-tag 163
                exit
                remote-attachment-circuit AC-5_v6 create
                    eth-tag 165
                exit
                evi 3
                vxlan bgp 1 vxlan-instance 1
                    send-tunnel-encap     # default
                    ecmp 2
                    no shutdown
                exit
            exit
            sap 1/1/1:3 create
            exit
            no shutdown
        exit
```

The service configuration on PE-5 is similar, as follows:

```
# on PE-5:
configure
    service
        system
            vxlan
                tunnel-termination 10.0.5.1 fpe 1 create
                tunnel-termination 2001::5:1 fpe 1 create
            exit
        exit
        epipe 3 name "Epipe-3" customer 1 create
            vxlan-src-vtep 2001::5:1
            vxlan instance 1 vni 3 create
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-5_v6 create
                    eth-tag 165
                exit
                remote-attachment-circuit AC-3_v6 create
                    eth-tag 163
                exit
                evi 3
                vxlan bgp 1 vxlan-instance 1
                    send-tunnel-encap     # default
                    ecmp 2
                    no shutdown
                exit
            exit
            spoke-sdp 56:3 create
                no shutdown
            exit
            no shutdown
        exit
```

Verification

The connectivity between the CEs that are emulated by VPRN 13 is verified as follows:

```
*A:PE-1# ping router 13 172.16.13.63 rapid
PING 172.16.13.63 56 data bytes
!!!!!
---- 172.16.13.63 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 4.00ms, avg = 4.31ms, max = 4.86ms, stddev = 0.318ms
```

```
*A:PE-1# ping router 13 2001:db8::13:63 rapid
PING 2001:db8::13:63 56 data bytes
!!!!!
---- 2001:db8::13:63 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 4.13ms, avg = 4.36ms, max = 4.50ms, stddev = 0.130ms
```

On PE-3, the VXLAN destination for Epipe 3 is the non-system IPv6 address 2001::5:1 on PE-5, as follows:

```
*A:PE-3# show service id 3 vxlan destinations
```

```
===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address                            Egress VNI          Oper      Vxlan
                                                            State     Type
-------------------------------------------------------------------------------
2001::5:1                               3                   Up        evpn
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN VXLAN ES Dest
===============================================================================
I Eth Seg Id                    TEP Address       VNI       Last Changed
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

The following BGP-EVPN information for Epipe 3 on PE-3 includes the EVI and the AC names and Ethernet tags.

```
*A:PE-3# show service id 3 bgp-evpn

===============================================================================
BGP EVPN Table
===============================================================================
EVI                : 3                  Creation Origin    : manual
Local AC Name      : AC-3_v6
Eth Tag            : 163
Endpoint           : (Not Specified)
Ingress Label      : 0
Remote AC Name     : AC-5_v6
Eth Tag            : 165
Endpoint           : (Not Specified)

===============================================================================
BGP EVPN VXLAN Information
===============================================================================
Admin Status       : Enabled            Bgp Instance       : 1
Vxlan Instance     : 1
Max Ecmp Routes    : 1
Default Route Tag  : none
Send EVPN Encap    : Enabled
===============================================================================
```

PE-3 received the following BGP-EVPN AD per-EVI route with RD 192.0.2.5:3 and next-hop 2001::5:1.

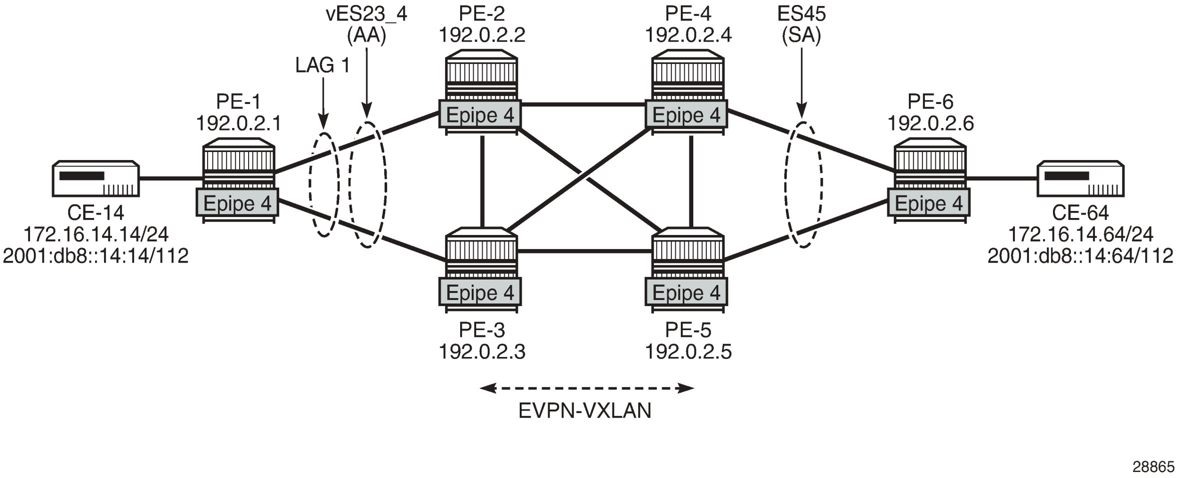
```
*A:PE-3# show router bgp routes evpn auto-disc rd 192.0.2.5:3
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Flag  Route Dist.         ESI                           NextHop
```

```
      Tag                                                 Label
-------------------------------------------------------------------------------
u*>i  192.0.2.5:3        ESI-0                            2001::5:1
      165                                                 VNI 3
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

AA and SA multi-homed EVPN-VXLAN Epipe using system IPv4 addresses

Figure 157: EVPN-VXLAN Epipe 4 with AA MH and SA MH using system IPv4 addresses shows the example topology for EVPN-VXLAN Epipe 4 with AA MH ES "vES23_4" between PE-2 and PE-3 and SA MH ES "ES45" between PE-4 and PE-5.

Figure 157: EVPN-VXLAN Epipe 4 with AA MH and SA MH using system IPv4 addresses



Configuration of Epipe 4

On PE-1, Epipe 4 is configured as follows:

```
# on PE-1:
configure
    service
        epipe 4 name "Epipe-4" customer 1 create
            sap 1/2/1:4 create
                no shutdown
            exit
            sap lag-1:4 create
                no shutdown
            exit
            no shutdown
        exit
```

On PE-2 and PE-3, the AA MH ES "vES23_4" is configured as a virtual ES for LAG 1 and dot1q-tag 4, so it only affects Epipe 4.

```
# on PE-2:
configure
    service
        system
            bgp-evpn
                ethernet-segment "vES23_4" virtual create
                    esi 01:00:00:00:00:23:00:00:00:04
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 1
                    dot1q
                        q-tag-range 4
                    exit
                    no shutdown
                exit
            exit
        exit
```

On PE-2 and PE-3, Epipe 4 is configured as follows. The system IPv4 address is used as VXLAN termination, the local AC Ethernet tag is 123, and the remote AC Ethernet tag is 145.

```
# on PE-2:
configure
    service
        epipe 4 name "Epipe-4" customer 1 create
            vxlan instance 1 vni 4 create
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-23 create
                    eth-tag 123
                exit
                remote-attachment-circuit AC-45 create
                    eth-tag 145
                exit
                evi 4
                vxlan bgp 1 vxlan-instance 1
                    send-tunnel-encap     # default
                    ecmp 2
                    no shutdown
                exit
            exit
            sap lag-1:4 create
                no shutdown
            exit
            no shutdown
        exit
```

On PE-4 and PE-5, the SA MH ES "ES45" is configured with a GRE SDP toward PE-6: SDP 460 on PE-4 and SDP 560 on PE-6. The following is the configuration of "ES45" on PE-4:

```
# on PE-4:
configure
    service
```

```
        sdp 460 create
            description "GRE SDP for SA MH"
            far-end 192.0.2.6
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        system
            bgp-evpn
                ethernet-segment "ES45"  create
                    esi 01:00:00:00:00:45:00:00:00:04
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 460
                    no shutdown
                exit
            exit
        exit
```

On PE-4, Epipe 4 is configured as follows. The configuration on PE-5 is similar, but with spoke-SDP 560:4 instead.

```
# on PE-4:
configure
    service
        epipe 4 name "Epipe-4" customer 1 create
            vxlan instance 1 vni 4 create
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-45 create
                    eth-tag 145
                exit
                remote-attachment-circuit AC-23 create
                    eth-tag 123
                exit
                evi 4
                vxlan bgp 1 vxlan-instance 1
                    send-tunnel-encap     # default
                    ecmp 2
                    no shutdown
                exit
            exit
            spoke-sdp 460:4 create
                no shutdown
            exit
            no shutdown
        exit
```

On PE-6, Epipe 4 is configured as follows:

```
# on PE-6:
configure
    service
        epipe 4 name "Epipe-4" customer 1 create
            endpoint "EP" create
            exit
            sap 1/2/1:4 create
```

```
                no shutdown
            exit
            spoke-sdp 640:4 endpoint "EP" create
                no shutdown
            exit
            spoke-sdp 650:4 endpoint "EP" create
                no shutdown
            exit
            no shutdown
        exit
```

Verification

The connectivity between the CEs emulated by VPRN 14 can be verified as follows:

```
*A:PE-1# ping router 14 172.16.14.64 rapid
PING 172.16.14.64 56 data bytes
!!!!!
---- 172.16.14.64 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 3.84ms, avg = 5.12ms, max = 9.41ms, stddev = 2.15ms
```

```
*A:PE-1# ping router 14 2001:db8::14:64 rapid
PING 2001:db8::14:64 56 data bytes
!!!!!
---- 2001:db8::14:64 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 3.80ms, avg = 5.19ms, max = 9.89ms, stddev = 2.36ms
```

The following BGP-EVPN information for Epipe 4 includes the EVI and the AC names and Ethernet tags:

```
*A:PE-2# show service id 4 bgp-evpn

===============================================================================
BGP EVPN Table
===============================================================================
EVI                : 4                  Creation Origin    : manual
Local AC Name      : AC-23
Eth Tag            : 123
Endpoint           : (Not Specified)
Ingress Label      : 0
Remote AC Name     : AC-45
Eth Tag            : 145
Endpoint           : (Not Specified)

===============================================================================
BGP EVPN VXLAN Information
===============================================================================
Admin Status       : Enabled            Bgp Instance       : 1
Vxlan Instance     : 1
Max Ecmp Routes    : 2
Default Route Tag  : none
Send EVPN Encap    : Enabled
===============================================================================
```

PE-4 received the following BGP-EVPN ES route with ESI 01:00:00:00:00:45:00:00:00:04 from PE-5:

```
*A:PE-4# show router bgp routes evpn eth-seg
===============================================================================
 BGP Router ID:192.0.2.4        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Eth-Seg Routes
===============================================================================
Flag  Route Dist.         ESI                           NextHop
      OrigAddr
-------------------------------------------------------------------------------
u*>i  192.0.2.5:0         01:00:00:00:00:45:00:00:00:04 192.0.2.5
      192.0.2.5

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

Furthermore, PE-4 received the following AD per-EVI (with Ethernet tag 123 or 145) and AD per-ES (MAX-ET) routes for Epipe 4 from its three BGP peers. The ESI is non-zero for multi-homed services.

```
*A:PE-4# show router bgp routes evpn auto-disc
===============================================================================
 BGP Router ID:192.0.2.4        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Auto-Disc Routes
===============================================================================
Flag  Route Dist.         ESI                           NextHop
      Tag                                               Label
-------------------------------------------------------------------------------
---snip---

u*>i  192.0.2.2:4         01:00:00:00:00:23:00:00:00:04 192.0.2.2
      123                                               VNI 4

u*>i  192.0.2.2:4         01:00:00:00:00:23:00:00:00:04 192.0.2.2
      MAX-ET                                            LABEL 0

u*>i  192.0.2.3:4         01:00:00:00:00:23:00:00:00:04 192.0.2.3
      123                                               VNI 4

u*>i  192.0.2.3:4         01:00:00:00:00:23:00:00:00:04 192.0.2.3
      MAX-ET                                            LABEL 0

u*>i  192.0.2.5:4         01:00:00:00:00:45:00:00:00:04 192.0.2.5
      145                                               VNI 4

u*>i  192.0.2.5:4         01:00:00:00:00:45:00:00:00:04 192.0.2.5
      MAX-ET                                            LABEL 0

-------------------------------------------------------------------------------
```

In AA MH ESs, the DF for VPLS services is the forwarder for Broadcast, Unknown unicast, and Multicast (BUM) traffic. In Epipes, however, all traffic is treated as unicast. The following tools commands on PE-2 and PE-3 show that DF is not applicable for AA MH ES "vES23_4.

```
*A:PE-2# tools dump service system bgp-evpn ethernet-segment "vES23_4" evi 4 df

[06/29/2021 09:59:05] All Active VPWS - DF N/A
```

```
*A:PE-3# tools dump service system bgp-evpn ethernet-segment "vES23_4" evi 4 df

[06/29/2021 09:59:03] All Active VPWS - DF N/A
```

The following command on PE-2 shows no DF candidates for ES "vES23_4", even though PE-2 (as well as PE-3) is considered as DF (DF=yes):

```
*A:PE-2# show service system bgp-evpn ethernet-segment name "vES23_4" evi 4

===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
4             4             0                   yes 06/29/2021 09:38:25
===============================================================================

===============================================================================
DF Candidates                           Time Added
-------------------------------------------------------------------------------
No entries found
===============================================================================
```

In the SA MH ES "ES45", PE-4 is DF out of a list of two candidates, as follows:

```
*A:PE-4# show service system bgp-evpn ethernet-segment name "ES45" evi 4

===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
4             4             0                   yes 06/29/2021 09:54:31
===============================================================================

===============================================================================
DF Candidates                           Time Added
-------------------------------------------------------------------------------
192.0.2.4                               06/29/2021 09:54:44
192.0.2.5                               06/29/2021 09:54:44
-------------------------------------------------------------------------------
Number of entries: 2
===============================================================================
```

On NDF PE-5, the spoke-SDP is operationally down with flag StandbyForMHProtocol, as follows:

```
*A:PE-5# show service id 4 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
```

```
-------------------------------------------------------------------------------
560:4             Spok     192.0.2.6       Up      Down      524282    524281
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

```
*A:PE-5# show service id 4 sdp detail | match "Flags"
Flags              : StandbyForMHProtocol
```

The following command on PE-2 shows that the VXLAN destination for Epipe 4 is the ES "ES45" with ESI 01:00:00:00:00:45:00:00:00:04 and TEP address 192.0.2.4, which is the system IP address of the DF.

```
*A:PE-2# show service id 4 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address                            Egress VNI          Oper      Vxlan
                                                            State     Type
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================

===============================================================================
BGP EVPN VXLAN ES Dest
===============================================================================
I Eth Seg Id                    TEP Address       VNI       Last Changed
-------------------------------------------------------------------------------
1 01:00:00:00:00:45:00:00:00:04 192.0.2.4         4         06/29/2021 09:54:47
-------------------------------------------------------------------------------
===============================================================================
```

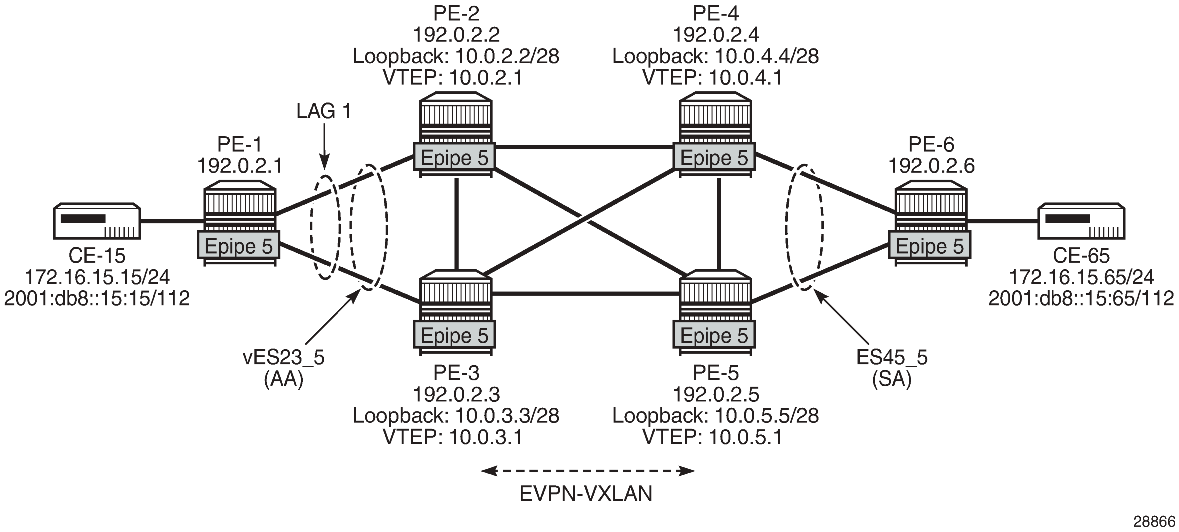
On PE-2, the following command shows that BGP-EVPN AD per-EVI routes with Ethernet tag 145 from PE-4 (RD 192.0.2.4:4) are sent with primary flag P=1 and AD per-EVI routes with Ethernet tag 145 from PE-5 (RD 192.0.2.5:4) are sent with primary flag P=0 and backup flag B=1.

```
*A:PE-3# show router bgp routes evpn auto-disc tag 145 detail
                                        | match expression "C:|Route Dist"
Community      : target:64500:4 l2-attribute:MTU: 1514 C: 0 P: 1 B: 0
Route Dist.    : 192.0.2.4:4
---snip---
Community      : target:64500:4 l2-attribute:MTU: 1514 C: 0 P: 0 B: 1
Route Dist.    : 192.0.2.5:4
---snip---
```

AA and SA multi-homed EVPN-VXLAN Epipe using non-system IPv4 addresses

Figure 158: EVPN-VXLAN Epipe 5 with AA MH and SA MH using non-system IPv4 addresses shows the example topology for EVPN-VXLAN Epipe 5 with AA MH ES "vES23_5" between PE-2 and PE-3 and SA MH ES "ES45_5" between PE-4 and PE-5.

Figure 158: EVPN-VXLAN Epipe 5 with AA MH and SA MH using non-system IPv4 addresses



The configuration of Epipe 5 on PE-1 is similar to the configuration of Epipe 4 on PE-1, so it is not shown here. The same applies for Epipe 5 on PE-6.

On PE-2, VTEP 10.0.2.1 is used instead of the system IP address. The ES must include two additional parameters for the DF selection: **es-orig-ip** and **route-next-hop**, which are both equal to the VTEP. Without these parameters, the DF selection will not work. The **es-orig-ip** command modifies the originator IP address of the ES route and the **route-next-hop** modifies the next-hop of the AD per-ES routes for the ES. The service configuration on PE-2 is as follows:

```
# on PE-2:
configure
    service
        system
            vxlan
                tunnel-termination 10.0.2.1 fpe 1 create
                tunnel-termination 2001::2:1 fpe 1 create
            exit
            bgp-evpn
                ethernet-segment "vES23_5" virtual create
                    esi 01:00:00:00:00:23:00:00:00:05
                    es-orig-ip 10.0.2.1
                    route-next-hop 10.0.2.1
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 1
                    dot1q
                        q-tag-range 5
                    exit
                    no shutdown
                exit
            exit
        exit
        epipe 5 name "Epipe-5" customer 1 create
            vxlan-src-vtep 10.0.2.1
            vxlan instance 1 vni 5 create
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-23_2 create
                    eth-tag 223
                exit
                remote-attachment-circuit AC-45_2 create
                    eth-tag 245
                exit
                evi 5
                vxlan bgp 1 vxlan-instance 1
                    ecmp 2
                    no shutdown
                exit
            exit
            sap lag-1:5 create
                no shutdown
            exit
            no shutdown
        exit
```

The service configuration on PE-3 is similar.

On PE-4, the service configuration is as follows:

```
# on PE-4:
configure
    service
        sdp 465 create
            far-end 192.0.2.6
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        system
            vxlan
                tunnel-termination 10.0.4.1 fpe 1 create
            exit
            bgp-evpn
                ethernet-segment "ES45_5" create
                    esi 01:00:00:00:00:45:00:00:00:05
                    es-orig-ip 10.0.4.1
                    route-next-hop 10.0.4.1
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 465
                    no shutdown
                exit
            exit
        exit
        epipe 5 name "Epipe-5" customer 1 create
            vxlan-src-vtep 10.0.4.1
            vxlan instance 1 vni 5 create
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-45_2 create
                    eth-tag 245
                exit
                remote-attachment-circuit AC-23_2 create
                    eth-tag 223
                exit
                evi 5
                vxlan bgp 1 vxlan-instance 1
                    ecmp 2
                    no shutdown
                exit
            exit
            spoke-sdp 465:5 create
                no shutdown
            exit
            no shutdown
        exit
```

In the AA MH ES, both PE-2 and PE-3 are DF. PE-4 receives BGP-EVPN autodiscovery routes with Ethernet tag 223 from PE-2 and PE-3 with the primary flag set to 1, as follows:

```
*A:PE-4# show router bgp routes evpn auto-disc tag 223 detail
                                        | match expression "C:|Route Dist"
Community      : target:64500:5 l2-attribute:MTU: 1514 C: 0 P: 1 B: 0
Route Dist.    : 192.0.2.2:5
Community      : target:64500:5 l2-attribute:MTU: 1514 C: 0 P: 1 B: 0
Route Dist.    : 192.0.2.2:5
Community      : target:64500:5 l2-attribute:MTU: 1514 C: 0 P: 1 B: 0
Route Dist.    : 192.0.2.3:5
Community      : target:64500:5 l2-attribute:MTU: 1514 C: 0 P: 1 B: 0
Route Dist.    : 192.0.2.3:5
```

The VXLAN destinations for Epipe 5 on PE-4 are the non-system TEP addresses 10.0.2.1 and 10.0.3.1 in ES "vES23_5" with ESI 01:00:00:00:00:23:00:00:00:05, as follows:

```
*A:PE-4# show service id 5 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address                            Egress VNI          Oper      Vxlan
                                                            State     Type
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================

===============================================================================
BGP EVPN VXLAN ES Dest
===============================================================================
I Eth Seg Id                    TEP Address       VNI       Last Changed
-------------------------------------------------------------------------------
1 01:00:00:00:00:23:00:00:00:05 10.0.2.1          5         06/29/2021 10:06:33
1 01:00:00:00:00:23:00:00:00:05 10.0.3.1          5         06/29/2021 10:06:33
-------------------------------------------------------------------------------
===============================================================================
```

In the SA MH ES, PE-5 is DF and PE-4 is NDF. PE-2 receives BGP-EVPN autodiscovery routes with Ethernet tag 245 from PE-4 with backup flag 1 and from PE-5 with primary flag 1, as follows:

```
*A:PE-2# show router bgp routes evpn auto-disc tag 245 detail
                                         | match expression "C:|Route Dist"
Community      : target:64500:5 l2-attribute:MTU: 1514 C: 0 P: 0 B: 1
Route Dist.    : 192.0.2.4:5
Community      : target:64500:5 l2-attribute:MTU: 1514 C: 0 P: 0 B: 1
```

```
Route Dist.    : 192.0.2.4:5
Community      : target:64500:5 l2-attribute:MTU: 1514 C: 0 P: 1 B: 0
Route Dist.    : 192.0.2.5:5
Community      : target:64500:5 l2-attribute:MTU: 1514 C: 0 P: 1 B: 0
Route Dist.    : 192.0.2.5:5
```

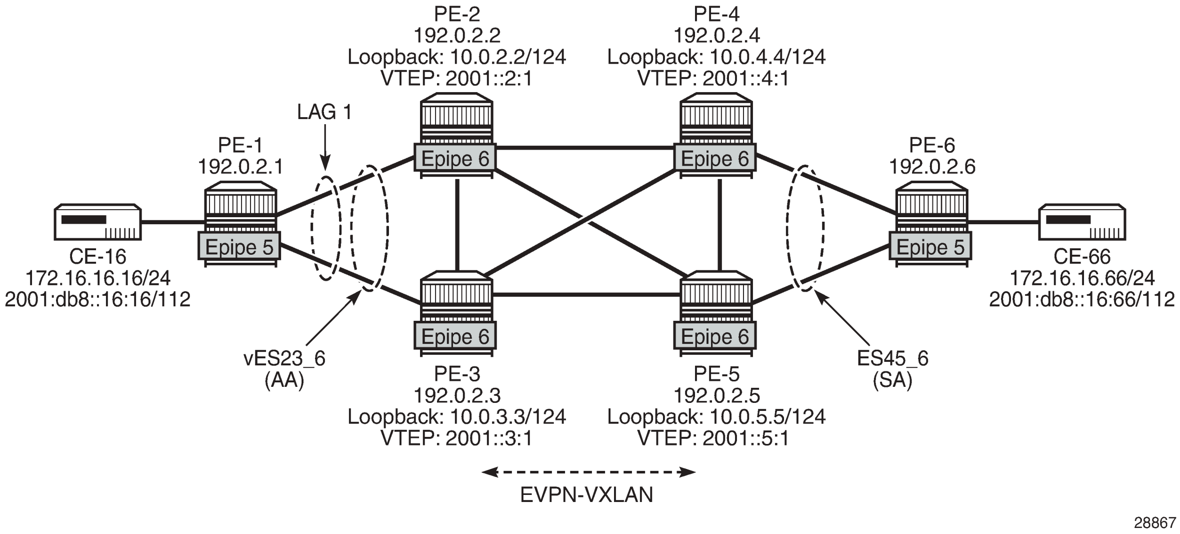
The VXLAN destination for Epipe 5 on PE-2 is the non-system TEP address 10.0.5.1 of DF PE-5 in ES "ES45_5" with ESI 01:00:00:00:00:45:00:00:00:05, as follows:

```
*A:PE-2# show service id 5 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address                            Egress VNI          Oper      Vxlan
                                                            State     Type
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================

===============================================================================
BGP EVPN VXLAN ES Dest
===============================================================================
I Eth Seg Id                    TEP Address       VNI       Last Changed
-------------------------------------------------------------------------------
1 01:00:00:00:00:45:00:00:00:05 10.0.5.1          5         06/29/2021 10:07:30
-------------------------------------------------------------------------------
===============================================================================
```

AA and SA multi-homed EVPN-VXLAN Epipe using non-system IPv6 addresses

Figure 159: EVPN-VXLAN Epipe 6 with AA MH and SA MH using non-system IPv6 addresses shows the example topology for EVPN-VXLAN Epipe 6 with AA MH ES "vES23_6" between PE-2 and PE-3 and SA MH ES "ES45_6" between PE-4 and PE-5.

Figure 159: EVPN-VXLAN Epipe 6 with AA MH and SA MH using non-system IPv6 addresses



The service configuration on PE-2 is as follows:

```
# on PE-2:
configure
    service
        system
            vxlan
                tunnel-termination 2001::2:1 fpe 1 create
            exit
            bgp-evpn
                ethernet-segment "vES23_6" virtual create
                    esi 01:00:00:00:00:23:00:00:00:06
                    es-orig-ip 2001::2:1
                    route-next-hop 2001::2:1
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 1
                    dot1q
                        q-tag-range 6
                    exit
                    no shutdown
                exit
            exit
        exit
        epipe 6 name "Epipe-6" customer 1 create
            vxlan-src-vtep 2001::2:1
            vxlan instance 1 vni 6 create
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-23_v6 create
                    eth-tag 623
                exit
                remote-attachment-circuit AC-45_v6 create
                    eth-tag 645
                exit
                evi 6
                vxlan bgp 1 vxlan-instance 1
                    send-tunnel-encap     # default
                    ecmp 2
                    no shutdown
                exit
            exit
            sap lag-1:6 create
                no shutdown
            exit
            no shutdown
        exit
```

The service configuration on PE-4 is as follows:

```
# on PE-4:
configure
    service
        sdp 466 create
            far-end 192.0.2.6
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        system
            vxlan
                tunnel-termination 10.0.4.1 fpe 1 create
                tunnel-termination 2001::4:1 fpe 1 create
            exit
            bgp-evpn
                ethernet-segment "ES45_6" create
                    esi 01:00:00:00:00:45:00:00:00:06
                    es-orig-ip 2001::4:1
                    route-next-hop 2001::4:1
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 466
                    no shutdown
                exit
            exit
        exit
        epipe 6 name "Epipe-6" customer 1 create
            vxlan-src-vtep 2001::4:1
            vxlan instance 1 vni 6 create
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-45_v6 create
                    eth-tag 645
                exit
                remote-attachment-circuit AC-23_v6 create
                    eth-tag 623
                exit
                evi 6
                vxlan bgp 1 vxlan-instance 1
                    send-tunnel-encap     # default
                    ecmp 2
                    no shutdown
                exit
            exit
            spoke-sdp 466:6 create
                no shutdown
            exit
            no shutdown
        exit
```

Conclusion

EVPN-VXLAN VPWS is similar to EVPN-MPLS VPWS, and can be used in networks without MPLS.

Fully Dynamic VSD Integration Model

This chapter provides information about fully dynamic virtualized service directory (VSD) integration model.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

Software requirements for this feature are SR OS Release 13.0.R4 or later and Nuage Virtualized Services Platform (VSP) release 3.2.R1 or later. This configuration was tested on SR OS Release 13.0.R4 and Nuage VSP release 3.2.R3.

Note:
Fully dynamic extensible messaging and presence protocol (XMPP) provisioning is not supported along with the dynamic business services feature in Release 13.0. Both features are mutually exclusive.

Note:
Provisioning of filter entries from Virtualized Services Directory (VSD) is not supported in SR OS 13.0.

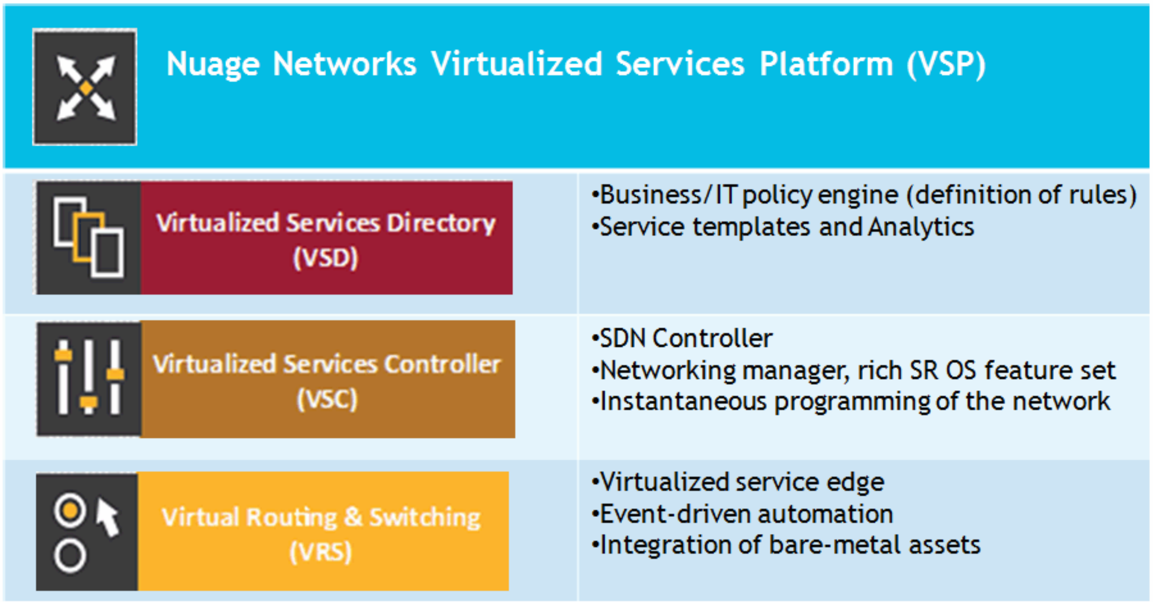
Nuage VSP conceptual knowledge and Nuage VSD operational knowledge are prerequisites. See the Nuage VSP user documentation for more information.

Overview

The Nuage VSP is a Software-Defined Networking (SDN) solution that provides data center (DC) network virtualization and automatically establishes connectivity between compute resources upon their creation. Leveraging programmable business logic and a powerful policy engine, the Nuage VSP provides an open and highly responsive solution that scales to meet the stringent needs of massive multi-tenant DCs. The Nuage VSP can be deployed over an existing DC IP network fabric, and has three main components:

Virtualized Services Directory (VSD), Virtualized Services Controller (VSC), and Virtual Routing and Switching (VRS), as displayed in Figure 160: Nuage VSP overview

Figure 160: Nuage VSP overview



Nuage Networks Virtualized Services Platform (VSP)

Virtualized Services Directory (VSD)

- •Business/IT policy engine (definition of rules)
- •Service templates and Analytics

Virtualized Services Controller (VSC)

- •SDN Controller
- •Networking manager, rich SR OS feature set
- •Instantaneous programming of the network

Virtual Routing & Switching (VRS)

- •Virtualized service edge
- •Event-driven automation
- •Integration of bare-metal assets

Virtualized Services Directory (VSD)

The Nuage VSD is a programmable policy and analytics engine. It provides a flexible and hierarchical network policy framework that enables IT administrators to define and enforce resource policies in a user-friendly manner.

The VSD contains a multi-tenant service directory, which supports role-based administration of users, computing, and network resources. The VSD also manages network resource assignments such as IP addresses and ACLs.

For the purpose of service assurance, the VSD allows the definition of sophisticated statistics rules, such as collection frequencies, rolling averages, and samples, as well as Threshold Crossing Alerts (TCA). When a TCA occurs, it will trigger an event that can be exported to external systems through a generic messaging bus. Statistics are aggregated over hours, days, and months, and stored in a Hadoop® analytics cluster to facilitate data mining and performance reporting.

The VSD runs as a number of processes in a virtual machine (VM) environment.

Virtualized Services Controller (VSC)

The Nuage VSC is an SDN controller. It functions as the robust network control plane for DCs, maintaining a full view of per-tenant network and service topologies. Through the VSC, virtual routing and switching constructs are established to program the network forwarding plane, the Nuage VRS, using the OpenFlow protocol.

The VSC communicates with the VSD policy engine using Extensible Messaging and Presence Protocol (XMPP). An ejabberd XMPP server/cluster is used to distribute messages between the VSD and VSC entities. Multiple VSC instances can be interconnected within and across DCs by leveraging Multi-Protocol Border Gateway Protocol (MP-BGP).

The VSC is based on the Service Router Operating System (SR OS) and runs in a virtual machine environment.

Virtual Routing and Switching (VRS)

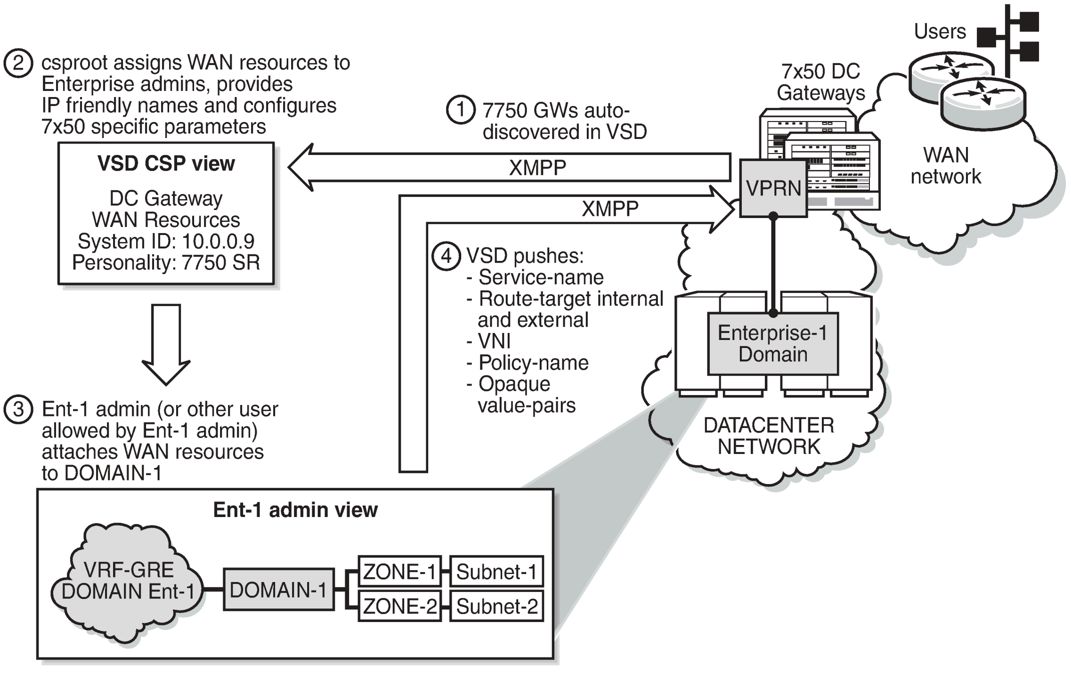
The Nuage VRS component is an enhanced Open vSwitch (OVS) implementation that constitutes the network forwarding plane. It encapsulates and de-encapsulates user traffic, enforcing L2 to L4 traffic policies as defined by the VSD. The VRS tracks Virtual Machine (VM) creation, migration, and deletion events in order to dynamically adjust network connectivity.

DC Gateway automated service provisioning

The first phase of VSD-SR OS integration was introduced in SR OS 12.0.R4. This phase included the development of an XMPP interface on the SR OS and the integration in the Nuage XMPP architecture. This so-called Static + Dynamic (S-D) provisioning model allows the auto-provisioning of VPLS and VPRN route targets, as well as VPLS VNI (VXLAN Network Identifiers) on the SR OS through the XMPP interface and the VSD interaction. The prerequisite in this model is the pre-configuration of the VPLS and VPRN services on the SR OS through CLI or SNMP. This model is intended to be used in DC Gateways where the WAN and the DC are managed by different administrative entities. The DC administrator will use VSD to “attach” the already configured VPLS or VPRN service to the L2 or L3 domain in the DC.

The second phase of VSD-SR OS integration was introduced in SR OS 13.0.R4. This phase supports the Fully Dynamic (F-D) provisioning model. The goal of this model is to avoid the prerequisite of pre-configuring the services on the SR OS existing in the S-D provisioning model, since this model assumes that the service is completely owned by the DC administrator. The entire service will be auto-generated on the SR OS as a result of the interaction with the VSD. Figure 161: DC Gateway fully dynamic provisioning workflow shows the workflow of the F-D provisioning model.

Figure 161: DC Gateway fully dynamic provisioning workflow



1. As soon as the XMPP server is configured on the SR OS DC Gateway, it is auto-discovered by the VSD.
2. The Cloud Service Provider (CSP) root user creates WAN services and assigns these resources to Enterprise administrators; for example, to the Ent-1 admin.
3. The Ent-1 admin sees the WAN service in the infrastructure resources and assigns permissions to certain user groups in Ent-1, who can consume these WAN resources by connecting to an L2/L3 domain.
4. As soon as the WAN service is added to an L2/L3 domain, the VSD pushes a list of parameters to the SR OS DC Gateway, which uses a python script to construct the configuration of the WAN service. The list of parameters sent to the 7750 routers can include:
   - **service-name** (Service ID field in the WAN Service GUI) - used as VSD domain in the CLI
   - **config-type** (Config Type field in the WAN Service GUI) - DYNAMIC for F-D XMPP provisioning
   - **service-type** (based on combination of the Service Type field and IRB check box in the WAN Service GUI) - possible values: L2DOMAIN, L2DOMAIN-IRB, VRF-GRE, or VRF-VXLAN
   - **name** (name of the L2 domain in the VSD to which the WAN service is assigned, or BackHaulSubnet in the case of service-type VRF-VXLAN)
   - **service-policy** (service Policy in the WAN Service GUI field) - should match the python policy configured on the SR OS DC Gateway
   - **vn-id** (VNI used for the Nuage overlay service) - dynamically supplied for VXLAN WAN services
   - **RT-I** (internal Route Target used for the Nuage overlay service)
   - **RT-E** (ext. Route Target in the WAN Service GUI field)

- **metadata** (list of opaque parameters supplied in the Metadata section of the WAN Service GUI)

The dynamic provisioning of parameters is provided for the following VSD domain types (configured in the SR OS DC Gateway):

l2-domain To attach a service at the gateway to an L2 (Ethernet) domain in the data center with no routing at the gateway, a VPLS service must be associated with a vsd-domain of type l2-domain.

l2-domain-irb To attach a service at the gateway to an L2 (Ethernet) domain in the data center with routing at the gateway, an R-VPLS service should be associated with a vsd-domain of type l2-domain-irb.

vrf-gre To attach a service at the gateway to an L3 domain (with GRE transport) in the data center, a VPRN service should be associated with a vsd-domain of type vrf-gre.

vrf-vxlan To attach a service at the gateway to an L3 domain (with VXLAN transport) in the data center, an R-VPLS service (with ip-route-advertisement enabled and linked to an EVPN-tunnel) should be associated with a vsd-domain of type vrf-vxlan.

This chapter will show examples of l2-domain, l2-domain-irb, and vrf-vxlan service type F-D provisioning, and focuses mostly on the SR OS DC Gateway configuration. For a more detailed F-D provisioning workflow on the VSD UI, refer to the VSP User Guide.

Python script

The XMPP parameters supplied by the VSD are parsed by a python script on the SR OS DC Gateway that dynamically provisions the VPLS and/or VPRN services provided for the Nuage overlay services.

The python script generates an executable CLI script based on the information received in the XMPP attributes. Three dynamic data service functions can be specified: **setup**, **modify**, or **teardown**. A fourth action, **revert**, is automatically invoked when the modify action fails:

- **setup** function: output = CLI to create a new dynamic data service.
- **teardown** function: output = CLI to delete an existing dynamic data service.
- **modify** function: output = CLI to change the parameters of an existing dynamic data service
- **revert** function: output = CLI to rollback the dynamic data service modify function actions in case of a modify failure

The python script uses the alc.dyn python module that contains a number of functions required to set up dynamic data services. To use the alc.dyn module, it must be imported into the python script:

```
from alc import dyn
```

The alc.dyn module contains a number of functions. Relevant alc.dyn functions for F-D XMPP provisioning are listed here:

- **dyn.action**(dictionary)
- **dyn.add_cli**(string)

- **dyn.select_free-id**(service-id)

The next sections provide a basic description of these functions. The alc.dyn module contains other functions that are not relevant for this feature. For a full list of the alc.dyn functions together with an extensive explanation of each function, refer to the RADIUS-Triggered Dynamic Data Service Provisioning chapter.

The trigger in the python script to execute a specific function is by calling the internal function **dyn.action**(d), where "d" is a python dictionary:

- d = { key : value, key : value, key : value, … , key : value }

For F-D XMPP provisioning, only 1 key:value pair is used and the key string must be set to "script".

The value is a tuple with the following comma separated values:

- (setup-function, modify-function, revert-function, teardown-function)

Setup and teardown functions are mandatory. Modify and revert functions are optional. If a modify function is defined, the revert function must also be defined. If no modify/revert function is required, the keyword **None** should be used instead.

The following two combinations are supported for F-D Dynamic XMPP provisioning:

1. without modify function:

```
d = {"script" : (setup_script, None, None, teardown_script)}
dyn.action(d)
```

2. with modify function (allows for changes in the WAN service on the VSD while the service is assigned to a domain):

```
d = {"script" : (setup_script, modify_script, revert_script, teardown_script)}
dyn.action(d)
```

When the configuration for a new service-name is received from the VSD, the vsd parameters and the opaque parameters string are concatenated into a single dictionary. The setup_script() is called and the dictionary is passed to the function. In this chapter, the dictionary will be named "vsdParams", but any other name would do. Within the python script:

- The VSD UI parameters are referenced as vsdParams['rt'], vsdParams['vni'], vsdParams['servicetype'], and so on.
- The metadata parameters are defined in an opaque string. For example, when the metadata string "rd=1:1,sap=1/1/1:1000" is supplied to the VSD WAN Service GUI, the format in the dictionary will be in the following format: "'metadata': 'rd=1:1,sap=1/1/1:1000 '".

To reference the metadata, the format is changed (trailing space is removed and parameters split up):

```
metadata = vsdParams['metadata']
    metadata = metadata.rstrip()
    metadata = dict(e.split('=') for e in metadata.split(','))
```

The individual metadata parameters can then be referenced in a similar way as the vsd parameters; for example, metadata['rd'], metadata['sap'], and so on.

When the startup script is executed, the **config>service>vsd>domain** is created outside the script context before running the actual script. The teardown script will remove the vsd domain. The domain-name is taken from the service-name supplied by the VSD ("Service ID" field in the WAN Service GUI - used as

VSD domain in the CLI). When testing the script with the **tools perform python-script** command, the domain-name is taken from the domain-name command parameter (see Testing the python script section).

When subsequent configuration messages are received from the VSD, the new parameter list is again generated from the VSD message and compared to the last parameter list that was successfully executed.

- If the two strings are identical, no action is taken.
- If there is a difference between the strings, the **modify_script()** function is called. For example, the **modify_script()** function is set up to handle a change in the service-mtu.

If a configuration message is received from the VSD for an existing service-name with no VSD parameters, the **teardown_script()** is called.

If a **setup_script()** fails, the **teardown_script()** is called.

To generate CLI output in the python script, an internal function, **dyn.add_cli**(output-string), is available. It adds the specified output-string to the CLI script. Python enables the use of triple quotes to specify strings that span multiple lines. For example:

```
from alc import dyn
    dyn.add_cli("""
configure
    service
        ies %(svc_id)s customer 1 create
            service-name "%(inst)s"
            description "%(inst)s"
            no shutdown
        exit
    exit
exit
""" % d)
```

An internal function, **dyn.select_free_id**("service-id"), is available to select a free (unused) service identifier in the service-range specified in the dynamic-services context (see the Configuration section). If no service-range is configured, the python script fails when **dyn.select_free_id**("service-id") is called. The service-id is made available again after a successful teardown (removal) of the service.

XMPP

The Extensible Messaging and Presence Protocol (XMPP) is an open technology for real-time communication, using XML (Extensible Markup Language) as the base format for exchanging information. XMPP provides a way to send small pieces of XML from one entity to another in near real time. Although initially intended for Instant Messaging applications, it can be easily extended to be used in a DC environment.

In the Nuage solution, each XMPP client, including the SR OS, is referred to with a JID(JabberID) in the following format: username@xmppserver.domain. The xmppserver.domainpoints to the XMPP server.

The Nuage VSP/SR OS DC Gateway solution uses the XMPP PubSub (Publish Subscribe) extension. This extension allows a user to subscribe to a node so that it can be notified whenever there is new or updated information available. The mechanism is used in this feature to auto-discover the username of the VSD JID. Additionally, the SR OS will subscribe to a separate PubSub for each DC Gateway, to discover updates on specific domains. Subscriptions are confirmed periodically (every 15 min).

The SR OS DC Gateway will periodically audit the VSD and request a DIFF list of F-D VSD domains. The VSD keeps a DIFF list of domains, which contains the F-D domain names for which the VSD has not received an info/query (IQ) request from the SR OS for a long time. The DC Gateway periodically checks the info for each of its deployed dynamic services with an IQ request (every 16-24 min). A DIFF or FULL domain list audit can also be triggered with the **tools perform service vsd fd-domain-sync <full> | <diff>** command.

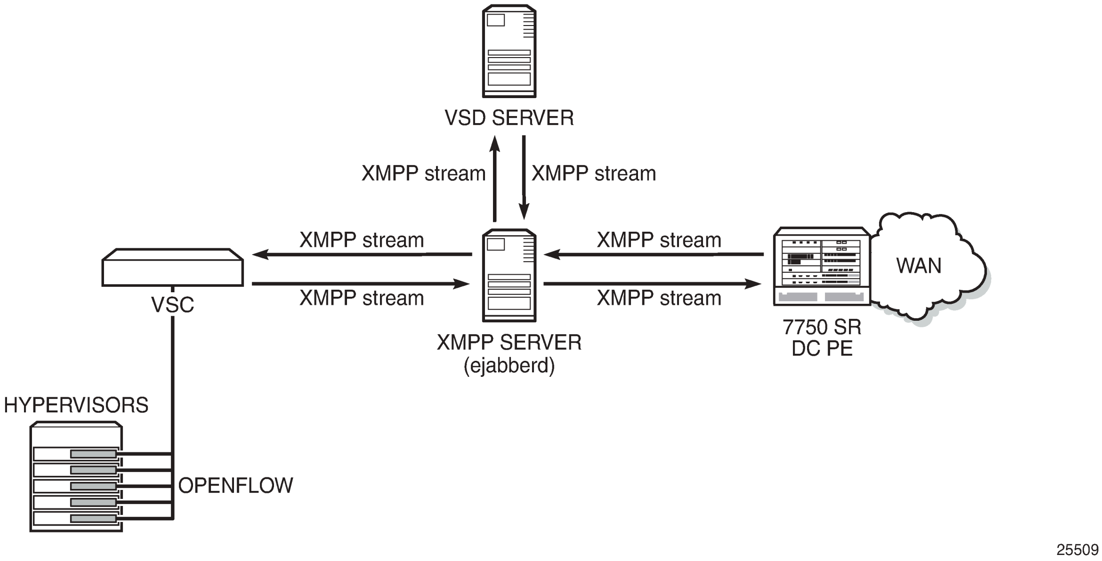
Configuration

This section describes the configuration that is required on the SR OS DC Gateway for F-D XMPP provisioning.

The following figure shows the basic setup used and also illustrates the XMPP architecture in the data center. Although the VSD and XMPP servers are represented by a single server, a cluster of VSD servers (using the same database) and/or XMPP servers will be a very common configuration in a data center.

It is assumed that underlying IP connectivity and an IGP has already been configured in this setup.

Figure 162: F-D XMPP provisioning setup



XMPP configuration

To receive configuration parameters from the VSD, the SR OS DC Gateway has to establish an XMPP client session with the XMPP server. Only one XMPP server can be configured.

When the XMPP server is properly configured, with no shutdown, the 7750 will try to establish a TCP session with the XMPP server through the management interface first. If it fails to establish communication, the 7750 will use in-band communication and will use its system IP as the source IP address.

To resolve the XMPP server fully qualified domain name (FQDN), provide a DNS server in the boot option file (bof) configuration and configure a dns-domain:

```
*A:pe-9>config>system# show bof
===============================================================================
BOF (Memory)
```

```
===============================================================================
---snip---
    primary-dns      138.203.39.47
    dns-domain       nuage.net
---snip---
===============================================================================
```

Then, configure the system-id of the DC Gateway that will be communicated to the VSD:

```
*A:pe-9>config>system# info
----------------------------------------------
#--------------------------------------------------
echo "vsd Configuration"
#--------------------------------------------------
        vsd
            system-id "pe9"
        exit
```

The next step is to configure the VSD server. The domain-name is the domain portion of the JID. The username is the username portion of the JID acting as an XMPP client. Ensure that the username uses all letters in lowercase (see SR OS 13.0 release notes). If no username is provided, an in-band registration will be provided, using the chassis MAC as username. The use of a password is optional:

```
*A:pe-9>config>system# info
#--------------------------------------------------
echo "Xmpp Configuration"
#--------------------------------------------------
        xmpp
            server vsd domain-name vsd1.nuage.net create username pe9
                no shutdown
            exit
        exit
----------------------------------------------
```

When the XMPP server has been configured, the state should move to "Functional":

```
*A:pe-9# show system xmpp server
===============================================================================
XMPP Server Table
===============================================================================
Name                            User Name          State
 XMPP FQDN                       Last State chgd    Admin State
-------------------------------------------------------------------------------
vsd                             pe9                Functional
 vsd1.nuage.net                  0d 00:37:01        inService
-------------------------------------------------------------------------------
No. of XMPP server's: 1
===============================================================================
```

XMPP Tx/Rx counters and other details can be obtained with the following command:

```
*A:pe-9# show system xmpp server "vsd"
===============================================================================
XMPP Server Table
===============================================================================
XMPP FQDN          : vsd1.nuage.net
XMPP Admin User    : pe9
XMPP Oper User     : pe9
State Lst Chg Since: 0d 00:07:43        State              : Functional
Admin State        : Up                 Connection Mode    : outOfBand
```

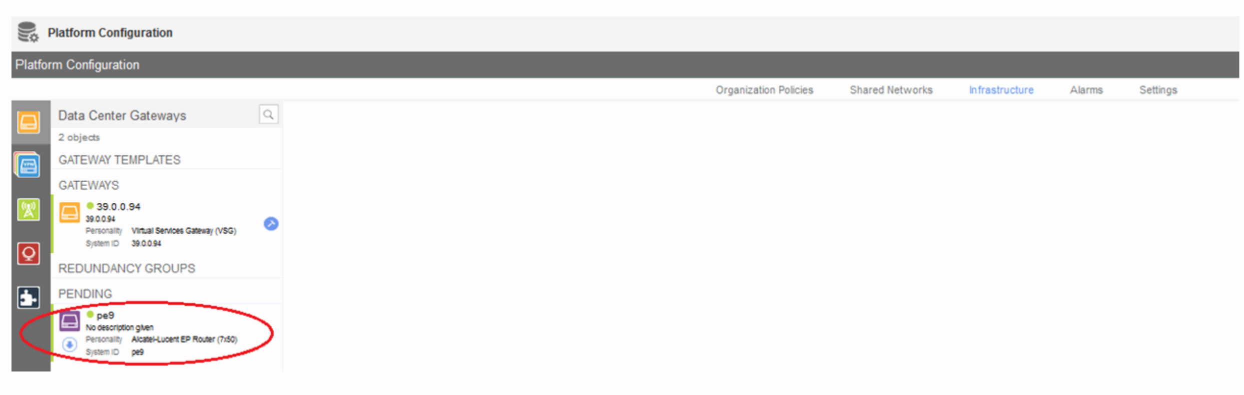
```
Auth Type           : md5
IQ Tx.              : 10                    IQ Rx.            : 10
IQ Error            : 0                     IQ Timed Out      : 0
IQ Min. Rtt         : 20 ms                 IQ Max. Rtt       : 80 ms
IQ Ack Rcvd.        : 10
Push Updates Rcvd   : 1                     VSD list Upd Rcvd  : 1
Msg Tx.             : 3                     Msg Rx.           : 3
Msg Ack. Rx.        : 3                     Msg Error         : 0
Msg Min. Rtt        : 0 ms                  Msg Max. Rtt      : 80 ms
Sub Tx.             : 1                     UnSub Tx.         : 0
Msg Timed Out       : 0
```

F-D XMPP provisioning uses the PubSub XMPP extension that allows each user to subscribe to a node, to be notified whenever that node gets new pieces of information or updated information.

The DC Gateway PubSub subscription state and subscriber name can be shown:

```
*A:pe-9# show system vsd
===============================================================================
VSD Information
===============================================================================
System Id                : pe9
GW Last Audit Tx Time    : 10/13/2015 13:56:34
Gateway Publish-Subscribe Information
-------------------------------------------------------------------------------
Subscribed               : True
Subscriber Name          : nuage_gateway_id_pe9
Last Subscription Time   : 10/13/2015 13:56:34
===============================================================================
```

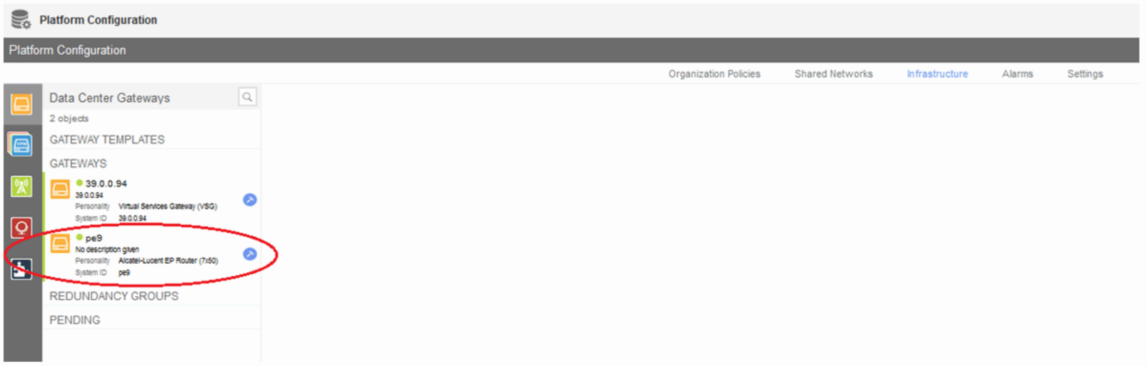
At the same time, the SR OS DC Gateway will be announced as a pending gateway in the VSD:



39.0.0.94
39.0.0.94
Personality Virtual Services Gateway (VSG)
System ID 39.0.0.94

pe9
No description given
Personality Alcatel-Lucent EP Router (7x50)
System ID pe9

The gateway will be promoted to the available gateways group by clicking on the arrow below the pending gateway icon:



BGP configuration

In the Nuage VSP solution, MP-BGP is used in the control plane to distribute MAC/IP information about the VMs. This information is distributed between the different VSCs, VSGs, and SR OS DC Gateways. Configure MP-BGP on the SR OS DC Gateway and the VSC/VSG (in this case, a VSG was used, but VSC is similar):

```
*A:pe-9>config>router# info
----------------------------------------------
#--------------------------------------------------
echo "IP Configuration"
#--------------------------------------------------
        interface "system"
            address 10.0.0.9/32
            no shutdown
        exit
---snip---
        autonomous-system 65000
---snip---
#--------------------------------------------------
echo "BGP Configuration"
#--------------------------------------------------
            min-route-advertisement 5
            rapid-withdrawal
            rapid-update evpn
            group "Nuage"
                family route-target evpn
                type internal
                neighbor 39.0.0.94
                exit
            exit
            no shutdown
```

```
*A:vsc1.nuage.net>config>router# info
----------------------------------------------
#--------------------------------------------------
echo "IP Configuration"
#--------------------------------------------------
---snip---
        interface "system"
            address 39.0.0.94/32
            no shutdown
        exit
        autonomous-system 65000
```

```
---snip---
#--------------------------------------------------
echo "BGP Configuration"
#--------------------------------------------------
        bgp
            family route-target evpn
            min-route-advertisement 5
            rapid-withdrawal
            rapid-update evpn
            group "internal"
                type internal
                neighbor 10.0.0.9
                    family evpn
                exit
            exit
            no shutdown
        exit
----------------------------------------------
```

In this setup, the family type "evpn" and "route-target" is used. The former is used to learn the EVPN route updates while the latter is restricting the SR OS to only learn those MB-BGP routes for which it has a route target configured.

Note:
To use vrf-gre domains, configure BGP family "vpn-ipv4" as well. Similarly, to use BGP-MH (for example, in case of redundant SR OS DC Gateways with L2-domains), the use of BGP family "l2vpn" is required.

Verify that BGP peering is in the operational state:

```
*A:pe-9# show router bgp summary
===============================================================================
 BGP Router ID:10.0.0.9         AS:65000       Local AS:65000
===============================================================================
BGP Admin State         : Up          BGP Oper State              : Up
---snip---
                   AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
39.0.0.94
                65000   76287    0 00h04m14s 0/0/0 (RouteTarget)
                         2634    0           0/0/0 (Evpn)
-------------------------------------------------------------------------------
```

Dynamic VSD Services range

F-D XMPP provisioning requires a reserved range of Service-IDs that can be used for dynamic data services. This configured range is no longer available for regular services configured via CLI/SNMP:

```
*A:pe-9>config>service# info
----------------------------------------------
        vsd
            service-range 64000 to 64999
        exit
```

```
*A:pe-9# show service vsd summary
===============================================================================
VSD Information
===============================================================================
Service Range
Start              : 64000                       End           : 64999
===============================================================================
No domain entries found
```

Python script

The python script that will build the dynamic services based on the VSD parameters obtained via XMPP can be stored locally on the CF or on a remote FTP server:

```
*A:pe-9>config>python# info
----------------------------------------------
        python-script "l2domain_services" create
            primary-url "ftp://*:*@138.203.15.48/./l2domain_service.py"
            no shutdown
        exit
```

The script is loaded into memory as soon as a no shutdown is performed and is reloaded with each **shutdown**/**no shutdown** action. Alternatively, a tools command can be used to reload the script:

```
*A:pe-9# tools perform python-script reload "l2domain_services"
```

In case incorrect python syntax is used in the script, an error message is displayed after a **no shutdown** of the script (or a reload with the tools command), with an indication of the line where the error is located.

The details of the python script can be inspected:

```
*A:pe-9# show python python-script "l2domain_services"
===============================================================================
Python script "l2domain_services"
===============================================================================
Description   : (Not Specified)
Admin state   : inService
Oper state    : inService
Action on fail: drop
Protection    : none
Primary URL   : ftp://*:*@138.203.15.48/./l2domain_service.py
Secondary URL : (Not Specified)
Tertiary URL  : (Not Specified)
Active URL    : primary
Last changed  : 10/13/2015 09:54:26
===============================================================================
```

The contents of the python script can also be viewed. The contents of the script are shown in the "Test the python script" section:

```
*A:pe-9# show python python-script "l2domain_services" source-in-use
===============================================================================
Python script "l2domain_services"
===============================================================================
Admin state   : inService
Oper state    : inService
Primary URL   : ftp://*:*@138.203.15.48/./l2domain_service.py
Secondary URL : (Not Specified)
Tertiary URL  : (Not Specified)
```

```
Active URL    : primary
-------------------------------------------------------------------------------
Source (dumped from memory)
-------------------------------------------------------------------------------
      1 from alc import dyn
      2
      3 # example of metadata to be added in VSD WAN Service: "rd=1:1,sap=1/1/1:1000"
      4
---snip---
      8 def setup_script(vsdParams):
---snip---
     70 def modify_script(vsdParams,setup_result):
---snip---
    106 def revert_script(vsdParams,setup_result):
---snip---
    129 def teardown_script(setupParams):
---snip---
    164 d = {"script" : (setup_script, modify_script, revert_script, teardown_script)}
    165
    166 dyn.action(d)
===============================================================================
```

A list of CLI command nodes that can be used with python script function dyn.add_cli is provided with the **tools dump service vsd-services command-list** command. In general, all the 'leaf' commands under the nodes shown in the tools dump command, can be used with the python script.

Further restriction of CLI commands is possible by creating a separate CLI user for the XMPP interface and associate that user with a profile where the commands are limited.

The CLI user for the XMPP interface is configurable:

```
config>system>security>cli-script>authorization>
        vsd
[no] cli-user <username>
```

Python policy

Python scripts are called by a python policy that will be referred to in the VSD WAN services GUI in the Service Policy field.

Create a python policy (that will be referred to in the VSD GUI) and link the python policy to the python script:

```
*A:pe-9>config>python# info
----------------------------------------------
        python-policy "py-l2" create
            description "Python script to create L2 domains"
            vsd script "l2domain_services"
        exit
```

```
*A:pe-9# show python python-script "l2domain_services" association
===============================================================================
Python Script Association
===============================================================================
Policy                           Type                                    Dir
-------------------------------------------------------------------------------
py-l2                            vsdAccessRequest                        ingress
```

```
-------------------------------------------------------------------------------
```

```
*A:pe-9# show python python-policy "py-l2"
===============================================================================
Python policy "py-l2"
===============================================================================
Description                 : Python script to create L2 domains
-------------------------------------------------------------------------------
Messages
-------------------------------------------------------------------------------
Type                                   Dir     Script
-------------------------------------------------------------------------------
vsdAccessRequest                       ingress l2domain_services
-------------------------------------------------------------------------------
```

Test the python script

The python script can be tested separately on the SR OS DC Gateway; even before connecting it to the Nuage setup.

Some notes about the python script and creating dynamic services:

- The VSD (and **tools evaluate-script** command) will provide some compulsory parameters like:
  - domain-name
  - domain type (l2-domain|vrf-gre|vrf-vxlan|l2-domain-irb)
  - action (setup or teardown)
  - vni
  - RT (internal and external route-targets are provided)
- The VSD (and **tools evaluate-script** command) can provide extra metadata that is supplied in a text string and comma separated. For example, metadata "rd=1:1,sap=1/1/1:1000".
- The RT format supplied by the VSD though XMPP (that is, "x:x") differs from the RT format that can be used with the **tools evaluate-script** command (for example, "target:x:x"). For that reason, it can be useful to add the following check in the script:

  ```
  if not rt.startswith ('target'):
       rt = "target:"+rt
  ```

- The VSD metadata string includes an empty space at the end. This can be removed with the following python command:

  ```
  metadata = metadata.rstrip()
  ```

- If a configuration message is received from the VSD for an existing service-name with no VSD parameters, or if a setup_script() fails, the teardown_script() is called.
- At any point in the script, you can add **print** commands to check the status/content of various parameters.

The python policy and script can then be tested with the tools evaluate-script command. Example syntax has been added into the scripts for convenience.

Before testing the script, it will be useful to enable the following debugging:

```
debug
    python
        python-script "l2domain_services"
            script-all-info
        exit
    exit
    vsd
        scripts
            event
                cli
                errors
                executed-cmd
                warnings
                state-change
            exit
        exit
    exit
exit
```

In this section, a python example script for an l2-domain is used. The contents of the script are shown in the following text format. Examples for l2-domain-irb and vrf-vxlan type domains are shown in dedicated sections:

```
from alc import dyn
# example of metadata to be added in VSD WAN Service: "rd=1:1,sap=1/1/1:1000"
# example of tools cli to test this script: tools perform service vsd evaluate-script domain-
name "l2dom1" type l2-domain action setup policy "py-l2" vni 1234 rt-i target:1:1 rt-e
 target:1:1 metadata "rd=1:1,sap=1/1/1:1000"

# teardown example cli: tools perform service vsd evaluate-script domain-name "l2dom1" type l2-
domain action teardown policy "py-l2" vni 1234 rt-i target:1:1 rt-e target:1:1

def setup_script(vsdParams):

    print ("These are the VSD params: " + str(vsdParams))
    servicetype = vsdParams['servicetype']
    vni = vsdParams['vni']
    rt = vsdParams['rt']

# add "target:" if provisioned by VSD (VSD uses x:x format whereas tools command uses
 target:x:x format)
    if not rt.startswith ('target'):
       rt = "target:"+rt
    metadata = vsdParams['metadata']

# remove trailing space at the end of the metadata
    metadata = metadata.rstrip()
    print ("VSD metadata" + str(metadata))


    metadata = dict(e.split('=') for e in metadata.split(','))
    print ("Modified metadata" + str(metadata))
    vplsSvc_id = dyn.select_free_id("service-id")
    print ("this is the free svc id picked up by the system: " + vplsSvc_id)
    if servicetype == "L2DOMAIN":

      rd = metadata['rd']
      sap_id = metadata['sap']
      print ('servicetype, VPLS id, rt, vni, rd, sap:', servicetype, vplsSvc_id, rt, vni, rd,
 sap_id)
```

```
      dyn.add_cli("""
        configure service
           vpls %(vplsSvc_id)s customer 1 create
              description vpls%(vplsSvc_id)s
              proxy-arp
                 dynamic-arp-populate
                 no shutdown
                 exit
              bgp
                 route-distinguisher %(rd)s
                 route-target %(rt)s
              exit
              vxlan vni %(vni)s create
              exit
              bgp-evpn
                  evi %(vplsSvc_id)s
                  vxlan
                      no shutdown
                  exit
              exit
              service-name evi%(vplsSvc_id)s
              sap %(sap_id)s create
              exit
              no shutdown
              exit
            exit
        exit
      """ % {'vplsSvc_id' : vplsSvc_id, 'vni' : vsdParams['vni'], 'rt' : rt, 'rd' :
 metadata['rd'], 'sap_id' : sap_id})
      # L2DOMAIN returns setupParams: vplsSvc_id, servicetype, vni, sap
      return {'vplsSvc_id' : vplsSvc_id, 'servicetype' : servicetype, 'vni' : vni, 'sap_id' :
 sap_id}
#----------------------------------------------------------------------------------------------
--
def modify_script(vsdParams,setup_result):

    print ("These are the setup_result params for modify_script: " + str(setup_result))
    print ("These are the VSD params for modify_script: " + str(vsdParams))

    # remove trailing space at the end of the metadata
    metadata = vsdParams['metadata'].rstrip()

    print ("VSD metadata" + str(metadata))
    metadata = dict(e.split('=') for e in metadata.split(','))
    print ("Modified metadata" + str(metadata))


    # updating the setup_result dict
    setup_result.update(metadata)
    params = setup_result

    print ("The updated params from metadata and return from the setup result: " + str(params))

    svc_mtu = params['svc-mtu']

    dyn.add_cli("""
      configure service
          vpls %(vplsSvc_id)s
            service-mtu %(svc-mtu)s
             exit
          exit
      exit
    """ %params )
    # Result is passed to teardown_script
    return params
#----------------------------------------------------------------------------------------------
--
def revert_script(vsdParams,setup_result):
    print ("These are the setup_result params for revert_script: " + str(setup_result))
    print ("These are the VSD params for revert_script: " + str(vsdParams))

    # When modify fails, the revert is called and then the teardown is called.
    # It is recommended to revert to same value as used in setup for the attributes modified in
 modify_script.

    params = setup_result

    dyn.add_cli("""
      configure service
          vpls %(vplsSvc_id)s
            service-mtu 2000
             exit
          exit
      exit
    """ %params )

    # Result is passed to teardown_script
    return params

#----------------------------------------------------------------------------------------------
--
def teardown_script(setupParams):
    print ("These are the teardown_script setupParams: " + str(setupParams))
    servicetype = setupParams['servicetype']
    if servicetype == "L2DOMAIN":
      dyn.add_cli("""
        configure service
            vpls %(vplsSvc_id)s
               no description
               proxy-arp shut
               no proxy-arp
               bgp-evpn
                   vxlan
                       shut
                   exit
                   no evi
                   exit
               no vxlan vni %(vni)s
               bgp
                  no route-distinguisher
                  no route-target
               exit
               no bgp
               no bgp-evpn
               sap %(sap_id)s
                  shutdown
                  exit
               no sap %(sap_id)s
               shutdown
               exit
               no vpls %(vplsSvc_id)s
            exit
        exit
      """ % {'vplsSvc_id' : setupParams['vplsSvc_id'], 'vni' : setupParams['vni'], 'sap_id' :
 setupParams['sap_id']})
      return setupParams
```

```
d = {"script" : (setup_script, modify_script, revert_script, teardown_script)}

dyn.action(d)
```

The script can be tested with the following command:

```
*A:pe-9# tools perform service vsd evaluate-script domain-name "l2dom1" type l2-domain action
 setup policy "py-l2" vni 1234 rt-i target:1:1 rt-e target:1:1 metadata "rd=1:1,sap=1/1/1:1000"
1 2015/10/15 09:51:16.08 UTC MINOR: DEBUG #2001 Base dyn-script req=setup
"dyn-script req=setup: l2dom1
  state=init->waiting-for-setup
"
2 2015/10/15 09:51:16.08 UTC MINOR: DEBUG #2001 Base dyn-script req=setup
"dyn-script req=setup: l2dom1
  state=waiting-for-setup->generating-setup
"
3 2015/10/15 09:51:16.08 UTC MINOR: DEBUG #2001 Base Python Output
"Python Output: l2domain_services
These are the VSD params: {'rt': 'target:1:1', 'rte': 'target:1:1', 'domain': ''
, 'servicetype': 'L2DOMAIN', 'vni': '1234', 'metadata': 'rd=1:1,sap=1/1/1:1000 '
}
VSD metadatard=1:1,sap=1/1/1:1000
Modified metadata{'rd': '1:1', 'sap': '1/1/1:1000'}
this is the free svc id picked up by the system: 64000
('servicetype, VPLS id, rt, vni, rd, sap:', 'L2DOMAIN', '64000', 'target:1:1', '
1234', '1:1', '1/1/1:1000')
"
4 2015/10/15 09:51:16.08 UTC MINOR: DEBUG #2001 Base Python Result
"Python Result: l2domain_services
"
5 2015/10/15 09:51:16.08 UTC MINOR: DEBUG #2001 Base dyn-script req=setup
"dyn-script req=setup: l2dom1
  state=generating-setup->executing-setup
"
6 2015/10/15 09:51:16.08 UTC MINOR: DEBUG #2001 Base dyn-script cli 1/1
"dyn-script cli 1/1: script:l2dom1(cli 705 dict 0->123)
        configure service
           vpls 64000 customer 1 create
              description vpls64000
              proxy-arp
                 dynamic-arp-populate
                 no shut
                 exit
              bgp
                 route-distinguisher 1:1
                 route-target target:1:1
              exit
              vxlan vni 1234 create
              exit
              bgp-evpn
                  evi 64000
                  vxlan
                      no shut
                  exit
              exit
              service-name evi64000
              sap 1/1/1:1000 create
              exit
              no shutdown
              exit
            exit
        exit
```

```
      "
7 2015/10/15 09:51:16.08 UTC MINOR: DEBUG #2001 Base dyn-script setup
"dyn-script setup: l2dom1 script:l2dom1 line 2
 configure service"
Success
---snip---
24 2015/10/15 09:51:16.08 UTC MINOR: DEBUG #2001 Base dyn-script req=setup
"dyn-script req=setup: l2dom1
  state=executing-setup->established
"
```

At this moment a new VSD domain has been created as well as a new service:

```
*A:pe-9# show service vsd domain
===============================================================================
VSD Domain Table
===============================================================================
Name                                    Type           Origin    Admin
-------------------------------------------------------------------------------
l2dom1                                  l2Domain       vsd       inService
-------------------------------------------------------------------------------
```

```
*A:pe-9# show service vsd domain "l2dom1" association
============================================================
Service VSD Domain
============================================================
Svc Id      Svc Type  Domain Type    Domain Admin   Origin
------------------------------------------------------------
64000       vpls      l2Domain       inService      vsd
------------------------------------------------------------
```

```
*A:pe-9# show service vsd domain "l2dom1"
===============================================================================
VSD Information
===============================================================================
Name               : l2dom1
Description        : l2dom1
Type               : l2Domain                    Admin State   : inService
Last Error To Vsd  : (Not Specified)
Last Error From Vsd: (Not Specified)
Statistics
-------------------------------------------------------------------------------
Last Cfg Chg Evt   : 10/14/2015 16:02:54          Cfg Chg Evts  : 1
Last Cfg Update    : 10/14/2015 16:02:54          Cfg Upd Rcvd  : 1
Last Cfg Done      : 10/14/2015 16:02:54Cfg Success        : 1                            Cfg
 Failed    : 0
Last Recd Params   : script = {'domain' : '', 'vn
                   : i' : '1234', 'rt' : 'target:
                   : 1:1', 'rte' : 'target:1:1',
                   : 'servicetype' : 'L2DOMAIN',
                   : 'metadata' : 'rd=1:1,sap=1/1
                   : /1:1000 '}
Last Exec Params   : script = {'domain' : '', 'vn
                   : i' : '1234', 'rt' : 'target:
                   : 1:1', 'rte' : 'target:1:1',
                   : 'servicetype' : 'L2DOMAIN',
                   : 'metadata' : 'rd=1:1,sap=1/1
                   : /1:1000 '}
===============================================================================
```

```
*A:pe-9# show service service-using
```

```
===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
64000        VPLS      Up   Up   1          evi64000
2147483648   IES       Up   Down 1          _tmnx_InternalIesService
2147483649   intVpls   Up   Down 1          _tmnx_InternalVplsService
-------------------------------------------------------------------------------
```

Note:
Service ID 2147483648 and 2147483649 are internal services that are always present on the SR OS. They are not relevant for this feature and will be truncated in other output examples in this document.

```
*A:pe-9# show service id 64000 all
===============================================================================
Service Detailed Information
===============================================================================
Service Id        : 64000               Vpn Id            : 0
Service Type      : VPLS
Name              : evi64000
Description       : vpls64000
Customer Id       : 1                   Creation Origin   : vsd
Last Status Change: 10/14/2015 16:02:54
Last Mgmt Change  : 10/14/2015 16:02:54
Etree Mode        : Disabled
Admin State       : Up                  Oper State        : Up
MTU               : 1514                Def. Mesh VC Id   : 64000
SAP Count         : 1                   SDP Bind Count    : 0
---snip---
VSD Domain        : l2dom1
---snip---
-------------------------------------------------------------------------------
BGP Information
-------------------------------------------------------------------------------
Vsi-Import        : None
Vsi-Export        : None
Route Dist        : 1:1
Oper Route Dist   : 1:1
Oper RD Type      : configured
Rte-Target Import : 1:1                 Rte-Target Export : 1:1
Oper RT Imp Origin: configured          Oper RT Import    : 1:1
Oper RT Exp Origin: configured          Oper RT Export    : 1:1
PW-Template Id    : None
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
SAP 1/1/1:1000
-------------------------------------------------------------------------------
Service Id         : 64000
SAP                : 1/1/1:1000               Encap             : q-tag
Description        : (Not Specified)
Admin State        : Up                       Oper State        : Up
---snip---
```

Note:
You cannot see the dynamic VSD services in the configuration nor can you edit their configuration under normal circumstances (this is discussed further in the next section).

```
*A:pe-9>config>service# info
----------------------------------------------
        customer 1 create
            description "Default customer"
        exit
        vsd
            service-range 64000 to 64999
        exit
----------------------------------------------
```

```
*A:pe-9# configure service vpls 64000
MINOR: CLI Modification of services created by a dynamic script is not allowed.
```

The service can be modified by adding/changing a service-mtu to the metadata. This will trigger the modify-script function in the python script. The following basic script is only an example of how a modify-script function operates. The script could be extended to modify other parameters as well; however this is out of the scope of this chapter:

```
*A:pe-9# tools perform service vsd evaluate-script domain-name "l2dom1" type l2-domain action
 modify policy "py-l2" vni 1234 rt-i target:1:1 rt-e target:1:1 metadata "rd=1:1,sap=1/1/
1:1000,svc-mtu=2222"
25 2015/10/15 09:51:22.44 UTC MINOR: DEBUG #2001 Base dyn-script req=modify
"dyn-script req=modify: l2dom1
  state=established->waiting-for-modify
"
26 2015/10/15 09:51:22.44 UTC MINOR: DEBUG #2001 Base dyn-script req=modify
"dyn-script req=modify: l2dom1
  state=waiting-for-modify->generating-modify
"
27 2015/10/15 09:51:22.44 UTC MINOR: DEBUG #2001 Base Python Output
"Python Output: l2domain_services
These are the setup_result params for modify_script: {'servicetype': 'L2DOMAIN',
 'vplsSvc_id': '64000', 'vni': '1234', 'sap_id': '1/1/1:1000'}
These are the VSD params for modify_script: {'rt': 'target:1:1', 'rte': 'target:
1:1', 'domain': '', 'servicetype': 'L2DOMAIN', 'vni': '1234', 'metadata': 'rd=1:
1,sap=1/1/1:1000,svc-mtu=2222 '}
VSD metadatard=1:1,sap=1/1/1:1000,svc-mtu=2222
Modified metadata{'rd': '1:1', 'sap': '1/1/1:1000', 'svc-mtu': '2222'}
The updated params from metadata and return from the setup result: {'rd': '1:1',
 'servicetype': 'L2DOMAIN', 'svc-mtu': '2222', 'sap_id': '1/1/1:1000', 'sap': '1
/1/1:1000', 'vplsSvc_id': '64000', 'vni': '1234'}
"
Success
28 2015/10/15 09:51:22.44 UTC MINOR: DEBUG #2001 Base Python Result
"Python Result: l2domain_services
"
29 2015/10/15 09:51:22.44 UTC MINOR: DEBUG #2001 Base dyn-script req=modify
"dyn-script req=modify: l2dom1
  state=generating-modify->executing-modify
"
*A:pe-9#
30 2015/10/15 09:51:22.44 UTC MINOR: DEBUG #2001 Base dyn-script cli 1/1
"dyn-script cli 1/1: script:l2dom1(cli 123 dict 123->203)
      configure service
          vpls 64000
            service-mtu 2222
```

```
             exit
          exit
      exit
    "
31 2015/10/15 09:51:22.44 UTC MINOR: DEBUG #2001 Base dyn-script modify
"dyn-script modify: l2dom1 script:l2dom1 line 2
 configure service"
---snip---
36 2015/10/15 09:51:22.44 UTC MINOR: DEBUG #2001 Base dyn-script req=commit
"dyn-script req=commit: l2dom1
  state=waiting-for-commit->established
"
```

The service-mtu has now been changed to 2222:

```
show service id 64000 all | match MTU
MTU               : 2222                Def. Mesh VC Id   : 64000
Admin MTU          : 9212                     Oper MTU          : 9212
```

The service can be removed with the teardown script:

```
*A:pe-9# tools perform service vsd evaluate-script domain-name "l2dom1" type l2-domain action
 teardown policy "py-l2" vni 1234 rt-i target:1:1 rt-e target:1:1
37 2015/10/15 09:51:29.80 UTC MINOR: DEBUG #2001 Base dyn-script req=teardown
"dyn-script req=teardown: l2dom1
  state=established->waiting-for-teardown
"
38 2015/10/15 09:51:29.80 UTC MINOR: DEBUG #2001 Base dyn-script req=teardown
"dyn-script req=teardown: l2dom1
  state=waiting-for-teardown->generating-teardown
"
39 2015/10/15 09:51:29.80 UTC MINOR: DEBUG #2001 Base Python Output
"Python Output: l2domain_services
These are the teardown_script setupParams: {'servicetype': 'L2DOMAIN', 'svc-mtu'
: '2222', 'sap_id': '1/1/1:1000', 'vplsSvc_id': '64000', 'vni': '1234', 'rd': '1
:1', 'sap': '1/1/1:1000'}
"
40 2015/10/15 09:51:29.80 UTC MINOR: DEBUG #2001 Base Python Result
"Python Result: l2domain_services
"
41 2015/10/15 09:51:29.80 UTC MINOR: DEBUG #2001 Base dyn-script req=teardown
"dyn-script req=teardown: l2dom1
  state=generating-teardown->executing-teardown
"
42 2015/10/15 09:51:29.80 UTC MINOR: DEBUG #2001 Base dyn-script cli 1/1
"dyn-script cli 1/1: script:l2dom1(cli 709 dict 203->0)
        configure service
            vpls 64000
               no description
               proxy-arp shut
               no proxy-arp
               bgp-evpn
                   vxlan
                       shut
                   exit
                   no evi
                   exit
               no vxlan vni 1234
               bgp
                  no route-distinguisher
                  no route-target
               exit
               no bgp
               no bgp-evpn
               sap 1/1/1:1000
                  shutdown
                  exit
               no sap 1/1/1:1000
               shutdown
               exit
               no vpls 64000
            exit
        exit
      "
43 2015/10/15 09:51:29.80 UTC MINOR: DEBUG #2001 Base dyn-script teardown
"dyn-script teardown: l2dom1 script:l2dom1 line 2
 configure service"
---snip---
63 2015/10/15 09:51:29.81 UTC MINOR: DEBUG #2001 Base dyn-script req=teardown
"dyn-script req=teardown: l2dom1
  state=executing-teardown->stopped
"
```

After the dynamic service has been torn down, the dynamic service and the VSD domain should not be present on the SR OS DC Gateway:

```
*A:pe-9# show service service-using
===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
2147483648   IES       Up   Down 1          _tmnx_InternalIesService
2147483649   intVpls   Up   Down 1          _tmnx_InternalVplsService
-------------------------------------------------------------------------------
Matching Services : 2
-------------------------------------------------------------------------------
===============================================================================
```

```
*A:pe-9# show service vsd domain
No domain entries found
```

Editing dynamic VSD services

As indicated in the previous section, the dynamic VSD services CLI configuration cannot be shown or edited normally. However, under certain circumstances, it might be necessary to inspect/change/remove the configuration of a dynamic VSD service; for example, when the python VSD script was not using the correct syntax and the creation/deletion of the dynamic VSD service failed.

It is possible to edit the dynamic VSD services configuration by entering the **enable-vsd-config** mode. First, create a password, which is required to enter this mode:

```
*A:pe-9# configure system security password
*A:pe-9>config>system>security>password# vsd-password *****
```

Then, enter the enable-vsd-config mode. You will be asked for the previously configured password:

```
*A:pe-9# enable-vsd-config
Password:
```

Now you can edit the dynamic VSD services configuration and change/add/remove configuration:

```
*A:pe-9# configure service
*A:pe-9>config>service# info
----------------------------------------------
        customer 1 create
            description "Default customer"
        exit
        vsd
            domain l2dom1 type l2-domain create
                description "l2dom1"
                no shutdown
            exit
            service-range 64000 to 64999
        exit
        vpls 64000 customer 1 create
            description "vpls64000"
            vxlan vni 1234 create
            exit
            bgp
                route-distinguisher 1:1
            exit
            bgp-evpn
                evi 64000
                vxlan
                    no shutdown
                exit
                mpls
                    shutdown
                exit
            exit
            proxy-arp
                dynamic-arp-populate
                no shutdown
            exit
            stp
                shutdown
            exit
            service-name "evi64000"
            sap 1/1/1:1000 create
            exit
            vsd-domain "l2dom1"
            no shutdown
        exit
----------------------------------------------
```

In the enable-vsd-config mode, only dynamic VSD services can be edited, not regular CLI-based services:

```
*A:pe-9# configure service vpls 100 customer 1 create
MINOR: SVCMGR #1201 Invalid service-id - not reserved
```

After inspecting/editing the dynamic VSD services configuration, you should exit this mode again:

```
*A:pe-9# no enable-vsd-config
```

L2 VXLAN

An example python script for F-D XMPP provisioning of an L2 VXLAN type service (l2-domain) was provided in the "Testing the python script" section. In this section, the same script is used for provisioning via the VSD.

To dynamically provision this type of service, a few things must be configured on the VSD: (screenshots of this workflow are available in the VSP User Guide)

Create an L2 WAN service in the VSD:

- under Platform Config/Infrastructure, select the DC Gateway to add a WAN service
- select Service Type "layer 2" (no IRB)
- select "Dynamic" configuration type to allow Fully Dynamic provisioning
- under Service Policy, provide the python policy configured on the DC Gateway
- provide a Name and Service-ID (the Service-ID will be the name of the dynamically created service domain on the DC Gateway)

Add the metadata to the WAN service:

- right-click the WAN service and select "inspect"
- a dialog box appears, select the "Metadata" tag and add the metadata info "rd=1:1,sap=1/1/1:1000"

Add permissions to Enterprise1. The WAN service should now be visible in Enterprise1. Add permissions for a group of users to use this WAN service. Instantiate the L2 domain and attach the WAN service.

As soon as the WAN service is attached to the L2 domain, the VSD will send a notification via XMPP to the DC Gateway about the new Service-ID. The VSD will send an XMPP IQ request to the VSD to obtain the VSD service parameters.

Note:
There is an 8 to 12 s delay. The command **tools perform service vsd domain refresh-config** can be used to expedite the request.

As soon as the DC Gateway receives this information, the python policy mentioned in the VSD service parameters is triggered and the VSD parameters are passed to the associated python script. The python script will then construct and execute the same configuration CLI and trigger similar debug information as is shown in the section "Testing the python script".

After the python script has completed successfully, the service can be inspected in a similar way as before.

- **show service service-using**
- **show service id** <*service-id*> **all**
- enter enable-vsd-config mode if required and inspect the service config

MAC addresses can now be learned in the Nuage VSC/VSG and sent via MP-BGP to the DC Gateway:

```
*A:pe-9# show service id 64000 fdb detail
===============================================================================
Forwarding Database, Service 64000
===============================================================================
ServId    MAC               Source-Identifier        Type     Last Change
                                                     Age
-------------------------------------------------------------------------------
64000     1e:50:01:01:00:01 vxlan:                   Evpn     10/14/15 16:10:44
```

```
                            39.0.0.94:1006636
-------------------------------------------------------------------------------
```

In case HyperVisors (HVs) with VRS are deployed or a Host vPORT has been connected to a VSG, the EVPN MAC/Route (type 2) will also include the IP address of the VM/Host, in which case the DC Gateway can perform a proxy-arp function. For more information about EVPN-VXLAN features, refer to the chapters EVPN for VXLAN Tunnels (Layer 2) and EVPN for VXLAN Tunnels (Layer 3).

```
*A:pe-9# show service id 64000 proxy-arp detail
-------------------------------------------------------------------------------
Proxy Arp
-------------------------------------------------------------------------------
Admin State       : enabled
Dyn Populate      : enabled
Age Time          : disabled            Send Refresh      : disabled
Table Size        : 250                 Total             : 1
Static Count      : 0                   EVPN Count        : 1
Dynamic Count     : 0                   Duplicate Count   : 0
Dup Detect
-------------------------------------------------------------------------------
Detect Window     : 3 mins              Num Moves         : 5
Hold down         : 9 mins
Anti Spoof MAC    : None
EVPN
-------------------------------------------------------------------------------
Garp Flood        : enabled             Req Flood         : enabled
-------------------------------------------------------------------------------
===============================================================================
VPLS Proxy Arp Entries
===============================================================================
IP Address          Mac Address         Type      Status    Last Update
-------------------------------------------------------------------------------
10.32.78.100        1e:50:01:01:00:01   evpn      active    07/22/2015 10:05:25
-------------------------------------------------------------------------------
Number of entries : 1
===============================================================================
```

The svc-ID belonging to the Service-ID configured on the VSD can be easily obtained:

```
*A:pe-9# show service vsd domain "L2-service-1" association
============================================================
Service VSD Domain
============================================================
Svc Id      Svc Type  Domain Type    Domain Admin   Origin
------------------------------------------------------------
64000       vpls      l2Domain       inService      vsd
------------------------------------------------------------
Number of entries: 1
```

The associated RT/RD/VNI information can then be displayed with the **show service id** <*service-id*> **all** command, as shown previously.

An overview of the VTEPs that the DC Gateway shares this service with, and the corresponding VNIs is available:

```
*A:pe-9# show service id 64000 vxlan
===============================================================================
VPLS VXLAN, Ingress VXLAN Network Id: 1006636
===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address           Egress VNI     Num. MACs    Mcast   Oper State   L2 PBR
-------------------------------------------------------------------------------
39.0.0.94              1006636        0            Yes     Up           No
-------------------------------------------------------------------------------
```

An overview of all the Service-IDs and associated VNIs that the DC Gateway has in common with a VSC or VSG can be shown with the following command:

```
*A:pe-9# show service vxlan 39.0.0.94
===============================================================================
VXLAN Tunnel Endpoint: 39.0.0.94
===============================================================================
Egress VNI                    Service Id               Oper State
-------------------------------------------------------------------------------
1006636                       64000                    Up
-------------------------------------------------------------------------------
===============================================================================
```

Relevant MAC/IP/VNI/RT/NH information is also in the EVPN BGP RIB:

```
*A:pe-9# show router bgp routes evpn mac
===============================================================================
 BGP Router ID:10.0.0.9         AS:65000       Local AS:65000
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Ip Address
                                            NextHop
                                            Label1
-------------------------------------------------------------------------------
u*>i  65534:39790         1e:50:01:01:00:01 ESI-0
      1006636             Static            10.32.78.100
                                            39.0.0.94
                                            VNI 1006636
-------------------------------------------------------------------------------
```

The WAN service can be detached from the L2 domain in the VSD GUI, if required.

This triggers similar debug information on the DC Gateway as the **tools perform service vsd evaluate teardown** command shown in the "Testing the python script" section.

After deleting the WAN service in the VSD GUI, the VPLS service and the service domain is removed from the DC Gateway.

L2 VXLAN IRB

An example python script for F-D XMPP provisioning of an L2 VXLAN IRB type service (l2-domain-irb) is as follows:

```
from alc import dyn
# example of metadata to be added in VSD WAN Service: "rd=2:2,sap=1/1/1:1000,vprnAS=
65000,vprnRD=65000:1,vprnRT=target:65000:1,vprnLo=1.1.1.1,irbGW=10.32.78.1/24"
```

```
# example of tools cli to test this script: tools perform service vsd evaluate-script
 domain-name "l2domIRB1" type l2-domain-irb action setup policy "py-l2-irb" vni  1234 rt-i
 target:2:2 rt-e target:2:2 metadata "rd=2:2,sap=1/1/1:1000,vprnAS=65000,vprnRD=65000:1,vprnRT=
target:65000:1,vprnLo=1.1.1.1,irbGW=10.32.78.1/24"
# teardown example cli: tools perform service vsd evaluate-script domain-name "l2domIRB1" type
 l2-domain-irb action teardown policy "py-l2-irb" vni 1234 rt-i target:2:2  rt-e target:2:2
def setup_script(vsdParams):

    print ("These are the VSD params: " + str(vsdParams))
    servicetype = vsdParams['servicetype']
    vni = vsdParams['vni']
    rt = vsdParams['rt']
# add "target:" if provisioned by VSD (VSD uses x:x format whereas tools command uses
 target:x:x format)
    if not rt.startswith ('target'):
       rt = "target:"+rt
    metadata = vsdParams['metadata']
# remove trailing space at the end of the metadata
    metadata = metadata.rstrip()
    print ("VSD metadata" + str(metadata))
    metadata = dict(e.split('=') for e in metadata.split(','))
    print ("Modified metadata" + str(metadata))
    vplsSvc_id = dyn.select_free_id("service-id")
    vprnSvc_id = dyn.select_free_id("service-id")
    print ("this are the free svc ids picked up by the system: VPLS:" + vplsSvc_id + " + VPRN:"
 + vprnSvc_id)
    if servicetype == "L2DOMAIN-IRB":

      rd = metadata['rd']
      sap_id = metadata['sap']
      vprn_AS = metadata ['vprnAS']
      vprn_RD = metadata ['vprnRD']
      vprn_RT = metadata ['vprnRT']
      vprn_Lo = metadata ['vprnLo']
      irb_GW = metadata ['irbGW']
      print ('servicetype, VPLS id, rt, vni, rd, sap, VPRN id, vprn_AS, vprn_RD, vprn_RT, vprn_
Lo, irb_GW:', servicetype, vplsSvc_id, rt, vni, rd, sap_id, vprnSvc_id, vprn_AS, vprn_RD, vprn_
RT, vprn_Lo, irb_GW)
      dyn.add_cli("""
        configure service
           vpls %(vplsSvc_id)s customer 1 create
              allow-ip-int-bind
              exit
              description vpls%(vplsSvc_id)s
              bgp
                 route-distinguisher %(rd)s
                 route-target %(rt)s
              exit
              vxlan vni %(vni)s create
              exit
              bgp-evpn
                  evi %(vplsSvc_id)s
                  vxlan
                      no shut
                  exit
              exit
              service-name vpls%(vplsSvc_id)s
              sap %(sap_id)s create
              exit
              no shutdown
              exit
             exit
        exit
        configure service
```

```
           vprn %(vprnSvc_id)s customer 1 create
              autonomous-system %(vprn_AS)s
              route-distinguisher %(vprn_RD)s
              vrf-target %(vprn_RT)s
              interface "irbvpls-%(vplsSvc_id)s" create
                 address %(irb_GW)s
                 vpls "vpls%(vplsSvc_id)s"
                 exit
              exit
              interface "lo1" create
                 address %(vprn_Lo)s/32
                 loopback
              exit
              no shutdown
           exit
        exit
      """ % {'vplsSvc_id' : vplsSvc_id, 'vprnSvc_id' : vprnSvc_id, 'vni' : vsdParams['vni'],
 'rt' : rt, 'rd' : metadata['rd'], 'sap_id' : sap_id, 'vprn_AS' : vprn_AS, 'vprn_RD' : vprn_RD,
 'vprn_RT' : vprn_RT, 'vprn_Lo' : vprn_Lo, 'irb_GW' : irb_GW})
      # L2DOMAIN-IRB returns setupParams: vplsSvc_id, vprnSvc_id, servicetype, vni, sap, vprn_
AS, vprn_RD, vprn_RT, vprn_Lo
      return {'vplsSvc_id' : vplsSvc_id, 'vprnSvc_id' : vprnSvc_id, 'servicetype' :
 servicetype, 'vni' : vni, 'sap_id' : sap_id, 'vprn_AS' : vprn_AS, 'vprn_RD' : vprn_RD, 'vprn_
RT' : vprn_RT, 'vprn_Lo' : vprn_Lo, 'irb_GW': irb_GW}
#----------------------------------------------------------------------------------------------
--
def teardown_script(setupParams):
    print ("These are the teardown_script setupParams: " + str(setupParams))
    servicetype = setupParams['servicetype']
    if servicetype == "L2DOMAIN-IRB":
      dyn.add_cli("""
        configure service
            vpls %(vplsSvc_id)s
               no description
               bgp-evpn
                   vxlan
                       shut
                   exit
                   no evi
                   exit
               no vxlan vni %(vni)s
               bgp
                  no route-distinguisher
                  no route-target
               exit
               no bgp
               no bgp-evpn
               sap %(sap_id)s
                  shutdown
                  exit
               no sap %(sap_id)s
               shutdown
               exit
            no vpls %(vplsSvc_id)s
            vprn %(vprnSvc_id)s
               interface lo1 shutdown
               no interface lo1
               interface "irbvpls-%(vplsSvc_id)s"
                  no vpls
                  shutdown
                  exit
               no interface "irbvpls-%(vplsSvc_id)s"
               shutdown
            exit
        no vprn %(vprnSvc_id)s
        exit
      """ % {'vplsSvc_id' : setupParams['vplsSvc_id'], 'vprnSvc_id' : setupParams['vprnSvc_
id'], 'vni' : setupParams['vni'], 'sap_id' : setupParams['sap_id']})
      return setupParams
d = {"script" : (setup_script, None, None, teardown_script)}
dyn.action(d)
```

The python script and policy are configured in a similar way as the previous example:

```
*A:pe-9# configure python
*A:pe-9>config>python# info
----------------------------------------------
        python-script "l2domain-irb_services" create
            primary-url "ftp://*:*@138.203.15.48/./l2domainIRB_service.py"
            no shutdown
        exit
        python-policy "py-l2-irb" create
            description "Python script to create L2-IRB domains"
            vsd script "l2domain-irb_services"
        exit
----------------------------------------------
```

It is also possible to create a python script that covers different domain types. The relevant part of the script is then addressed by using the following if-statement in the script:

```
    servicetype = vsdParams.get('servicetype')
        if servicetype == "L2DOMAIN-IRB":
           ---snip---
```

On the VSD, the following has to be provided:

(screenshots of this workflow are available in the VSP User Guide)

Create an L2 WAN service in the VSD:

- under Platform Config/Infrastructure, select the DC Gateway to add a WAN service
- select Service Type "layer 2" and select "IRB"
- select "Dynamic" configuration type to allow Fully Dynamic provisioning
- under Service Policy, provide the python policy configured on the DC Gateway
- provide a Name and Service-ID (the Service-ID will be the name of the dynamically created service domain on the DC Gateway)

Add the metadata to the WAN service:

- right-click the WAN service and select "inspect"
- a pop-up dialog box appears; select the "Metadata" tag and add the metadata info; for example,

  "rd=2:2,sap=1/1/1:1000,vprnAS=65000,vprnRD=65000:1,vprnRT=target:65000:1,vprnLo=1.1.1.1, irbGW=10.32.78.1/24"

Add permissions to Enterprise1. The WAN service should now be visible in Enterprise1.

Add permissions for a group of users to use this WAN service. Instantiate an L2 domain and attach the WAN service.

After the script has completed, there should be two new services created:

```
*A:pe-9# show service service-using
```

```
===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
64000        VPLS      Up   Up   1          vpls64000
64001        VPRN      Up   Up   1
```

```
*A:pe-9# show service id 64000 all
===============================================================================
Service Detailed Information
===============================================================================
Service Id        : 64000               Vpn Id            : 0
Service Type      : VPLS
Name              : vpls64000
Description       : vpls64000
Customer Id       : 1                   Creation Origin   : vsd
Last Status Change: 07/22/2015 11:15:36
Last Mgmt Change  : 07/22/2015 11:15:36
Etree Mode        : Disabled
Admin State       : Up                  Oper State        : Up
MTU               : 1514                Def. Mesh VC Id   : 64000
SAP Count         : 1                   SDP Bind Count    : 0
---snip---
VSD Domain        : L2-IRB-Service-1
---snip---

-------------------------------------------------------------------------------
BGP Information
-------------------------------------------------------------------------------
Vsi-Import        : None
Vsi-Export        : None
Route Dist        : 2:2
Oper Route Dist   : 2:2
Oper RD Type      : configured
Rte-Target Import : 65534:6985          Rte-Target Export : 65534:6985
Oper RT Imp Origin: configured          Oper RT Import    : 65534:6985
Oper RT Exp Origin: configured          Oper RT Export    : 65534:6985
PW-Template Id    : None
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
SAP 1/1/1:1000
-------------------------------------------------------------------------------
Service Id         : 64000
SAP                : 1/1/1:1000               Encap             : q-tag
Description        : (Not Specified)
Admin State        : Up                       Oper State        : Up
---snip---
===============================================================================
VPLS VXLAN, Ingress VXLAN Network Id: 1006636
===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address           Egress VNI     Num. MACs    Mcast   Oper State   L2 PBR
-------------------------------------------------------------------------------
39.0.0.94              1006636        0            Yes     Up           No
-------------------------------------------------------------------------------
---snip---
```

```
*A:pe-9# show service id 64001 all
===============================================================================
Service Detailed Information
===============================================================================
Service Id        : 64001               Vpn Id            : 0
Service Type      : VPRN
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : vsd
Last Status Change: 07/22/2015 11:15:36
Last Mgmt Change  : 07/22/2015 11:15:36
Admin State       : Up                  Oper State        : Up

Route Dist.       : 65000:1             VPRN Type         : regular
Oper Route Dist   : 65000:1
Oper RD Type      : configured
AS Number         : 65000               Router Id         : 10.0.0.9
ECMP              : Enabled             ECMP Max Routes   : 1
---snip---
-------------------------------------------------------------------------------
Interface
-------------------------------------------------------------------------------
If Name           : irbvpls-64000
Admin State       : Up                  Oper (v4/v6)      : Up/Down
Protocols         : None
IP Addr/mask      : 10.32.78.1/24       Address Type      : Primary
---snip---
Routed VPLS Details
VPLS Name        : vpls64000
Binding Status   : Up
---snip---
-------------------------------------------------------------------------------
Interface
-------------------------------------------------------------------------------
If Name           : lo1
Admin State       : Up                  Oper (v4/v6)      : Up/Down
Protocols         : None
IP Addr/mask      : 1.1.1.1/32          Address Type      : Primary
```

The dynamically created configuration can be inspected in enable-vsd-config mode (only enter the enable-vsd-config mode when absolutely required):

```
*A:pe-9>config>service# info
----------------------------------------------
        customer 1 create
            description "Default customer"
        exit
        vsd
            domain L2-IRB-Service-1 type l2-domain-irb create
                description "L2-IRB-Service-1"
                no shutdown
            exit
            service-range 64000 to 64999
        exit
        vpls 64000 customer 1 create
            description "vpls64000"
            allow-ip-int-bind
            exit
            vxlan vni 1006636 create
            exit
            bgp
                route-distinguisher 2:2
            exit
            bgp-evpn
                evi 64000
                vxlan
                    no shutdown
                exit
                mpls
                    shutdown
                exit
            exit
            stp
                shutdown
            exit
            service-name "vpls64000"
            sap 1/1/1:1000 create
            exit
            vsd-domain "L2-IRB-Service-1"
            no shutdown
        exit
        vprn 64001 customer 1 create
            autonomous-system 65000
            route-distinguisher 65000:1
            auto-bind-tunnel
                resolution any
            exit
            vrf-target target:65000:1
            interface "irbvpls-64000" create
                address 10.32.78.1/24
                vpls "vpls64000"
                exit
            exit
            interface "lo1" create
                address 1.1.1.1/32
                loopback
            exit
            vsd-domain "L2-IRB-Service-1"
            no shutdown
        exit
----------------------------------------------
```

MAC addresses can now be learned in the Nuage VSP/VSG and sent via MP-BGP to the DC Gateway:

```
*A:pe-9# show service id 64000 fdb detail
===============================================================================
Forwarding Database, Service 64000
===============================================================================
ServId    MAC               Source-Identifier        Type     Last Change
                                                     Age
-------------------------------------------------------------------------------
64000     1e:50:01:01:00:01 vxlan:                   Evpn     07/22/15 11:26:15
                            39.0.0.94:1006636
64000     1e:e2:ff:00:f9:3d cpm                      Intf     07/22/15 11:15:36
-------------------------------------------------------------------------------
```

The second MAC address is the GW-MAC that the VM or VSG-connected host will use to reach the interface on the VPRN:

```
*A:pe-9# show router 64001 route-table
===============================================================================
Route Table (Service: 64001)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
1.1.1.1/32                                    Local   Local     00h17m10s  0
```

```
       lo1                                                       0
10.32.78.0/24                                 Local   Local     00h17m10s  0
       irbvpls-64000                                             0
```

```
*A:pe-9# show router 64001 arp
===============================================================================
ARP Table (Service: 64001)
===============================================================================
IP Address      MAC Address       Expiry    Type   Interface
-------------------------------------------------------------------------------
10.32.78.1      1e:e2:ff:00:f9:3d 00h00m00s Oth[I] irbvpls-64000
10.32.78.100    1e:50:01:01:00:01 03h53m22s Dyn[I] irbvpls-64000
1.1.1.1         1e:e2:ff:00:00:00 00h00m00s Oth    lo1
```

Similar commands as shown in the previous section are available to obtain relevant information:

- **show service vsd domain "L2-IRB-Service-1" association** to obtain svc-IDs
- **show service id <id> all** to obtain RT/RD/VNI values
- **show service id <id> vxlan** to obtain VTEPs in the VPLS service
- **show service vxlan <vtep-ip>** to obtain svc-ID and VNI information
- **show router bgp routes evpn mac** to obtain MAC/IP/VNI/RT/NH information

The VM or VSG-connected Host should be able to ping the loopback interface of the VPRN service:

```
*A:ce1# ping 1.1.1.1 source 10.32.78.100
PING 1.1.1.1 56 data bytes
64 bytes from 1.1.1.1: icmp_seq=1 ttl=64 time=1.67ms.
64 bytes from 1.1.1.1: icmp_seq=2 ttl=64 time=1.83ms.
```

L3 VXLAN

An example python script for F-D XMPP provisioning of an L3 VXLAN type service (vrf-vxlan) is:

```
from alc import dyn

# example of metadata to be added in VSD WAN Service: "rd=3:3,vprnAS=65000,vprnRD=
65000:1,vprnRT=target:65000:1,vprnLo=1.1.1.1"

# example of tools cli to test this script: tools perform service vsd evaluate-script domain-
name "l3dom1" type vrf-vxlan action setup policy "py-vrf-vxlan" vni 1234 rt-i target:3:3 rt-e
 target:3:3 metadata "rd=3:3,vprnAS=65000,vprnRD=65000:1,vprnRT=target:65000:1,vprnLo=1.1.1.1"

# teardown example cli: tools perform service vsd evaluate-script domain-name "l3dom1" type
 vrf-vxlan action teardown policy "py-vrf-vxlan" vni 1234 rt-i target:3:3 rt-e target:3:3

def setup_script(vsdParams):

    print ("These are the VSD params: " + str(vsdParams))
    servicetype = vsdParams['servicetype']
    vni = vsdParams['vni']
    rt = vsdParams['rt']

# add "target:" if provisioned by VSD (VSD uses x:x format whereas tools command uses
 target:x:x format)
    if not rt.startswith ('target'):
       rt = "target:"+rt
    metadata = vsdParams['metadata']
```

```
# remove trailing space at the end of the metadata
    metadata = metadata.rstrip()
    print ("VSD metadata" + str(metadata))
    metadata = dict(e.split('=') for e in metadata.split(','))
    print ("Modified metadata" + str(metadata))
    vplsSvc_id = dyn.select_free_id("service-id")
    vprnSvc_id = dyn.select_free_id("service-id")
    print ("this are the free svc ids picked up by the system: VPLS:" + vplsSvc_id + " + VPRN:"
 + vprnSvc_id)

    if servicetype == "VRF-VXLAN":

      rd = metadata['rd']
      vprn_AS = metadata ['vprnAS']
      vprn_RD = metadata ['vprnRD']
      vprn_RT = metadata ['vprnRT']
      vprn_Lo = metadata ['vprnLo']
      print ('servicetype, VPLS id, rt, vni, rd, VPRN id, vprn_AS, vprn_RD, vprn_RT, vprn_Lo:',
 servicetype, vplsSvc_id, rt, vni, rd, vprnSvc_id, vprn_AS, vprn_RD, vprn_RT, vprn_Lo)
      dyn.add_cli("""
        configure router policy-options
           begin
             community _VSD_%(vplsSvc_id)s members %(rt)s
             policy-statement vsi_import_%(vplsSvc_id)s
                entry 10
                   from
                      family evpn
                      community _VSD_%(vplsSvc_id)s
                      exit
                   action accept
                   exit
                exit
             exit
             policy-statement vsi_export_%(vplsSvc_id)s
                entry 10
                   from
                      family evpn
                      exit
                      action accept
                      community add _VSD_%(vplsSvc_id)s
                      exit
                   exit
              exit
           commit
        exit

        configure service
           vpls %(vplsSvc_id)s customer 1 create
              allow-ip-int-bind
                 exit
              description vpls%(vplsSvc_id)s
              bgp
                 route-distinguisher %(rd)s
                 vsi-import vsi_import_%(vplsSvc_id)s
                 vsi-export vsi_export_%(vplsSvc_id)s
                 exit
              vxlan vni %(vni)s create
                 exit
              bgp-evpn
                 ip-route-advertisement
                 vxlan
                     no shut
                     exit
                 exit
              service-name vpls%(vplsSvc_id)s
              no shutdown
              exit
            exit
        exit

        configure service
           vprn %(vprnSvc_id)s customer 1 create
              autonomous-system %(vprn_AS)s
              route-distinguisher %(vprn_RD)s
              vrf-target %(vprn_RT)s
              interface "vpls-%(vplsSvc_id)s" create
                 vpls "vpls%(vplsSvc_id)s" evpn-tunnel
                 exit
              interface "lo1" create
                 address %(vprn_Lo)s/32
                 loopback
                 exit
              no shutdown
              exit
        exit

      """ % {'vplsSvc_id' : vplsSvc_id, 'vprnSvc_id' : vprnSvc_id, 'vni' : vsdParams['vni'],
 'rt' : rt, 'rd' : metadata['rd'], 'vprn_AS' : vprn_AS, 'vprn_RD' : vprn_RD, 'vprn_RT' : vprn_
RT, 'vprn_Lo' : vprn_Lo})
      # VRF-VXLAN returns setupParams: vplsSvc_id, vprnSvc_id, servicetype, vni, vprn_AS, vprn_
RD, vprn_RT, vprn_Lo
      return {'vplsSvc_id' : vplsSvc_id, 'vprnSvc_id' : vprnSvc_id, 'servicetype' :
 servicetype, 'vni' : vni, 'vprn_AS' : vprn_AS, 'vprn_RD' : vprn_RD, 'vprn_RT' : vprn_RT,
 'vprn_Lo' : vprn_Lo}
#----------------------------------------------------------------------------------------------
--

def teardown_script(setupParams):
    print ("These are the teardown_script setupParams: " + str(setupParams))
    servicetype = setupParams['servicetype']
    if servicetype == "VRF-VXLAN":
      dyn.add_cli("""
        configure service
            vpls %(vplsSvc_id)s
               no description
               bgp-evpn
                   vxlan
                       shut
                       exit
                   no evi
                   exit
               no vxlan vni %(vni)s
               bgp
                  no route-distinguisher
                  no route-target
                  exit
               no bgp
               no bgp-evpn
               shutdown
               exit
            no vpls %(vplsSvc_id)s
            vprn %(vprnSvc_id)s
               interface lo1 shutdown
               no interface lo1
               interface "vpls-%(vplsSvc_id)s"
                  vpls "vpls%(vplsSvc_id)s"
                     no evpn-tunnel
                     exit
                  no vpls
                  shutdown
                  exit
               no interface "vpls-%(vplsSvc_id)s"
               shutdown
            exit
        no vprn %(vprnSvc_id)s
        exit
        configure router policy-options
            begin
            no community _VSD_%(vplsSvc_id)s
            no policy-statement vsi_import_%(vplsSvc_id)s
            no policy-statement vsi_export_%(vplsSvc_id)s
            commit
        exit

      """ % {'vplsSvc_id' : setupParams['vplsSvc_id'], 'vprnSvc_id' : setupParams['vprnSvc_
id'], 'vni' : setupParams['vni']})
      return setupParams

d = {"script" : (setup_script, None, None, teardown_script)}

dyn.action(d)
```

The python script and policy are configured in a similar way as the previous example:

```
*A:pe-9# configure python
*A:pe-9>config>python# info
----------------------------------------------
        python-script "vrf-vxlan_services" create
            primary-url "ftp://*:*@138.203.15.48/./vrf-vxlan_service.py"
            no shutdown
        exit
        python-policy "py-vrf-vxlan" create
            description "Python script to create vrf-vxlan domains"
            vsd script "l3vxlan"
        exit
----------------------------------------------
```

The following steps are required on the VSD to provision the WAN service:

(screenshots of this workflow are available in the VSP User Guide)

Create an L3 WAN service in the VSD:

- under Platform Config / Infrastructure, select the DC Gateway to add a WAN service
- select Service Type "layer 3"
- select "Dynamic" configuration type to allow Fully Dynamic provisioning
- under Service Policy, provide the python policy configured on the DC Gateway
- provide a Name and Service-ID (the Service-ID will be the name of the dynamically created service domain on the DC Gateway)

Add the metadata to the WAN service:

- right-click the WAN service and select "inspect"
- a pop-up dialog box appears; select the "Metadata" tag and add the metadata info; for example, "rd=3:3,vprnAS=65000,vprnRD=65000:1,vprnRT=target:65000:1,vprnLo=1.1.1.1"

Add permissions to Enterprise1. The WAN service should now be visible in Enterprise1.

Add permissions for a group of users to use this WAN service. Instantiate an L3 domain and attach the WAN service.

After the script has completed, there should be two new services:

```
*A:pe-9# show service service-using
===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
64000        VPLS      Up   Up   1          vpls64000
64001        VPRN      Up   Up   1
```

```
*A:pe-9# show service id 64000 all
===============================================================================
Service Detailed Information
===============================================================================
Service Id        : 64000               Vpn Id            : 0
Service Type      : VPLS
Name              : vpls64000
Description       : vpls64000
Customer Id       : 1                   Creation Origin   : vsd
Last Status Change: 10/14/2015 18:38:44
Last Mgmt Change  : 10/14/2015 18:38:44
Etree Mode        : Disabled
Admin State       : Up                  Oper State        : Up
MTU               : 1514                Def. Mesh VC Id   : 64000
SAP Count         : 0                   SDP Bind Count    : 0
---snip---
VSD Domain        : L3-service-1
---snip---

-------------------------------------------------------------------------------
BGP Information
-------------------------------------------------------------------------------
Vsi-Import        : vsi_import_64000
Vsi-Export        : vsi_export_64000
Route Dist        : 3:3
Oper Route Dist   : 3:3
Oper RD Type      : configured
Rte-Target Import : None                Rte-Target Export : None
Oper RT Imp Origin: vsi                 Oper RT Import    : None
Oper RT Exp Origin: vsi                 Oper RT Export    : None
PW-Template Id    : None
-------------------------------------------------------------------------------
---snip---
===============================================================================
VPLS VXLAN, Ingress VXLAN Network Id: 119281

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address           Egress VNI     Num. MACs    Mcast   Oper State   L2 PBR
-------------------------------------------------------------------------------
39.0.0.94              119281         1            No      Up           No
-------------------------------------------------------------------------------
---snip---
```

The script dynamically creates a VSI-import and VSI-export policy and links it to an RT that was dynamically created by the VSC/VSG:

```
*A:pe-9# show router policy
===============================================================================
Route Policies
===============================================================================
Policy                            Description
-------------------------------------------------------------------------------
vsi_export_64000
vsi_import_64000
-------------------------------------------------------------------------------
Policies : 2
===============================================================================
```

```
*A:pe-9# show router policy "vsi_import_64000"
    entry 10
        from
            community "_VSD_64000"
            family evpn
        exit
        action accept
        exit
    exit
```

```
*A:pe-9# show router policy "vsi_export_64000"
    entry 10
        from
            family evpn
        exit
        action accept
            community add "_VSD_64000"
        exit
    exit
```

```
*A:pe-9# show router policy community "_VSD_64000"
community "_VSD_64000" members "65534:38619"
```

```
*A:pe-9# show service id 64001 all
===============================================================================
Service Detailed Information
===============================================================================
Service Id        : 64001               Vpn Id            : 0
Service Type      : VPRN
Name              : (Not Specified)
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : vsd
Last Status Change: 10/14/2015 18:38:44
Last Mgmt Change  : 10/14/2015 18:38:44
Admin State       : Up                  Oper State        : Up

Route Dist.       : 65000:1             VPRN Type         : regular
Oper Route Dist   : 65000:1
Oper RD Type      : configured
AS Number         : 65000               Router Id         : 10.0.0.9
ECMP              : Enabled             ECMP Max Routes   : 1
Auto Bind Tunnel
Resolution        : any
---snip---
Vrf Target        : target:65000:1
```

```
---snip---
-------------------------------------------------------------------------------
Interface
-------------------------------------------------------------------------------
If Name           : vpls-64000
Admin State       : Up                  Oper (v4/v6)      : Up/Down
Protocols         : None

IP Addr/mask      : Not Assigned
---snip---
Routed VPLS Details
VPLS Name        : vpls64000
Binding Status   : Up
---snip---
-------------------------------------------------------------------------------
Interface
-------------------------------------------------------------------------------
If Name           : lo1
Admin State       : Up                  Oper (v4/v6)      : Up/Down
Protocols         : None
IP Addr/mask      : 1.1.1.1/32          Address Type      : Primary
```

The dynamically created configuration can be inspected in enable-vsd-config mode (only enter the enable-vsd-config mode when required):

```
*A:pe-9>config>service# info
----------------------------------------------
        customer 1 create
            description "Default customer"
        exit
        vsd
            domain L3-service-1 type vrf-vxlan create
                description "L3-service-1"
                no shutdown
            exit
            service-range 64000 to 64999
        exit
        vpls 64000 customer 1 create
            description "vpls64000"
            allow-ip-int-bind
            exit
            vxlan vni 119281 create
            exit
            bgp
                route-distinguisher 3:3
                vsi-export "vsi_export_64000"
                vsi-import "vsi_import_64000"
            exit
            bgp-evpn
                ip-route-advertisement
                vxlan
                    no shutdown
                exit
                mpls
                    shutdown
                exit
            exit
            stp
                shutdown
            exit
            service-name "vpls64000"
            vsd-domain "L3-service-1"
            no shutdown
        exit
        vprn 64001 customer 1 create
            autonomous-system 65000
            route-distinguisher 65000:1
            auto-bind-tunnel
                resolution any
            exit
            vrf-target target:65000:1
            interface "vpls-64000" create
                vpls "vpls64000"
                    evpn-tunnel
                exit
            exit
            interface "lo1" create
                address 1.1.1.1/32
                loopback
            exit
            vsd-domain "L3-service-1"
            no shutdown
        exit
----------------------------------------------
```

The EVPN tunnel NH-MAC addresses can now be learned in the Nuage VSP/VSG and sent via MP-BGP to the DC Gateway:

```
*A:pe-9# show service id 64000 fdb detail
Forwarding Database, Service 64000
===============================================================================
ServId    MAC               Source-Identifier        Type     Last Change
                                                     Age
-------------------------------------------------------------------------------
64000     00:00:27:00:00:5e vxlan:                   Evpn     07/22/15 13:38:47
                            39.0.0.94:119281
64000     1e:e2:ff:00:f9:3d cpm                      Intf     07/22/15 13:38:44
-------------------------------------------------------------------------------
```

The first MAC entry address is the tunnel NH-MAC for the VSG and the second is the address for the DC Gateway for EVPN-tunnel service 64000:

```
*A:pe-9# show router 64001 route-table
===============================================================================
1.1.1.1/32                                    Local   Local     00h09m08s  0
       lo1                                                          0
10.32.78.0/24                                 Remote  BGP EVPN  00h09m05s  169
       vpls-64000 (ET-00:00:27:00:00:5e)                            0
10.32.78.100/32                               Remote  BGP EVPN  00h00m04s  169
       vpls-64000 (ET-00:00:27:00:00:5e)                            0
-------------------------------------------------------------------------------
```

The following commands are useful to obtain relevant information:

- **show service vsd domain "L3-service-1" association** to obtain svc-IDs
- **show service id <id> all** to obtain RT/RD/VNI values
- **show service id <id> vxlan** to obtain VTEPs in the VPLS service
- **show service vxlan <vtep-ip>** to obtain svc-id and VNI information

Relevant EVPN tunnel NH-MAC/VNI/RT/NH information is also in the EVPN BGP RIB:

```
*A:pe-9# show router bgp routes evpn mac detail
```

```
---snip---

Modified Attributes

Network        : N/A
Nexthop        : 39.0.0.94
From           : 39.0.0.94
Res. Nexthop   : 192.168.39.94
Local Pref.    : 200                    Interface Name : toNuage
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
Community      : target:65534:38619 bgp-tunnel-encap:VXLAN
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 39.0.0.94
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : MAC
ESI            : ESI-0
Tag            : 119281
IP Address     : N/A
Route Dist.    : 65534:29625
Mac Address    : 00:00:27:00:00:5e
MPLS Label1    : VNI 119281             MPLS Label2    : N/A
Route Tag      : 0
---snip---
```

VM/VSG-connected Host and network information is also in the EVPN BGP RIB:

```
*A:pe-9# show router bgp routes evpn ip-prefix detail
---snip---
Modified Attributes

Network        : N/A
Nexthop        : 39.0.0.94
From           : 39.0.0.94
Res. Nexthop   : 192.168.39.94
Local Pref.    : 200                    Interface Name : toNuage
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None
Connector      : None
Community      : target:65534:38619 ext:30b:220000000000
                 ext:30b:100b00b00000 bgp-tunnel-encap:VXLAN
                 mac-nh:00:00:27:00:00:5e
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 39.0.0.94
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : IP-PREFIX
ESI            : N/A
Tag            : 119281
Gateway Address: 00:00:27:00:00:5e
Prefix         : 10.32.78.100/32
Route Dist.    : 39.0.0.94:10269
MPLS Label     : VNI 119281
Route Tag      : 0
---snip---
```

```
Modified Attributes

Network        : N/A
Nexthop        : 39.0.0.94
From           : 39.0.0.94
Res. Nexthop   : 192.168.39.94
Local Pref.    : 200                   Interface Name : toNuage
Aggregator AS  : None                  Aggregator     : None
Atomic Aggr.   : Not Atomic            MED            : 0
AIGP Metric    : None
Connector      : None
Community      : target:65534:38619 ext:30b:220000000000
                 ext:30b:100b00b00000 bgp-tunnel-encap:VXLAN
                 mac-nh:00:00:27:00:00:5e
Cluster        : No Cluster Members
Originator Id  : None                  Peer Router Id : 39.0.0.94
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : IP-PREFIX
ESI            : N/A
Tag            : 119281
Gateway Address: 00:00:27:00:00:5e
Prefix         : 10.32.78.0/24
Route Dist.    : 39.0.0.94:10269
MPLS Label     : VNI 119281
Route Tag      : 0
---snip---
```

The VM or VSG-connected Host should be able to ping the loopback interface of the VPRN service:

```
*A:ce1# ping 1.1.1.1 source 10.32.78.100
PING 1.1.1.1 56 data bytes
64 bytes from 1.1.1.1: icmp_seq=1 ttl=64 time=1.67ms.
64 bytes from 1.1.1.1: icmp_seq=2 ttl=64 time=1.83ms.
```

Troubleshooting and debug commands

When testing/troubleshooting F-D XMPP provisioning, the following show/tools/debug commands can be useful:

- tools perform service vsd evaluate-script
- tools perform service vsd fd-domain-sync <full|diff>
- tools perform service vsd domain refresh-config
- tools perform python-script reload
- tools dump service vsd-services command-list
- debug python python-script
- debug vsd scripts event/instance
- debug system xmpp
- debug router bgp update
- show service vsd domain
- show service vsd script

- show service vsd summary
- show system vsd
- show xmpp vsd
- show python python-policy <name> {association}
- show python python-script <name> {association|source-in-use}
- show service vxlan [<vtep-ip>]
- show service service-using {<service-type>}
- show service id route-table
- show service id fdb detail
- show service id proxy-arp detail
- show router [<router-instance>] route-table
- show router [<router-instance>] arp
- show router bgp routes bgp <mac|ip-prefix|inclusive-mcast> {detail}
- log-id 99

Conclusion

The fully dynamic VSD integration model allows for automated provisioning of breakout services on the SR OS DC Gateway. Different domain types (l2-domain/l2-domain-irb/vrf-vxlan/vrf-gre) are supported. This chapter has shown how to construct, load, and test python scripts for this feature. It also described how to configure WAN services on the VSD and how to verify the dynamically created services.

Inter-AS Model C for VLL

This chapter describes advanced inter-AS model C for Virtual Leased Line (VLL) configurations.

Topics in this chapter include:

Applicability

This chapter was initially written for SR OS Release 8.0.R4. The CLI in the current edition corresponds to SR OS Release 20.10.R2.

Overview

SR OS supports RFC 3107, *Carrying Label Information in BGP-4*, including VLL/VPLS. BGP SDPs can also be used with PBB-VPLS services.

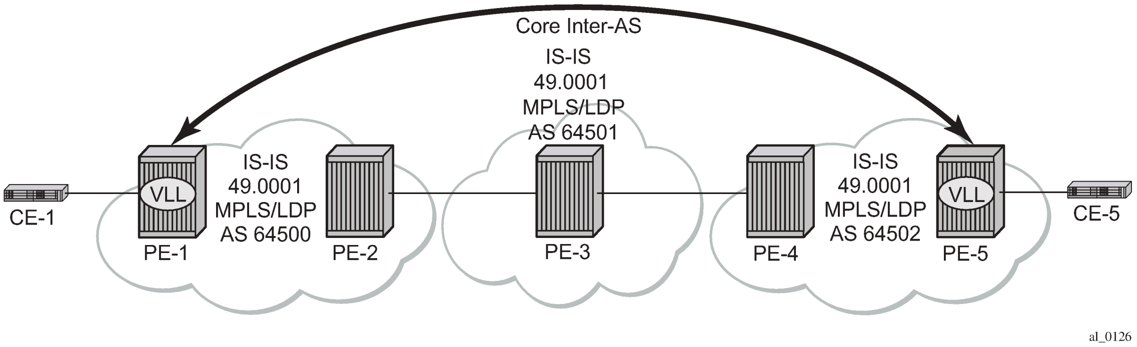
Internet service providers are looking for mechanisms to implement the VLL and VPLS services across Autonomous Systems (ASs). Service providers may have inter-AS operation as a consequence of delivering inter-provider VLL/VPLS or because they use multiple ASs as a result of acquisitions and mergers.

The objective of this chapter is to describe the interconnection of VLL services across multiple ASs, using inter-AS model C. Inter-AS Model C involves eBGP redistribution of internal system addresses to the neighboring AS using labeled IPv4 routes.

Example topology

Figure 163: Example topology – Inter-AS model C for VLL shows the example topology used for Inter-AS Model C VLL.

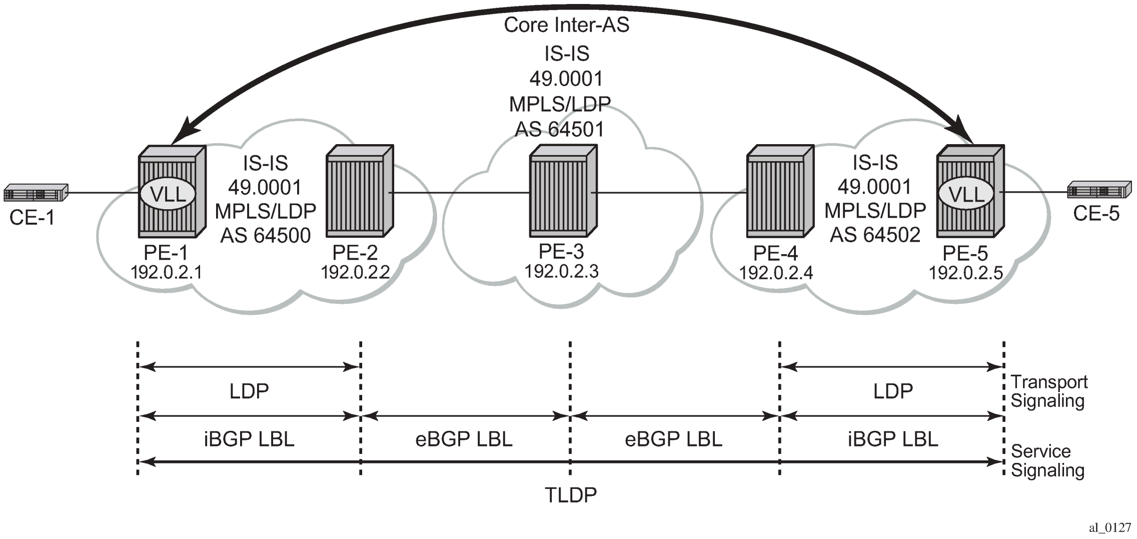
Figure 163: Example topology – Inter-AS model C for VLL



The example topology shown in Figure 163: Example topology – Inter-AS model C for VLL consists of three sites in different ASs with each site using SR OS nodes.

AS 64500 contains PE-1 and PE-2, AS 64501 contains PE-3, and AS 64502 contains PE-4 and PE-5. There is a business customer with two remote locations, Site A and Site B, with Customer Edge (CE) devices CE-1 connected to the AS 64500 via PE-1 and CE-5 connected to the AS 64502 via PE-5. A VLL Epipe service is configured between PE-1 and PE-5 to connect site A and site B.

Figure 164: Inter-AS model C for VLL



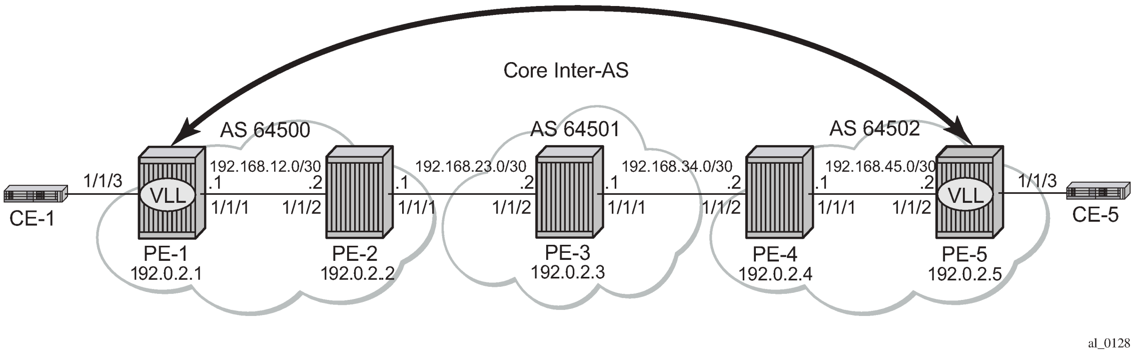
Configuration

This section describes all of the relevant configuration tasks for the detailed setup shown in Figure 165: Network setup configuration. In this particular example, the following protocols are assumed to be already configured.

- IS-IS as the IGP with all the nodes being level Level1/Level 2.

- LDP as the MPLS protocol to signal the transport tunnels within AS 64500 and AS 64502.

Figure 165: Network setup configuration



BGP configuration

A BGP tunnel must be established between PE-1 and PE-5, therefore, labeled BGP routes must be exchanged for prefixes 192.0.2.1/32 and 192.0.2.5/32 across the ASs. The following shows the BGP configuration — iBGP and eBGP — required for the PE routers to implement an Inter-AS VLL.

The BGP configuration on PE-3 in AS 64501 is as follows:

```
# on PE-3:
configure
    router Base
        autonomous-system 64501
        bgp
            rapid-withdrawal
            split-horizon
            group "EBGP"
                local-as 64501
                neighbor 192.168.23.1
                    family label-ipv4
                    peer-as 64500
                exit
                neighbor 192.168.34.2
                    family label-ipv4
                    peer-as 64502
                exit
            exit
        exit
```

The address family **label-ipv4** must be configured so that MPLS labels are carried along with MP-BGP Network Layer Reachability Information (NLRIs), see chapter *Separate BGP RIBs for Labeled Routes*. The setting **split-horizon** is optional and prevents that a received route is sent back to the originator, which might result in multiple routes for a certain prefix.

To export the prefixes of the nodes where the Epipe is configured (PE-1 and PE-5) to another AS, a common scenario is to advertise the prefix to be exported within the AS as labeled BGP. Therefore, an export policy is defined for prefix192.0.2.1/32 on PE-1 and this prefix will be advertised to the ASBR in AS 64500, in this case to PE-2.

On PE-2, the labeled BGP route for prefix 192.0.2.1/32 is inactive, because the IGP route for that prefix is preferred. No export policy needs to be configured in the Autonomous System Border Router (ASBR) PE-2 for the EBGP session with PE-3 in AS 64501. Rather, the setting **advertise-inactive** will allow the inactive labeled BGP routes from AS 64500 to be advertised to PE-3 in AS 64501.

Likewise, an export policy will be configured on PE-5 to advertise prefix 192.0.2.5/32 to ASBR PE-4 in AS 64502. On PE-4, BGP is configured with **advertise-inactive** to advertise the labeled BGP route to its EBGP peer, PE-3.

The advantage of this approach is that labeled BGP is used end-to-end between PE-1 and PE-5 and no IGP routes are to be redistributed into BGP, which would be the case if no local BGP labeled routes were advertised within AS 64500 or AS 64502 and only IGP routes were defined within these ASs.

The ASBRs PE-2, PE-3, and PE-4 will swap the BGP labels. PE-3 will advertise the labeled BGP routes learned from AS 64500 to AS 64502 and vice versa and the ASBRs will advertise these labeled routes for remote PE prefixes to their BGP peers. Eventually, PE-1 will have learned a labeled BGP route for prefix 192.0.2.5/32 and PE-5 will have learned a labeled BGP route for prefix 192.0.2.1/32 and a VLL Epipe can be established between PE-1 and PE-5.

The BGP configuration of ASBR PE-2 in AS 64500 is as follows:

```
# on PE-2:
configure
    router Base
        autonomous-system 64500
        bgp
            rapid-withdrawal
            split-horizon
            group "EBGP"
                local-as 64500
                neighbor 192.168.23.2
                    family label-ipv4
                    peer-as 64501
                    advertise-inactive
                exit
            exit
            group "IBGP"
                neighbor 192.0.2.1
                    family label-ipv4
                    next-hop-self
                    peer-as 64500
                exit
            exit
        exit
```

The BGP configuration of ASBR PE-4 in AS 64502 is as follows:

```
# on PE-4:
configure
    router Base
         autonomous-system 64502
         bgp
            rapid-withdrawal
            split-horizon
            group "EBGP"
                local-as 64502
                neighbor 192.168.34.1
                    family label-ipv4
                    peer-as 64501
                    advertise-inactive
                exit
```

```
             exit
             group "IBGP"
                 neighbor 192.0.2.5
                     family label-ipv4
                     next-hop-self
                     peer-as 64502
                 exit
             exit
         exit
```

PE-1 and PE-5 are the PEs to which the CEs are connected in AS 64500 and AS 64502. PE-1 and PE-5 advertise their system prefixes as labeled BGP routes to their BGP peers within the AS.

The BGP configuration of PE-1 is as follows:

```
# on PE-1:
configure
    router Base
         autonomous-system 64500
         bgp
            rapid-withdrawal
            split-horizon
            group "IBGP"
                export "export-PE-1"
                neighbor 192.0.2.2
                    family label-ipv4
                    next-hop-self
                    peer-as 64500
                exit
            exit
        exit
```

The BGP configuration of PE-5 in AS 64502 is as follows:

```
# on PE-5:
configure
    router Base
         autonomous-system 64502
         bgp
            rapid-withdrawal
            split-horizon
            group "IBGP"
                export "export-PEsys"
                neighbor 192.0.2.4
                    family label-ipv4
                    next-hop-self
                    peer-as 64502
                exit
            exit
        exit
```

Policy configuration

The export policies on PE-1 and PE-5 advertise the system addresses to the remote AS.

The export policy on PE-1 has a prefix list that only contains prefix 192.0.2.1/32 as follows:

```
# on PE-1:
configure
    router Base
```

```
        policy-options
            begin
            prefix-list "PE-1"
                prefix 192.0.2.1/32 exact
            exit
            policy-statement "export-PE-1"
                entry 10
                    from
                        prefix-list "PE-1"
                    exit
                    action accept
                    exit
                exit
            exit
            commit
```

A similar export policy can be configured for prefix 192.0.2.5/32 on PE-5. However, the export policy on PE-5 is slightly different: the policy has a prefix list that can be applied for prefixes on multiple PEs, but in this case, only prefix 192.0.2.5/32 will be exported:

```
# on PE-5:
configure
    router Base
        policy-options
            begin
            prefix-list "PEsys"
                prefix 192.0.2.0/29 longer
            exit
            policy-statement "export-PEsys"
                entry 10
                    from
                        protocol direct
                        prefix-list "PEsys"
                    exit
                    action accept
                    exit
                exit
            exit
            commit
```

The same policy could have been applied on PE-1.

Service configuration

Once BGP is configured, the configuration requires the service to be defined (Epipe 1). The focus here is a VLL service, however, it is also possible to have a similar configuration with VPLS services.

The following shows the service level configuration on PE-1:

```
# on PE-1:
configure
     service
        sdp 15 mpls create
            far-end 192.0.2.5
            bgp-tunnel
            no shutdown
        exit
        epipe 1 name "Epipe 1" customer 1 create
            description "Tunnel-PE-1-PE-5"
            sap 1/1/3:1 create
            exit
            spoke-sdp 15:1 create
            exit
            no shutdown
        exit
```

The following CLI shows the service level configuration on PE-5:

```
# on PE-5:
configure
    service
        sdp 51 mpls create
            far-end 192.0.2.1
            bgp-tunnel
            no shutdown
        exit
        epipe 1 name "Epipe 1" customer 1 create
            description "Tunnel-PE-5-PE-1"
            sap 1/1/3:1 create
            exit
            spoke-sdp 51:1 create
            exit
            no shutdown
        exit
```

Show commands and troubleshooting

On PE-5, BGP tunnels exist to the remote AS system addresses that are using LDP as a transport mechanism and the configuration of end-to-end SDPs over which T-LDP service labels are exchanged.

In the following sections, the same commands are launched on the nodes in the following order: first on PE-1 and PE-5; then on PE-3, and finally, on PE-2 and PE-4.

Show commands and troubleshooting on PE-1

The following shows information about SDP 15 on PE-1:

```
*A:PE-1# show service sdp

============================================================================
Services: Service Destination Points
============================================================================
SdpId  AdmMTU  OprMTU  Far End          Adm  Opr         Del     LSP   Sig
----------------------------------------------------------------------------
15     0       1552    192.0.2.5        Up   Up          MPLS    B     TLDP
----------------------------------------------------------------------------
Number of SDPs : 1
----------------------------------------------------------------------------
Legend: R = RSVP, L = LDP, B = BGP, M = MPLS-TP, n/a = Not Applicable
        I = SR-ISIS, O = SR-OSPF, T = SR-TE, F = FPE
============================================================================
```

On PE-1, the VLL Epipe service is up, as follows:

```
*A:PE-1# show service service-using

===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
1            Epipe     Up   Up   1          Epipe 1
2147483648   IES       Up   Down 1          _tmnx_InternalIesService
2147483649   intVpls   Up   Down 1          _tmnx_InternalVplsService
-------------------------------------------------------------------------------
Matching Services : 3
-------------------------------------------------------------------------------
===============================================================================
```

Two LDP sessions have been established from PE-1: a link LDP session with neighbor PE-2 in AS 64500 and a targeted LDP session with PE-5 in AS 64502, as follows:

```
*A:PE-1# show router ldp session ipv4

==============================================================================
LDP IPv4 Sessions
==============================================================================
Peer LDP Id         Adj Type  State         Msg Sent  Msg Recv  Up Time
------------------------------------------------------------------------------
192.0.2.2:0         Link      Established   109       111       0d 00:04:31
192.0.2.5:0         Targeted  Established   21        23        0d 00:01:20
------------------------------------------------------------------------------
No. of IPv4 Sessions: 2
==============================================================================
```

The route table on PE-1 shows that the system IP address of PE-5 is reachable using a BGP tunnel:

```
*A:PE-1# show router route-table

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                  Local   Local     00h06m03s  0
       system                                                       0
192.0.2.2/32                                  Remote  ISIS      00h04m48s  15
       192.168.12.2                                                 10
192.0.2.5/32                                  Remote  BGP_LABEL 00h01m46s  170
       192.0.2.2 (tunneled)                                         10
192.168.12.0/30                               Local   Local     00h06m03s  0
       int-PE-1-PE-2                                                0
-------------------------------------------------------------------------------
No. of Routes: 4
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The following tunnel-table on PE-1 shows the details of the LDP, SDP, and BGP tunnels.

```
*A:PE-1# show router tunnel-table

===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination           Owner     Encap TunnelId  Pref   Nexthop        Metric
```

```
   Color
-------------------------------------------------------------------------------
192.0.2.2/32          ldp       MPLS  65537     9      192.168.12.2   10
192.0.2.5/32          sdp       MPLS  15        5      192.0.2.5      0
192.0.2.5/32          bgp       MPLS  262145    12     192.0.2.2      1000
-------------------------------------------------------------------------------
Flags: B = BGP or MPLS backup hop available
       L = Loop-Free Alternate (LFA) hop available
       E = Inactive best-external BGP route
       k = RIB-API or Forwarding Policy backup hop
===============================================================================
```

The service details for Epipe 1 on PE-1 are as follows:

```
*A:PE-1# show service id 1 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : Epipe
MACSec enabled    : no
Name              : Epipe 1
Description       : Tunnel-PE-1-PE-5
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 01/21/2021 16:01:07
Last Mgmt Change  : 01/21/2021 16:00:53
Test Service      : No
Admin State       : Up                  Oper State        : Up
MTU               : 1514
Vc Switching      : False
SAP Count         : 1                   SDP Bind Count    : 1
Per Svc Hashing   : Disabled
Vxlan Src Tep Ip  : N/A
Force QTag Fwd    : Disabled
Oper Group        : <none>

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/3:1                              q-tag        1518    1518    Up   Up
sdp:15:1 S(192.0.2.5)                    Spok         0       1552    Up   Up
===============================================================================
```

ICMP is used to verify the IP connectivity from PE-1 to the system IP address of PE-5:

```
*A:PE-1# ping 192.0.2.5
PING 192.0.2.5 56 data bytes
64 bytes from 192.0.2.5: icmp_seq=1 ttl=64 time=1.91ms.
64 bytes from 192.0.2.5: icmp_seq=2 ttl=64 time=2.06ms.
64 bytes from 192.0.2.5: icmp_seq=3 ttl=64 time=2.02ms.
64 bytes from 192.0.2.5: icmp_seq=4 ttl=64 time=2.01ms.
64 bytes from 192.0.2.5: icmp_seq=5 ttl=64 time=2.02ms.

---- 192.0.2.5 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 1.91ms, avg = 2.01ms, max = 2.06ms, stddev = 0.050ms
```

Show commands and troubleshooting on PE-5

The same commands on PE-5 result in the following output:

```
*A:PE-5# show service sdp

=============================================================================
Services: Service Destination Points
=============================================================================
SdpId  AdmMTU  OprMTU  Far End          Adm   Opr        Del     LSP   Sig
-----------------------------------------------------------------------------
51     0       1552    192.0.2.1        Up    Up         MPLS    B     TLDP
-----------------------------------------------------------------------------
Number of SDPs : 1
-----------------------------------------------------------------------------
Legend: R = RSVP, L = LDP, B = BGP, M = MPLS-TP, n/a = Not Applicable
        I = SR-ISIS, O = SR-OSPF, T = SR-TE, F = FPE
=============================================================================
```

```
*A:PE-5# show service service-using

================================================================================
Services
================================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
--------------------------------------------------------------------------------
1            Epipe     Up   Up   1          Epipe 1
2147483648   IES       Up   Down 1          _tmnx_InternalIesService
2147483649   intVpls   Up   Down 1          _tmnx_InternalVplsService
--------------------------------------------------------------------------------
Matching Services : 3
--------------------------------------------------------------------------------
================================================================================
```

```
*A:PE-5# show router ldp session ipv4

===============================================================================
LDP IPv4 Sessions
===============================================================================
Peer LDP Id         Adj Type  State         Msg Sent  Msg Recv  Up Time
-------------------------------------------------------------------------------
192.0.2.1:0         Targeted  Established   52        53        0d 00:04:07
192.0.2.4:0         Link      Established   185       188       0d 00:07:57
-------------------------------------------------------------------------------
No. of IPv4 Sessions: 2
===============================================================================
```

```
*A:PE-5# show router route-table

================================================================================
Route Table (Router: Base)
================================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
--------------------------------------------------------------------------------
192.0.2.1/32                                  Remote  BGP_LABEL 00h05m47s  170
       192.0.2.4 (tunneled)                                         10
192.0.2.4/32                                  Remote  ISIS      00h08m13s  15
       192.168.45.1                                                 10
```

```
192.0.2.5/32                                  Local   Local     00h08m19s  0
       system                                                       0
192.168.45.0/30                               Local   Local     00h08m19s  0
       int-PE-5-PE-4                                                0
-------------------------------------------------------------------------------
No. of Routes: 4
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

```
*A:PE-5# show router tunnel-table

===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination           Owner     Encap TunnelId  Pref   Nexthop        Metric
   Color
-------------------------------------------------------------------------------
192.0.2.1/32          sdp       MPLS  51        5      192.0.2.1      0
192.0.2.1/32          bgp       MPLS  262145    12     192.0.2.4      1000
192.0.2.4/32          ldp       MPLS  65537     9      192.168.45.1   10
-------------------------------------------------------------------------------
Flags: B = BGP or MPLS backup hop available
       L = Loop-Free Alternate (LFA) hop available
       E = Inactive best-external BGP route
       k = RIB-API or Forwarding Policy backup hop
===============================================================================
```

```
*A:PE-5# show service id 1 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : Epipe
MACSec enabled    : no
Name              : Epipe 1
Description       : Tunnel-PE-5-PE-1
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 01/21/2021 16:01:07
Last Mgmt Change  : 01/21/2021 16:00:49
Test Service      : No
Admin State       : Up                  Oper State        : Up
MTU               : 1514
Vc Switching      : False
SAP Count         : 1                   SDP Bind Count    : 1
Per Svc Hashing   : Disabled
Vxlan Src Tep Ip  : N/A
Force QTag Fwd    : Disabled
Oper Group        : <none>

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                                Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/3:1                               q-tag        1518    1518    Up   Up
sdp:51:1 S(192.0.2.1)                     Spok         0       1552    Up   Up
```

```
===============================================================================
```

```
*A:PE-5# ping 192.0.2.1
PING 192.0.2.1 56 data bytes
64 bytes from 192.0.2.1: icmp_seq=1 ttl=64 time=1.83ms.
64 bytes from 192.0.2.1: icmp_seq=2 ttl=64 time=2.06ms.
64 bytes from 192.0.2.1: icmp_seq=3 ttl=64 time=2.01ms.
64 bytes from 192.0.2.1: icmp_seq=4 ttl=64 time=2.08ms.
64 bytes from 192.0.2.1: icmp_seq=5 ttl=64 time=2.15ms.

---- 192.0.2.1 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 1.83ms, avg = 2.03ms, max = 2.15ms, stddev = 0.107ms
```

On PE-5, the BGP route to the system IP address of PE-1 can been seen with PE-4 as the next hop:

```
*A:PE-5# show router bgp routes label-ipv4
===============================================================================
 BGP Router ID:192.0.2.5        AS:64502       Local AS:64502
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP Routes
===============================================================================
Flag  Network                                            LocalPref   MED
      Nexthop (Router)                                   Path-Id     IGP Cost
      As-Path                                                        Label
-------------------------------------------------------------------------------
u*>i  192.0.2.1/32                                       100         None
      192.0.2.4                                          None        10
      64501 64500                                                    524285
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

On PE-5, the FIB on slot 1 shows that the system IP address of PE-1 is reachable using BGP over an LDP transport to PE-4:

```
*A:PE-5# show router fib 1

===============================================================================
FIB Display
===============================================================================
Prefix [Flags]                                              Protocol
  NextHop
-------------------------------------------------------------------------------
192.0.2.1/32                                                BGP_LABEL
  192.0.2.4 (Transport:LDP)
192.0.2.4/32                                                ISIS
  192.168.45.1 (int-PE-5-PE-4)
192.0.2.5/32                                                LOCAL
  192.0.2.5 (system)
192.168.45.0/30                                             LOCAL
  192.168.45.0 (int-PE-5-PE-4)
-------------------------------------------------------------------------------
Total Entries : 4
-------------------------------------------------------------------------------
```

```
===============================================================================
```

Show commands on PE-3

The **show** commands on router PE-3 in AS 64501 are as follows:

```
*A:PE-3# show router bgp summary all

===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
Description
ServiceId          AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
192.168.23.1
Def. Instance  64500       22    0 00h09m12s 1/1/1 (Lbl-IPv4)
                           22    0
192.168.34.2
Def. Instance  64502       23    0 00h09m04s 1/1/1 (Lbl-IPv4)
                           25    0

-------------------------------------------------------------------------------
```

```
*A:PE-3# show router bgp routes label-ipv4
===============================================================================
 BGP Router ID:192.0.2.3        AS:64501       Local AS:64501
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP Routes
===============================================================================
Flag  Network                                            LocalPref   MED
      Nexthop (Router)                                   Path-Id     IGP Cost
      As-Path                                                        Label
-------------------------------------------------------------------------------
u*>i  192.0.2.1/32                                       None        None
      192.168.23.1                                       None        0
      64500                                                          524285
u*>i  192.0.2.5/32                                       None        None
      192.168.34.2                                       None        0
      64502                                                          524284
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

The BGP labels are swapped at PE-3, as follows:

```
*A:PE-3# show router bgp inter-as-label

===============================================================================
BGP Inter-AS labels
```

```
Flags: B - entry has backup, P - entry is promoted
===============================================================================
NextHop                       Received       Advertised     Label
                              Label          Label          Origin
-------------------------------------------------------------------------------
192.168.23.1                  524285         524287         External
192.168.34.2                  524284         524286         External
-------------------------------------------------------------------------------
Total Labels allocated:   2
===============================================================================
```

The routing table on PE-3 includes BGP labeled routes to PE-1 and PE-5, as follows:

```
*A:PE-3# show router route-table

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                  Remote  BGP_LABEL 00h10m02s  170
       192.168.23.1                                                 0
192.0.2.3/32                                  Local   Local     00h12m42s  0
       system                                                       0
192.0.2.5/32                                  Remote  BGP_LABEL 00h09m23s  170
       192.168.34.2                                                 0
192.168.23.0/30                               Local   Local     00h12m42s  0
       int-PE-3-PE-2                                                0
192.168.34.0/30                               Local   Local     00h12m42s  0
       int-PE-3-PE-4                                                0
-------------------------------------------------------------------------------
No. of Routes: 5
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

Show commands on PE-2

The commands on PE-2 are as follows:

```
*A:PE-2# show router bgp summary all

===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
Description
ServiceId          AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
192.0.2.1
Def. Instance  64500       36    0 00h16m24s 1/0/1 (Lbl-IPv4)
                           36    0
192.168.23.2
Def. Instance  64501       36    0 00h16m19s 1/1/1 (Lbl-IPv4)
```

```
                           36    0

-------------------------------------------------------------------------------
```

The BGP labels are swapped by PE-2 as follows:

```
*A:PE-2# show router bgp inter-as-label

===============================================================================
BGP Inter-AS labels
Flags: B - entry has backup, P - entry is promoted
===============================================================================
NextHop                       Received       Advertised     Label
                              Label          Label          Origin
-------------------------------------------------------------------------------
192.0.2.1                     524285         524285         Internal
192.168.23.2                  524286         524284         External
-------------------------------------------------------------------------------
Total Labels allocated:   2
===============================================================================
```

```
*A:PE-2# show router route-table

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                  Remote  ISIS      00h13m43s  15
       192.168.12.1                                                 10
192.0.2.2/32                                  Local   Local     00h13m50s  0
       system                                                       0
192.0.2.5/32                                  Remote  BGP_LABEL 00h11m01s  170
       192.168.23.2                                                 0
192.168.12.0/30                               Local   Local     00h13m50s  0
       int-PE-2-PE-1                                                0
192.168.23.0/30                               Local   Local     00h13m50s  0
       int-PE-2-PE-3                                                0
-------------------------------------------------------------------------------
No. of Routes: 5
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

Show commands on PE-4

The **show** commands on PE-4 are the following:

```
*A:PE-4# show router bgp summary all

===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
```

```
Description
ServiceId          AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
192.0.2.5
Def. Instance  64502       29    0 00h12m43s 1/0/1 (Lbl-IPv4)
                           29    0
192.168.34.1
Def. Instance  64501       29    0 00h12m53s 1/1/1 (Lbl-IPv4)
                           30    0

-------------------------------------------------------------------------------
```

```
*A:PE-4# show router bgp routes label-ipv4
===============================================================================
 BGP Router ID:192.0.2.4        AS:64502       Local AS:64502
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP Routes
===============================================================================
Flag  Network                                            LocalPref   MED
      Nexthop (Router)                                   Path-Id     IGP Cost
      As-Path                                                        Label
-------------------------------------------------------------------------------
u*>i  192.0.2.1/32                                       None        None
      192.168.34.1                                       None        0
      64501 64500                                                    524287
*i    192.0.2.5/32                                       100         None
      192.0.2.5                                          None        10
      No As-Path                                                     524285
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

```
*A:PE-4# show router bgp inter-as-label

===============================================================================
BGP Inter-AS labels
Flags: B - entry has backup, P - entry is promoted
===============================================================================
NextHop                       Received       Advertised     Label
                              Label          Label          Origin
-------------------------------------------------------------------------------
192.0.2.5                     524285         524284         Internal
192.168.34.1                  524287         524285         External
-------------------------------------------------------------------------------
Total Labels allocated:   2
===============================================================================
```

Conclusion

The BGP tunnel-based SDP binding is allowed for VLL and VPLS services, including PBB-VPLS. Using RFC 3107, it is possible to implement inter-AS Model C VLLs.

The example used in this chapter illustrates the configuration of an Inter-AS VLL providing access to CE sites. Troubleshooting commands also have been shown to verify all the procedures.

L2 Multicast in EVPN-MPLS VPRN R-VPLS with All-Active Multi-Homing

This chapter provides information about L2 Multicast in EVPN-MPLS VPRN R-VPLS with All-Active Multi-Homing.

Topics in this chapter include:

Applicability

The information and configuration in this chapter are based on SR OS Release 23.7.R1.

Overview

IPv4 multicast traffic can be forwarded from an EVPN-MPLS service into an attached R-VPLS service in which the receiving devices are using EVPN all-active multi-homing.

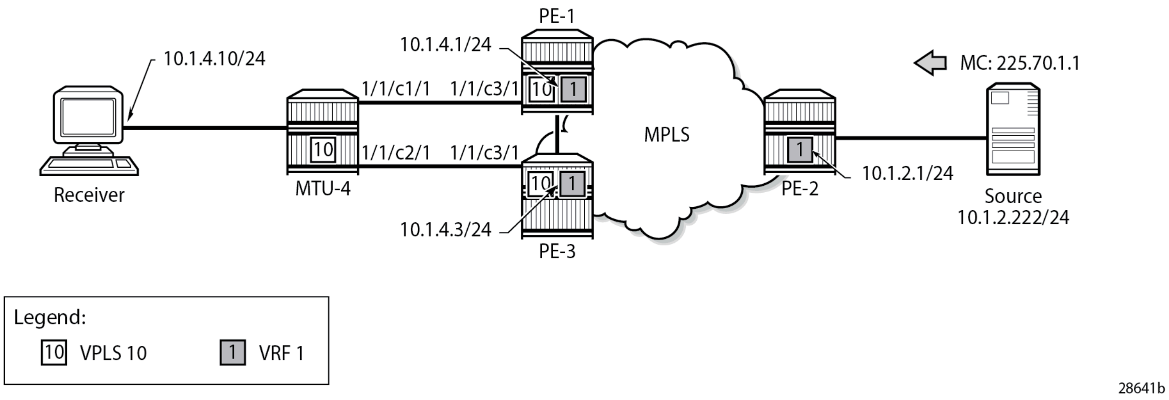
The routed service to which the R-VPLS service attaches can be an IES or a VPRN service. In this way, IPv4 multicast traffic can be transported using native IP for the IES case or NG-MVPN technologies for the VPRN case.

This feature requires:

- IGMP support on the R-VPLS IP interface
- Forwarding IPv4 multicast traffic from the IP interface of a VPRN or IES to its EVPN-MPLS R-VPLS service
- IGMP snooping within the VPLS of the R-VPLS service
- IGMP snooping state synchronization based on the ESI label to synchronize the IGMP snooping state between the all-active (R-)VPLS LAG SAPs

The configuration used in this chapter is the NG-MVPN scenario as shown in Figure 166: Multicast From an EVPN-MPLS Service Into an R-VPLS With All-Active EVPN Multi-Homing.

Figure 166: Multicast From an EVPN-MPLS Service Into an R-VPLS With All-Active EVPN Multi-Homing



A multicast stream is emitted by the source connected to PE-2 with group address 225.70.1.1. A multicast receiver connected to MTU-4 joins group 225.70.1.1. MTU-4 is connected to PE-1 and PE-3 through an all-active multi-homing EVPN Ethernet segment comprising LAG 1. On MTU-4, LAG 1 comprises port 1/1/c1/1 and 1/1/c2/1, and this LAG is used in VPLS 10. On PE-1 and PE-3, VPLS 10 is interconnected with VPRN 1 through an Integrated Routing and Bridging (IRB) interface. VPRN 1 is defined in PE-1, PE-2, and PE-3, and uses NG-MVPN for transporting the multicast traffic through the core of the network. See the EVPN for MPLS Tunnels and EVPN for MPLS Tunnels in Routed VPLS chapters for more information about EVPN. See the "NG-MVPN Configuration with MPLS" and the "NG-MVPN Configuration with PIM" chapters in the *7450 ESS, 7750 SR, and 7950 XRS Layer 3 Services Advanced Configuration Guide for Classic CLI* for more information about NG-MVPN.

Configuration

The initial configuration on the PE nodes includes the following:

- Cards, MDAs, ports
- Router interfaces
- IS-IS (alternatively, OSPF can be used)
- MPLS tunnels between the PEs: LDP- or RSVP-based

BGP is required at the core of the network, using the VPN IPv4 and MVPN IPv4 address families between all PEs, for supporting unicast and multicast traffic on VPRN services, and additionally using the EVPN address family between PE-1 and PE-3 to support EVPN services. The BGP configurations for PE-1, PE-2, and PE-3 are as follows:

```
# on PE-1:
configure
    router
        autonomous-system 64496
        bgp
            family vpn-ipv4 mvpn-ipv4 evpn
            vpn-apply-import
            vpn-apply-export
            rapid-withdrawal
            rapid-update evpn
            group "iBGP"
                neighbor 192.0.2.2
                    peer-as 64496
                exit
                neighbor 192.0.2.3
                    peer-as 64496
                exit
            exit
            no shutdown
        exit
    exit
exit
```

```
# on PE-2:
configure
    router
        autonomous-system 64496
        bgp
            family vpn-ipv4 mvpn-ipv4
            vpn-apply-import
            vpn-apply-export
            rapid-withdrawal
            rapid-update evpn
            group "iBGP"
                neighbor 192.0.2.1
                    peer-as 64496
                exit
                neighbor 192.0.2.3
                    peer-as 64496
                exit
            exit
            no shutdown
        exit
    exit
exit
```

```
# on PE-3:
configure
    router
        autonomous-system 64496
        bgp
            family vpn-ipv4 mvpn-ipv4 evpn
            vpn-apply-import
            vpn-apply-export
            rapid-withdrawal
            rapid-update evpn
            group "iBGP"
                neighbor 192.0.2.1
                    peer-as 64496
                exit
                neighbor 192.0.2.2
                    peer-as 64496
                exit
            exit
            no shutdown
        exit
    exit
exit
```

The receiver connected to MTU-4 joins group 225.70.1.1, and the corresponding multicast stream is emitted by the source that is connected to PE-2. MTU-4 is connected to PE-1 and PE-3 through an all-active multi-homing EVPN Ethernet segment comprising LAG 1. The VPLS and the LAG on MTU-4 are defined as follows:

```
# on MTU-4:
configure
    service
        vpls 10 name "mcast-vpls" customer 1 create
            stp
                shutdown
            exit
            sap 1/2/c1/1 create
                no shutdown
            exit
            sap lag-1:10 create
                no shutdown
            exit
            igmp-snooping
                no shutdown
            exit
            no shutdown
        exit
    exit
exit

configure
    lag 1
        mode access
        encap-type dot1q
        port 1/1/c1/1
        port 1/1/c2/1
        no shutdown
    exit
exit
```

The all-active multi-homing Ethernet segment esi-13 is configured identically on PE-1 and PE-3, as follows. See the EVPN for MPLS Tunnels and EVPN for MPLS Tunnels in Routed VPLS chapters for more information.

```
# on PE-1 and PE-3:
configure
    service
        system
            bgp-evpn
                ethernet-segment "esi-13" create
                    esi 01:00:00:00:00:13:00:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode manual
                        manual
                            preference non-revertive create
                                value 30
                            exit
                        exit
                    exit
                    multi-homing all-active
                    lag 1
                    no shutdown
                exit
            exit
        exit
    exit
exit
```

The multi-homed access circuits of esi-13 are located on port 1/1/c3/1 for PE-1 and PE-3, so the LAG is configured identically, as follows:

```
# on PE-1 and PE-3:
configure
    lag 1
        mode access
        encap-type dot1q
        port 1/1/c3/1
        no shutdown
    exit
exit
```

Also, the EVPN VPLS service with ID 10 is configured identically on PE-1 and PE-3, as follows. The *mcast-vpls* name is needed to link VPLS 10 to VPRN 1 at a later stage, without requiring a physical loop or hairpin. The **allow-ip-int-bind** command enables the VPLS to become an R-VPLS. The **igmp-snooping** and **mrouter-port** commands are required for multicast to work correctly in an all-active multi-homed scenario.

```
# on PE-1 and PE-3:
configure
    service
        vpls 10 name "mcast-vpls" customer 1 create
            allow-ip-int-bind
                igmp-snooping
                    mrouter-port
                exit
            exit
            bgp
            exit
            bgp-evpn
                evi 111
                mpls bgp 1
                    ingress-replication-bum-label
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            igmp-snooping
                no shutdown
            exit
            sap lag-1:10 create
                no shutdown
            exit
            no shutdown
        exit
    exit
exit
```

The VPRN service with ID 1 provides the connection toward MTU-4 via VPLS 10, through the *int-MCAST-VPLS* interface with address 10.1.4.1/24 on PE-1, and with address 10.1.4.3/24 on PE-3. This L3 interface is linked to VPLS 10 with the **vpls "mcast-vpls"** command. The *int-MCAST-VPLS* interface is also included in the IGMP and PIM configurations of VPRN 1. The full configuration of VPRN 1 on PE-1 is as follows. The configuration of VPRN 1 on PE-3 is similar.

```
# on PE-1:
configure
    service
```

```
        vprn 1 name "VPRN 1" customer 1 create
            bgp-ipvpn
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    route-distinguisher 64496:1
                    vrf-target target:64496:1
                    no shutdown
                exit
            exit
            interface "int-MCAST-VPLS" create
                address 10.1.4.1/24
                vpls "mcast-vpls"
                exit
            exit
            interface "int-PE-1-CE-1" create
                address 10.1.1.1/24
                sap 1/2/c1/1 create
                exit
            exit
            interface "system" create
                address 192.0.2.101/32
                loopback
            exit
            igmp
                ssm-translate
                    grp-range 225.70.1.1 225.70.255.255
                        source 10.1.2.222
                    exit
                exit
                interface "int-MCAST-VPLS"
                    no shutdown
                exit
                interface "int-PE-1-CE-1"
                    no shutdown
                exit
                no shutdown
            exit
            pim
                interface "int-MCAST-VPLS"
                exit
                interface "system"
                exit
                no shutdown
            exit
            mvpn
                auto-discovery default
                c-mcast-signaling bgp
                mdt-type receiver-only
                provider-tunnel
                    inclusive
                        mldp
                            no shutdown
                        exit
                    exit
                    selective
                        mldp
                            no shutdown
                        exit
                        data-threshold 224.0.0.0/4 1
                    exit
                exit
                vrf-target unicast
                exit
            exit
            no shutdown
        exit
    exit
exit
```

The full configuration of VPRN 1 on PE-2 is as follows. The *int-PE-2-CE-2-source* interface provides the connection to the multicast source.

```
# on PE-2:
configure
    service
        vprn 1 name "VPRN 1" customer 1 create
            bgp-ipvpn
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    route-distinguisher 64496:1
                    vrf-target target:64496:1
                    no shutdown
                exit
            exit
            interface "int-PE-2-CE-2-source" create
                address 10.1.2.1/24
                sap 1/2/c1/1 create
                exit
            exit
            interface "system" create
                address 192.0.2.102/32
                loopback
            exit
            pim
                interface "int-PE-2-CE-2-source"
                exit
                interface "system"
                exit
                no shutdown
            exit
            mvpn
                auto-discovery default
                c-mcast-signaling bgp
                mdt-type sender-only
                provider-tunnel
                    inclusive
                        mldp
                            no shutdown
                        exit
                    exit
                    selective
                        mldp
                            no shutdown
                        exit
                        data-threshold 224.0.0.0/4 1
                    exit
                exit
                vrf-target unicast
                exit
            exit
            no shutdown
        exit
    exit
```

```
exit
```

Verification

The following command shows that *esi-13* is an all-active multi-homed Ethernet segment, on PE-1. The same command can be executed on PE-3.

```
*A:PE-1# show service system bgp-evpn ethernet-segment name "esi-13"

===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : esi-13
Eth Seg Type            : None
Admin State             : Enabled            Oper State         : Up
ESI                     : 01:00:00:00:00:13:00:00:00:01
Oper ESI                : 01:00:00:00:00:13:00:00:00:01
Auto-ESI Type           : None
AC DF Capability        : Include
Multi-homing            : allActive          Oper Multi-homing  : allActive
ES SHG Label            : 524282
Source BMAC LSB         : None
Lag Id                  : 1
ES Activation Timer     : 3 secs
Oper Group              : (Not Specified)
Svc Carving             : manual             Oper Svc Carving   : manual
Cfg Range Type          : lowest-pref

-------------------------------------------------------------------------------
DF Pref Election Information
-------------------------------------------------------------------------------
Preference     Preference     Last Admin Change        Oper Pref      Do No
Mode           Value                                   Value          Preempt
-------------------------------------------------------------------------------
non-revertive  30             07/20/2023 15:20:09      30             Disabled
-------------------------------------------------------------------------------
EVI Ranges: <none>
ISID Ranges: <none>
Vprn NextHop EVI Ranges : <none>
===============================================================================
```

The output from the following commands on PE-1 and PE-3 shows that for esi-13, PE-1 is Non-Designated Forwarder (NDF), whereas PE-3 is Designated Forwarder (DF).

```
*A:PE-1# show service id 10 ethernet-segment "esi-13"

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                           Status
-------------------------------------------------------------------------------
lag-1:10              esi-13                            NDF
===============================================================================
No sdp entries
No vxlan instance entries
```

```
*A:PE-3# show service id 10 ethernet-segment "esi-13"

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
lag-1:10              esi-13                           DF
===============================================================================
No sdp entries
No vxlan instance entries
```

A stream with group address 225.70.1.1 is started by the multicast source and joined by the multicast receiver connected to MTU-4. This stream is forwarded from PE-2 to PE-3; PE-1 is not involved in the forwarding.

PE-1 maintains IGMP state for group 225.70.1.1 in VPRN 1, and so does PE-3. PE-1 and PE-3 synchronize IGMP state using a data-driven mechanism. The forwarding list includes the *int-MCAST-VPLS* interface, as follows:

```
*A:PE-1# show router 1 igmp group 225.70.1.1 interfaces
===============================================================================
IGMP Interface Groups
===============================================================================

(*,225.70.1.1)                                            UpTime: 0d 00:01:56
    Fwd List  : int-MCAST-VPLS
-------------------------------------------------------------------------------
Entries : 1
===============================================================================
```

PE-1 maintains PIM state for group 225.70.1.1, as follows. The outgoing interfaces list is empty and the forwarding rate is zero; both are indications that PE-1 is not forwarding any multicast traffic.

```
*A:PE-1# show router 1 pim group 225.70.1.1 detail

===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 225.70.1.1
Source Address     : 10.1.2.222
RP Address         : 0
Advt Router        : 192.0.2.2
Flags              :                  Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 192.0.2.2
MRIB Src Flags     : remote
Keepalive Timer Exp: 0d 00:02:05
Up Time            : 0d 00:02:42       Resolved By        : rtable-u

Up JP State        : Not Joined        Up JP Expiry       : 0d 00:00:00
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 192.0.2.2
Incoming Intf      : mpls-if-73728
Outgoing Intf List :

Curr Fwding Rate   : 0.000 kbps
Forwarded Packets  : 0                  Discarded Packets  : 0
Forwarded Octets   : 0                  RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
```

```
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

PE-2 and PE-3 are forwarding the stream as indicated by the PIM state for this group, as follows:

```
*A:PE-2# show router 1 pim group 225.70.1.1 detail

===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 225.70.1.1
Source Address     : 10.1.2.222
RP Address         : 0
Advt Router        : 192.0.2.2
Flags              :                    Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 10.1.2.222
MRIB Src Flags     : direct
Keepalive Timer    : Not Running
Up Time            : 0d 00:02:15        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:00
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 10.1.2.222
Incoming Intf      : int-PE-2-CE-2-source
Outgoing Intf List : mpls-if-73728 (mpls-if-73729)

Curr Fwding Rate   : 9751.560 kbps
Forwarded Packets  : 51112              Discarded Packets  : 0
Forwarded Octets   : 75747984           RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

```
*A:PE-3# show router 1 pim group 225.70.1.1 detail

===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 225.70.1.1
Source Address     : 10.1.2.222
RP Address         : 0
Advt Router        : 192.0.2.2
Flags              :                    Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 192.0.2.2
MRIB Src Flags     : remote
Keepalive Timer Exp: 0d 00:02:04
Up Time            : 0d 00:02:44        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:16
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No
```

```
Rpf Neighbor       : 192.0.2.2
Incoming Intf      : mpls-if-73728
Incoming SPMSI Intf: mpls-if-73729
Outgoing Intf List : int-MCAST-VPLS

Curr Fwding Rate   : 9745.632 kbps
Forwarded Packets  : 74533              Discarded Packets  : 0
Forwarded Octets   : 110457906          RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

The outgoing interfaces on PE-2 and PE-3 are the *mpls-if-73728* PMSI interface and the *int-MCAST-VPLS* interfaces, respectively. The properties of the S-PMSI interface are as follows:

```
*A:PE-2# show router 1 pim tunnel-interface "mpls-if-73728" detail

===============================================================================
PIM Interface ipv4 mpls-if-73728
===============================================================================
Admin Status       : Up                 Oper Status        : Up
IPv4 Admin Status  : Up                 IPv4 Oper Status   : Up
DR                 : 192.0.2.2
Auto-created       : No
Transport Type     : MVPN-Pmsi

-------------------------------------------------------------------------------
PIM Group Source
-------------------------------------------------------------------------------
Group Address      : 225.70.1.1
Source Address     : 10.1.2.222
Interface          : mpls-if-73728      Type               : (S,G)
RP Address         : 0.0.0.0
Up Time            : 0d 00:02:23

Join Prune State   : Join               Expires            : Never
Prune Pend Expires : N/A

Assert State       : No Info
-------------------------------------------------------------------------------
Interfaces : 1
===============================================================================
```

The stream is received on the incoming PMSI interface *mpls-if-73728* on PE-3. The properties of this PMSI interface are as follows:

```
*A:PE-3# show router 1 pim tunnel-interface "mpls-if-73728" detail

===============================================================================
PIM Interface ipv4 mpls-if-73728
===============================================================================
Admin Status       : Up                 Oper Status        : Up
IPv4 Admin Status  : Up                 IPv4 Oper Status   : Up
DR                 : 192.0.2.2
Auto-created       : No
Transport Type     : MVPN-Pmsi
-------------------------------------------------------------------------------
Interfaces : 1
===============================================================================
```

PE-3 sends this multicast stream to MTU-4, which in turn sends it to the receiver that sent the join, so the path taken by the multicast stream runs via PE-2, PE-3, and MTU-4.

In the example from Figure 166: Multicast From an EVPN-MPLS Service Into an R-VPLS With All-Active EVPN Multi-Homing, and the commands and traces that follow, PE-1 is the active IGMP querier using address 10.1.4.1, sending out the queries across the L2 domain. The group queries are sent by PE-1 to PE-3 across the EVPN-MPLS tunnel because PE-3 is DF for *esi-13*, then forwarded onto MTU-4 to reach the (potential) receiver. MTU-4 relays the IGMP responses from the receiver to one of the links; in this example, the link between MTU-4 and PE-1. When the IGMP response for joining the 225.1.70.1 stream is received on PE-1, this event is signaled across the EVPN-MPLS tunnel because it is received over esi-13. This way, the IGMP state is synchronized between PE-3 and PE-1 in a data-driven way.

The basic IGMP snooping state for VPLS 10 on PE-1 and PE-3 is as follows. The output shows that IGMP snooping is enabled on ports sap:lag-1:10, rvpls, and evpn-mpls.

```
*A:PE-1# show service id 10 igmp-snooping base

===============================================================================
IGMP Snooping Base info for service 10
===============================================================================
Admin State : Up
Querier     : 10.1.4.1 on rvpls int-MCAST-VPLS
SBD service : N/A
Evpn-proxy  : Disabled
-------------------------------------------------------------------------------
Port                       Oper MRtr Pim  Send Max  Max  Max   MVR       Num
Id                         Stat Port Port Qrys Grps Srcs Grp   From-VPLS Grps
                                                         Srcs
-------------------------------------------------------------------------------
sap:lag-1:10               Up   No   No   No   None None None  Local     1
rvpls                      Up   Yes  No   N/A  N/A  N/A  N/A   N/A       N/A
evpn-mpls
                           Up   Yes  No   N/A  N/A  N/A  N/A   N/A       N/A
===============================================================================
```

```
*A:PE-3# show service id 10 igmp-snooping base

===============================================================================
IGMP Snooping Base info for service 10
===============================================================================
Admin State : Up
Querier     : 10.1.4.1 on evpn-mpls
SBD service : N/A
Evpn-proxy  : Disabled
-------------------------------------------------------------------------------
Port                       Oper MRtr Pim  Send Max  Max  Max   MVR       Num
Id                         Stat Port Port Qrys Grps Srcs Grp   From-VPLS Grps
                                                         Srcs
-------------------------------------------------------------------------------
sap:lag-1:10               Up   No   No   No   None None None  Local     1
rvpls                      Up   Yes  No   N/A  N/A  N/A  N/A   N/A       N/A
evpn-mpls
                           Up   Yes  No   N/A  N/A  N/A  N/A   N/A       N/A
===============================================================================
```

PE-1 sends the IGMP queries on VPRN 1 via the *int-MCAST-VPLS* interface, so the VPLS that is referenced in the *int-MCAST-VPLS* interface registers the ports on which the IGMP queries are received as multicast router ports. EVPN-MPLS tunnels are always multicast router ports. The following output displays the source addresses of the multicast routers:

```
*A:PE-1# show service id 10 igmp-snooping mrouters

===============================================================================
IGMP Snooping Multicast Routers for service 10
===============================================================================
MRouter          Port Id                    Up Time        Expires   Version
-------------------------------------------------------------------------------
10.1.4.1         rvpls                      0d 00:11:05    130s      3
-------------------------------------------------------------------------------
Number of mrouters: 1
===============================================================================
```

```
*A:PE-3# show service id 10 igmp-snooping mrouters

===============================================================================
IGMP Snooping Multicast Routers for service 10
===============================================================================
MRouter          Port Id                    Up Time        Expires   Version
-------------------------------------------------------------------------------
10.1.4.1         evpn-mpls                  0d 00:10:27    253s      3
-------------------------------------------------------------------------------
Number of mrouters: 1
===============================================================================
```

The IGMP snooping querier properties for VPLS 10 on PE-1 and PE-3 are as follows:

```
*A:PE-1# show service id 10 igmp-snooping querier

===============================================================================
IGMP Snooping Querier info for service 10
===============================================================================
Port Id                 : r-vpls int-MCAST-VPLS
IP Address              : 10.1.4.1
Expires                 : 148s
Up Time                 : 0d 00:10:46
Version                 : 3

General Query Interval  : 125s
Query Response Interval : 10.0s
Robust Count            : 2
===============================================================================

*A:PE-3# show service id 10 igmp-snooping querier

===============================================================================
IGMP Snooping Querier info for service 10
===============================================================================
Port Id                 : evpn-mpls
IP Address              : 10.1.4.1
Expires                 : 146s
Up Time                 : 0d 00:10:09
Version                 : 3

General Query Interval  : 125s
Query Response Interval : 10.0s
Robust Count            : 2
===============================================================================
```

IGMP snooping in VPLS 10 registers the reports in the IGMP snooper port database (port-db). The port-db can be displayed with a show command, and specifying a SAP limits the output generated by this command, as follows:

```
*A:PE-1# show service id 10 igmp-snooping port-db sap lag-1:10

===============================================================================
IGMP Snooping SAP lag-1:10 Port-DB for service 10
===============================================================================
Group Address   Mode    Type    From-VPLS  Up Time        Expires  Num   MC
                                                                   Src   Stdby
-------------------------------------------------------------------------------
225.70.1.1      exclude dynamic local      0d 00:04:05    never    0
-------------------------------------------------------------------------------
Number of groups: 1
===============================================================================

*A:PE-3# show service id 10 igmp-snooping port-db sap lag-1:10

===============================================================================
IGMP Snooping SAP lag-1:10 Port-DB for service 10
===============================================================================
Group Address   Mode    Type    From-VPLS  Up Time        Expires  Num   MC
                                                                   Src   Stdby
-------------------------------------------------------------------------------
225.70.1.1      exclude dynamic local      0d 00:04:06    250s     0
-------------------------------------------------------------------------------
Number of groups: 1
===============================================================================
```

IGMP snooping statistics show the number of received, transmitted, and forwarded IGMP messages per type, and also provide drop counts per error type, as follows:

```
*A:PE-1# show service id 10 igmp-snooping statistics

===============================================================================
IGMP Snooping Statistics for service 10
===============================================================================
Message Type            Received      Transmitted   Forwarded
-------------------------------------------------------------------------------
General Queries         1             0             12
Group Queries           0             0             0
Group-Source Queries    0             0             0
V1 Reports              0             0             0
V2 Reports              0             0             0
V3 Reports              6             3             3
V2 Leaves               0             0             0
Unknown Type            0             N/A           0
EVPN SMET Routes        0             0             N/A
-------------------------------------------------------------------------------
Drop Statistics
-------------------------------------------------------------------------------
Bad Length               : 0
Bad IP Checksum          : 0
Bad IGMP Checksum        : 0
Bad Encoding             : 0
No Router Alert          : 0
Zero Source IP           : 0
Wrong Version            : 0
Lcl-Scope Packets        : 0
Rsvd-Scope Packets       : 0
```

```
Send Query Cfg Drops     : 0
Import Policy Drops      : 0
Exceeded Max Num Groups  : 0
Exceeded Max Num Sources : 0
Exceeded Max Num Grp Srcs: 0
MCAC Policy Drops        : 0
MCS Failures             : 0

MVR From VPLS Cfg Drops  : 0
MVR To SAP Cfg Drops     : 0
===============================================================================

*A:PE-3# show service id 10 igmp-snooping statistics

===============================================================================
IGMP Snooping Statistics for service 10
===============================================================================
Message Type            Received      Transmitted   Forwarded
-------------------------------------------------------------------------------
General Queries         6             0             6
Group Queries           0             0             0
Group-Source Queries    0             0             0
V1 Reports              0             0             0
V2 Reports              0             0             0
V3 Reports              6             3             0
V2 Leaves               0             0             0
Unknown Type            0             N/A           0
EVPN SMET Routes        0             0             N/A
-------------------------------------------------------------------------------
Drop Statistics
-------------------------------------------------------------------------------
Bad Length               : 0
Bad IP Checksum          : 0
Bad IGMP Checksum        : 0
Bad Encoding             : 0
No Router Alert          : 0
Zero Source IP           : 0
Wrong Version            : 0
Lcl-Scope Packets        : 0
Rsvd-Scope Packets       : 0

Send Query Cfg Drops     : 0
Import Policy Drops      : 0
Exceeded Max Num Groups  : 0
Exceeded Max Num Sources : 0
Exceeded Max Num Grp Srcs: 0
MCAC Policy Drops        : 0
MCS Failures             : 0

MVR From VPLS Cfg Drops  : 0
MVR To SAP Cfg Drops     : 0
===============================================================================
```

Debug

Debugging is useful for troubleshooting purposes, and the debug configuration used on PE-1 and PE-3 for checking IGMP and IGMP snooping functionalities is as follows:

```
debug
    router "1"
        igmp
```

```
            packet mode egr-ingr-and-dropped
        exit
    exit
    service
        id 10
            igmp-snooping
                mode egr-ingr-and-dropped
                detail-level high
                sap lag-1:10
                evpn-mpls
            exit
        exit
    exit
exit
```

When group 225.70.1.1 is joined, the trace on PE-1 is as follows. Event 7 is the IGMPv3 join message for group 225.70.1.1 received on SAP lag-1:10 in VPLS 10 from the receiver. The reception of this message is synchronized across the EVPN-MPLS tunnel for VPLS 10, as indicated by event 8. Event 10 is the IGMPv3 join message as received on interface *int-MCAST-VPLS* by VPRN 1.

```
7 2023/07/20 15:28:09.540 CEST MINOR: DEBUG #2001 Base IGMP
"IGMP: RX packet on svc 10
  from chaddr 04:0f:ff:00:01:41
  Port  : sap lag-1:10
  SrcIp : 0.0.0.0
  DstIp : 224.0.0.22
  Raw pkt dump:
  22 00 f7 b6 00 00 00 01 04 00 00 00 e1 46 01 01
  Type  : V3 REPORT
    Num Group Records: 1
        Group Record Type: CHG_TO_EXCL (4), AuxDataLen 0, Num Sources 0
          Group Addr: 225.70.1.1

"

8 2023/07/20 15:28:09.540 CEST MINOR: DEBUG #2001 Base IGMP
"IGMP: TX packet on svc 10
  from chaddr 5e:00:00:16:04:0f
  send towards ES : esi-13
  Port  : evpn-mpls
  SrcIp : 0.0.0.0
  DstIp : 224.0.0.22
  Raw pkt dump:
  22 00 f7 b6 00 00 00 01 04 00 00 00 e1 46 01 01
  Type  : V3 REPORT
    Num Group Records: 1
        Group Record Type: CHG_TO_EXCL (4), AuxDataLen 0, Num Sources 0
          Group Addr: 225.70.1.1

"
---snip---

10 2023/07/20 15:28:09.541 CEST MINOR: DEBUG #2001 vprn1 IGMP[2]
"IGMP[2]: RX-PKT
[000 00:16:13.580] IGMP interface int-MCAST-VPLS [ifIndex 4] V3 PDU: 0.0.0.0 -> 224.0.0.22 pdu
Len 16
    Type: V3 REPORT maxrespCode 0x0 checkSum 0xf7b6
    Num Group Records: 1
        Group Record 0
        Type: CHG_TO_EXCL, AuxDataLen 0, Num Sources 0
        Mcast Addr: 225.70.1.1
        Source Address List
```

```
"
```

The trace on PE-3 is as follows. Event 8 is the reception of the snooping state synchronization across the EVPN-MPSL tunnel, and event 11 is the IGMPv3 join as received on interface *int-MCAST-VPLS* by VPRN 1.

```
8 2023/07/20 15:28:09.489 CEST MINOR: DEBUG #2001 Base IGMP
"IGMP: RX packet on svc 10
  from chaddr 04:0f:ff:00:01:41
  received via evpn-mpls on ES : esi-13
  Port  : sap lag-1:10
  SrcIp : 0.0.0.0
  DstIp : 224.0.0.22
  Raw pkt dump:
  22 00 f7 b6 00 00 00 01 04 00 00 00 e1 46 01 01
  Type  : V3 REPORT
    Num Group Records: 1
        Group Record Type: CHG_TO_EXCL (4), AuxDataLen 0, Num Sources 0
          Group Addr: 225.70.1.1

"
---snip---

11 2023/07/20 15:28:09.490 CEST MINOR: DEBUG #2001 vprn1 IGMP[2]
"IGMP[2]: RX-PKT
[000 00:16:06.580] IGMP interface int-MCAST-VPLS [ifIndex 4] V3 PDU: 0.0.0.0 -> 224.0.0.22 pdu
Len 16
    Type: V3 REPORT maxrespCode 0x0 checkSum 0xf7b6
    Num Group Records: 1
        Group Record 0
        Type: CHG_TO_EXCL, AuxDataLen 0, Num Sources 0
        Mcast Addr: 225.70.1.1
        Source Address List

"
```

Similar events are logged when the multicast receiver leaves the 225.70.1.1 group.

Conclusion

By connecting customers to EVPN-MPLS VPRN/IES routed services via an R-VPLS, service providers can offer IPv4 multicast services to customers in an all-active multi-homing scenario.

L2 Services with Auto-GRE Spoke-SDPs

This chapter provides information about L2 Services with Auto-GRE Spoke-SDPs.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS Release 16.0.R4, but the CLI in the current edition is based on SR OS Release 21.5.R1. Auto-GRE spoke-SDPs are supported in BGP-VPLS, BGP-AD, BGP-VPWS, and with FEC 129 spoke-SDPs in SR OS Release 16.0.R1, and later.

Overview

When the connectivity between nodes is IP-based (not MPLS), VPWS and VPLS services can use manually provisioned or auto-generated GRE transport tunnels. For auto-GRE transport tunnels, the signaling can be BGP or Targeted LDP (T-LDP). BGP signaling is more scalable than T-LDP, because T-LDP requires point-to-point sessions between communicating peers.

Auto-GRE spoke-SDPs can be used in the following services:

- BGP-VPLS with BGP signaling
- LDP VPLS using BGP-AD with T-LDP signaling
- BGP-VPWS with BGP signaling
- Dynamic Multi-segment Pseudowire (MS-PW) spoke-SDP Forwarding Equivalence Class (FEC) 129 with T-LDP signaling

PW templates for auto-GRE spoke-SDPs are configured with the creation-time parameter **auto-gre-sdp**. The **auto-gre-sdp** parameter can be combined with the creation-time parameter **prefer-provisioned-sdp**, but not with **use-provisioned-sdp** (because that might contradict the use of auto-GRE spoke-SDPs), as follows:

```
*A:PE-1>config>service# pw-template 3 name "PW3" ?
  - pw-template <policy-id> [create] [prefer-provisioned-sdp] [name <name>]
    [auto-gre-sdp]
  - no pw-template <policy-id>
  - pw-template <policy-id> use-provisioned-sdp [create] [name <name>]
---snip---
```

The auto-GRE SDP and SDP binding are created after a matching BGP route has been received. Subsequent requests for an auto-GRE SDP of the same type and to the same destination as an existing auto-GRE SDP will use the existing auto-GRE SDP.

Downstream fragmentation is allowed for auto-GRE SDPs by clearing the Don't Fragment (DF) bit in the GRE IP header. The following command controls fragmentation for a PW template:

```
configure
    service
        pw-template 40 name "PW40" auto-gre-sdp create
            allow-fragmentation
        exit
```

The following PW template parameters are not supported with GRE tunnels and will be ignored when a GRE SDP is auto-created:

- Hash label
- Entropy label
- SDP include/exclude (there is no mechanism to configure an SDP admin group for auto-GRE SDPs)

However, these parameters are relevant for provisioned MPLS SDPs when the PW template is configured with **prefer-provisioned-sdp**.

The **pw-template-binding** parameter in the **bgp** context of the L2 service allows to configure the PW template to be used. It is possible to define multiple PW template bindings within a service. The mechanism for selecting the PW template is as follows:

- In BGP-VPWS, BGP-VPLS, and BGP-AD services, the PW template binding selection is based on matching the configured import Route Targets (RTs) for a PW template binding with the RTs in the received routes.
- The binding with the first matching RT is chosen. If no import RTs are configured, the lowest PW template binding ID is used.
- It is not possible to add RTs to BGP-VPWS BGP updates using import or export policies, because they are ignored. However, the RT exported to select the destination service can be used on the receiving PE with PW template binding statement to influence the PW template to be selected; see the first use case in the Configuration section.
- If the selected PW template is configured with **prefer-provisioned-sdp** and an SDP with a matching far-end address exists, the system chooses the SDP with the lowest metric from the tunnel table. If multiple matching SDPs with the same metric occur, the highest SDP ID that is operationally up is chosen.

The following **tools** command allows for PW template bindings to change:

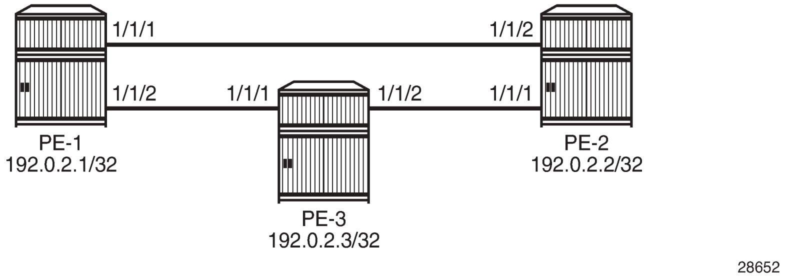
```
*A:PE-1# tools perform service id 1 eval-pw-template ?
  - eval-pw-template <policy-id> [allow-service-impact]

 <policy-id>          : [1..2147483647]
 <allow-service-imp*> : keyword
```

The policy ID refers to the PW template currently in use. With the **allow-service-impact** option, the current binding will be torn down and re-signaled.

Configuration

Figure 167: Example topology shows the example topology with three PEs in AS 64500. Services will be configured on PE-1 and PE-2, and PE-3 is the route reflector (RR).

Figure 167: Example topology



The initial configuration on the three PEs includes:

- Cards, MDAs, ports
- Router interfaces
- IS-IS as IGP (alternatively, OSPF can be used)

Auto-GRE spoke-SDPs are configured in the following use cases:

1. BGP-VPLS with BGP signaling
2. BGP-AD in VPLS with T-LDP signaling
3. BGP-VPWS with BGP signaling
4. Dynamic MS-PW spoke-SDP FEC 129 with T-LDP signaling

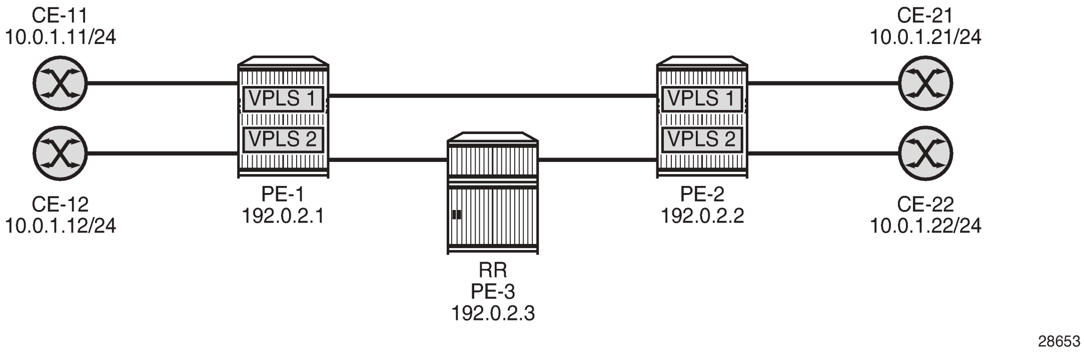
In the first three use cases (BGP-VPLS, BGP-AD, BGP-VPWS), BGP is configured for the L2-VPN address family; in the last use case (dynamic MS-PW), BGP is configured for the MS-PW address family.

In each of the use cases, two L2 services will be configured using different PW templates with **auto-gre-sdp**: one with **prefer-provisioned-sdp** and one without.

Auto-GRE spoke-SDPs in BGP-VPLS

Figure 168: BGP-VPLS with auto-GRE spoke-SDPs shows the example topology with BGP-VPLSs 1 and 2 configured on PE-1 and PE-2. BGP is configured for the L2-VPN address family with PE-3 as Route Reflector (RR). The CEs are emulated through VPRNs configured on the PEs and connected to the VPLSs via Port Cross-connect (PXC).

Figure 168: BGP-VPLS with auto-GRE spoke-SDPs



BGP configuration

For the BGP-VPLS, BGP-AD, and BGP-VPWS use cases, BGP is configured with the L2-VPN address family. The BGP configuration on PE-1 and PE-2 is identical, as follows:

```
# on PE-1, PE-2::
configure
    router Base
        autonomous-system 64500
        bgp
            rapid-withdrawal
            split-horizon
            group "WAN"
                family l2-vpn
                type internal
                neighbor 192.0.2.3
                exit
            exit
```

On RR PE-3, BGP is configured as follows:

```
configure
    router Base
        autonomous-system 64500
        bgp
            rapid-withdrawal
            split-horizon
            group "WAN"
                family l2-vpn
                cluster 192.0.2.3
                type internal
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.2
                exit
            exit
```

Service configuration

The configuration of BGP-VPLS services is described in the BGP VPLS chapter.

PW template 10 is configured with **auto-gre-sdp**; PW template 20 is configured with **prefer-provisioned-sdp** and **auto-gre-sdp**. Because only IP connectivity is present between the nodes (no MPLS), the provisioned SDP is GRE-based using BGP signaling (no T-LDP). VPLS 1 has PW template bindings with IDs 10 and 20; VPLS 2 is configured with PW template binding 20. The service configuration on PE-1 is as follows:

```
# on PE-1:
configure
    service
        sdp 12 create
            signaling bgp
            far-end 192.0.2.2
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        pw-template 10 name "PW10-auto-GRE" auto-gre-sdp create
        exit
        pw-template 20 name "PW20-auto-GRE_prefer-prov" prefer-provisioned-sdp
            auto-gre-sdp create
        exit
        vpls 1 name "BGP-VPLS-1" customer 1 create
            description "BGP-VPLS with auto-GRE spoke-SDP"
            bgp
                route-distinguisher 64500:1
                route-target export target:64500:1 import target:64500:1
                pw-template-binding 10
                exit
                pw-template-binding 20
                exit
            exit
            bgp-vpls
                max-ve-id 100
                ve-name "PE-1"
                    ve-id 1
                exit
                no shutdown
            exit
            stp
                shutdown
            exit
            sap pxc-10.a:1 create                # SAP to connect to CE-11
                no shutdown
            exit
            no shutdown
        exit
        vpls 2 name "BGP-VPLS-2" customer 1 create
            description "BGP-VPLS with auto-GRE spoke-SDP_prefer provisioned SDP"
            bgp
                route-distinguisher 64500:2
                route-target export target:64500:2 import target:64500:2
                pw-template-binding 20
                exit
            exit
            bgp-vpls
                max-ve-id 100
                ve-name "PE-1"
                    ve-id 1
                exit
                no shutdown
            exit
            stp
                shutdown
            exit
            sap pxc-10.a:2 create                # SAP to connect to CE-12
                no shutdown
            exit
            no shutdown
        exit
```

The service configuration on PE-2 is similar, but the VE name is "PE-2" and the VE ID equals 2 instead, as follows:

```
# on PE-2:
configure
    service
        sdp 21 create
            signaling bgp
            far-end 192.0.2.1
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        pw-template 10 name "PW10-auto-GRE" auto-gre-sdp create
        exit
        pw-template 20 name "PW20-auto-GRE_prefer-prov" prefer-provisioned-sdp
            auto-gre-sdp create
        exit
        vpls 1 name "BGP-VPLS-1" customer 1 create
            description "BGP-VPLS with auto-GRE spoke-SDP"
            bgp
                route-distinguisher 64500:1
                route-target export target:64500:1 import target:64500:1
                pw-template-binding 10
                exit
                pw-template-binding 20
                exit
            exit
            bgp-vpls
                max-ve-id 100
                ve-name "PE-2"
                    ve-id 2
                exit
                no shutdown
            exit
            stp
                shutdown
            exit
            sap pxc-10.a:1 create                # SAP to connect to CE-21
                no shutdown
            exit
            no shutdown
        exit
        vpls 2 name "BGP-VPLS-2" customer 1 create
            description "BGP-VPLS with auto-GRE spoke-SDP_prefer provisioned SDP"
            bgp
                route-distinguisher 64500:2
                route-target export target:64500:2 import target:64500:2
                pw-template-binding 20
                exit
            exit
            bgp-vpls
                max-ve-id 100
                ve-name "PE-2"
                    ve-id 2
                exit
                no shutdown
            exit
            stp
                shutdown
            exit
            sap pxc-10.a:2 create                # SAP to connect to CE-22
                no shutdown
            exit
            no shutdown
        exit
```

The following L2-VPN routes are received on PE-1: one for VPLS 1 with RD 64500:1 and another for VPLS 2 with RD 64500:2.

```
*A:PE-1# show router bgp routes l2-vpn
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP L2VPN Routes
===============================================================================
Flag  RouteType                   Prefix                             MED
      RD                          SiteId                             Label
      Nexthop                     VeId                   BlockSize   LocalPref
      As-Path                     BaseOffset             vplsLabelBa
                                                         se
-------------------------------------------------------------------------------
u*>i  VPLS                        -                      -           0
      64500:1                     -                                  -
      192.0.2.2                   2                      8           100
      No As-Path                  1                      524280
u*>i  VPLS                        -                      -           0
      64500:2                     -                                  -
      192.0.2.2                   2                      8           100
      No As-Path                  1                      524272
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

VPLS 1 is configured with two PW template bindings without import RT. Because the PW template binding with the lowest ID is preferred, PW template 10 is used and therefore, the following GRE SDP 32767 is auto-created:

```
*A:PE-1# show service id 1 sdp detail

===============================================================================
Services: Service Destination Points Details
===============================================================================
-------------------------------------------------------------------------------
```

```
 Sdp Id 32767:4294967295  -(192.0.2.2)
-------------------------------------------------------------------------------
Description     : (Not Specified)
SDP Id             : 32767:4294967295         Type              : BgpVpls
PW-Template Id     : 10
Split Horiz Grp    : (Not Specified)
Etree Root Leaf Tag: Disabled                 Etree Leaf AC     : Disabled
VC Type            : Ether                    VC Tag            : n/a
Admin Path MTU     : 0                        Oper Path MTU     : 8954
Delivery           : GRE
Far End            : 192.0.2.2                Tunnel Far End    : n/a
Oper Tunnel Far End: 192.0.2.2
---snip---

Admin State        : Up                       Oper State        : Up
MinReqd SdpOperMTU : 1514
Acct. Pol          : None                     Collect Stats     : Disabled
Ingress Label      : 524281                   Egress Label      : 524280
---snip---

Last Status Change : 06/23/2021 14:24:54      Signaling         : BGP
---snip---
```

VPLS 2 is configured with PW template binding 20, which prefers provisioned SDPs, so the provisioned SDP 12 is used, as follows:

```
*A:PE-1# show service id 2 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
12:4294967294    BgpVpls  192.0.2.2       Up      Up        524273    524272
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

In VPLS 1, the PW template binding selection can be changed by configuring a non-matching import RT to PW template 10, as follows:

```
# on PE-1:
configure
    service
        vpls "BGP-VPLS-1"
            bgp
                pw-template-binding 10 import-rt "target:64500:999"
                exit
            exit
```

This does not change the selected PW template during service operation and PW template 10 remains in use, as follows:

```
*A:PE-1# show service id 1 sdp detail | match "PW-Template"
PW-Template Id     : 10
```

The following **tools** command forces the system to re-evaluate the PW template binding:

```
*A:PE-1# tools perform service id 1 eval-pw-template 10 allow-service-impact
```

```
eval-pw-template succeeded for Svc 1 32767:4294967295 Policy 10
```

When the PW template binding is re-evaluated, PW template binding 20 is selected and the provisioned SDP 12 is used, as follows:

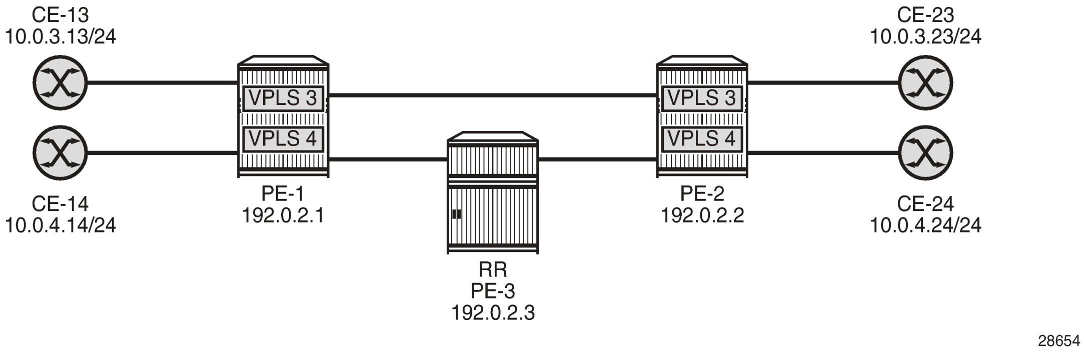
```
*A:PE-1# show service id 1 sdp detail | match "PW-Template"
PW-Template Id     : 20
```

```
*A:PE-1# show service id 1 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
12:4294967293    BgpVpls  192.0.2.2       Up      Up        524281    524280
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

Auto-GRE spoke-SDPs in LDP-VPLS using BGP-AD

Figure 169: LDP-VPLS using BGP-AD with auto-GRE Spoke-SDPs shows the example topology with VPLSs 3 and 4 configured with BGP-AD on PE-1 and PE-2. The BGP configuration is identical to the one for BGP-VPLS.

Figure 169: LDP-VPLS using BGP-AD with auto-GRE Spoke-SDPs



The following T-LDP session is configured between PE-1 and PE-2:

```
# on PE-1:
configure
    router Base
        ldp
            targeted-session
                peer 192.0.2.2
                    no shutdown
                exit
            exit
```

```
# on PE-2:
configure
    router Base
        ldp
            targeted-session
                peer 192.0.2.1
                    no shutdown
                exit
            exit
```

The following T-LDP signaled SDP is configured on PE-1 and PE-2:

```
# on PE-1:
configure
    service
        sdp 120 create
            far-end 192.0.2.2
            keep-alive
                shutdown
            exit
            no shutdown
        exit
```

```
# on PE-2:
configure
    service
        sdp 120 create
            far-end 192.0.2.1
            keep-alive
                shutdown
            exit
            no shutdown
        exit
```

The service configuration on PE-1 and PE-2 is as follows; see chapter LDP VPLS Using BGP Auto-Discovery for a description of BGP-AD in LDP VPLS. PW templates 10 and 20 are the same as in the preceding example.

```
# on PE-1, PE-2:
configure
    service
        pw-template 10 name "PW10-auto-GRE" auto-gre-sdp create
        exit
        pw-template 20 name "PW20-auto-GRE_prefer-prov" prefer-provisioned-sdp
                                                               auto-gre-sdp create
        exit
        vpls 3 name "BGP-AD VPLS-3" customer 1 create
            description "BGP-AD for LDP VPLS with auto-GRE spoke-SDP"
            bgp
                route-distinguisher 64500:3
                route-target export target:64500:3 import target:64500:3
                pw-template-binding 10
                exit
                pw-template-binding 20
                exit
            exit
            bgp-ad
                vpls-id 64500:3
                no shutdown
            exit
            stp
                shutdown
            exit
            sap pxc-10.a:3 create                # SAP to connect to CE-13 (PE-1) or CE-23 (PE-
2)
                no shutdown
            exit
            no shutdown
        exit
        vpls 4 name "BGP-AD VPLS-4" customer 1 create
            description "BGP-AD for LDP VPLS with auto-GRE spoke-SDP pref-prov-SDP"
            bgp
                route-distinguisher 64500:4
                route-target export target:64500:4 import target:64500:4
                pw-template-binding 20
                exit
            exit
            bgp-ad
                vpls-id 64500:4
                no shutdown
            exit
            stp
                shutdown
            exit
            sap pxc-10.a:4 create                # SAP to connect to CE-14 (PE-1) or CE-24 (PE-
2)
                no shutdown
            exit
            no shutdown
        exit
```

PE-1 has received the following L2-VPN BGP-AD routes:

```
*A:PE-1# show router bgp routes l2-vpn bgp-ad
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP L2VPN-AD Routes
===============================================================================
Flag  RouteType                   Prefix                                 MED
      RD                          SiteId                                 Label
      Nexthop                     VeId                    BlockSize      LocalPref
      As-Path                     BaseOffset              vplsLabelBa
                                                          se
-------------------------------------------------------------------------------
u*>i  AutoDiscovery               192.0.2.2               -              0
      64500:3                     -                                      -
      192.0.2.2                   -                       -              100
      No As-Path                  -                       -
u*>i  AutoDiscovery               192.0.2.2               -              0
      64500:4                     -                                      -
      192.0.2.2                   -                       -              100
      No As-Path                  -                       -
-------------------------------------------------------------------------------
Routes : 2
```

```
===============================================================================
```

The following shows the used SDPs on PE-1: BGP-signaled SDP 12 (used by VPLS 1 and 2) and T-LDP-signaled SDPs 120 and 32767.

```
*A:PE-1# show service sdp

============================================================================
Services: Service Destination Points
============================================================================
SdpId  AdmMTU  OprMTU  Far End          Adm  Opr         Del     LSP   Sig
----------------------------------------------------------------------------
12     0       8954    192.0.2.2        Up   Up          GRE     n/a   BGP
120    0       8954    192.0.2.2        Up   Up          GRE     n/a   TLDP
32767  0       8954    192.0.2.2        Up   Up          GRE     n/a   TLDP
----------------------------------------------------------------------------
Number of SDPs : 3
----------------------------------------------------------------------------
Legend: R = RSVP, L = LDP, B = BGP, M = MPLS-TP, n/a = Not Applicable
        I = SR-ISIS, O = SR-OSPF, T = SR-TE, F = FPE
============================================================================
```

The following shows that PW template 10 is used in VPLS 3 and that auto-GRE SDP 32767 is used, with T-LDP signaling:

```
*A:PE-1# show service id 3 sdp detail

===============================================================================
Services: Service Destination Points Details
===============================================================================
-------------------------------------------------------------------------------
 Sdp Id 32767:4294967292  -(192.0.2.2)
-------------------------------------------------------------------------------
Description     : (Not Specified)
SDP Id             : 32767:4294967292         Type              : BgpAd
PW-Template Id     : 10
AGI                : 64500:3                  SDP Bind Source   : bgp-l2vpn
Local AII          : 192.0.2.1
Remote AII         : 192.0.2.2
Split Horiz Grp    : (Not Specified)
Etree Root Leaf Tag: Disabled                 Etree Leaf AC     : Disabled
VC Type            : Ether                    VC Tag            : n/a
Admin Path MTU     : 0                        Oper Path MTU     : 8954
Delivery           : GRE
Far End            : 192.0.2.2                Tunnel Far End    : n/a
Oper Tunnel Far End: 192.0.2.2
---snip---

Admin State        : Up                       Oper State        : Up
---snip---

Last Status Change : 06/23/2021 14:30:31      Signaling         : TLDP
---snip---
```

The following shows that the T-LDP signaled GRE SDP 120 is used in VPLS 4, not the BGP-signaled GRE SDP 12:

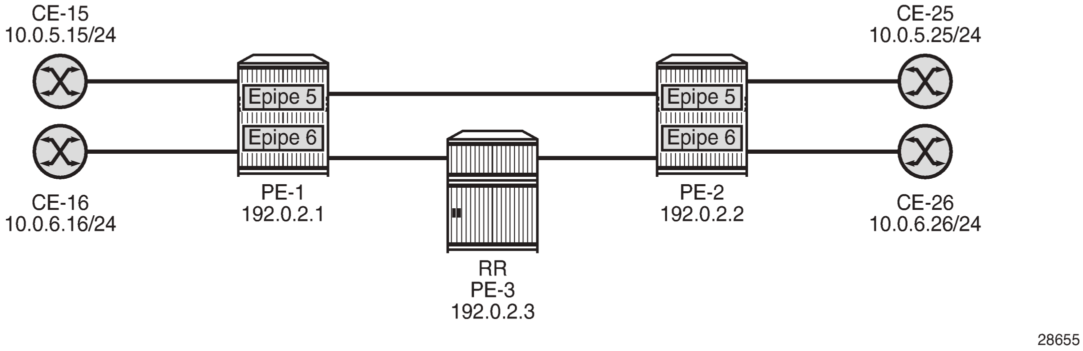
```
*A:PE-1# show service id 4 sdp

===============================================================================
Services: Service Destination Points
```

```
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
120:4294967291   BgpAd    192.0.2.2       Up      Up        524269    524269
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

Auto-GRE spoke-SDPs in BGP-VPWS

Figure 170: BGP-VPWS with auto-GRE spoke-SDPs shows the example topology with BGP-VPWS Epipes 5 and 6 on PE-1 and PE-2. The BGP configuration is identical to the one for BGP-VPLS.

Figure 170: BGP-VPWS with auto-GRE spoke-SDPs



Chapter BGP Virtual Private Wire Services describes the configuration of BGP VPWS. The configuration of Epipes 5 and 6 on PE-1 is as follows:

```
# on PE-1:
configure
    service
        pw-template 10 name "PW10-auto-GRE" auto-gre-sdp create
        exit
        pw-template 20 name "PW20-auto-GRE_prefer-prov" prefer-provisioned-sdp
                                                               auto-gre-sdp create
        exit
        epipe 5 name "BGP-VPWS-5" customer 1 create
            description "BGP-VPWS with auto-GRE spoke-SDP"
            bgp
                route-distinguisher 64500:5
                route-target export target:64500:5 import target:64500:5
                pw-template-binding 10
                exit
                pw-template-binding 20
                exit
            exit
            bgp-vpws
                ve-name "PE-1"
                    ve-id 1
                exit
                remote-ve-name "PE-2"
                    ve-id 2
                exit
                no shutdown
            exit
            sap pxc-10.a:5 create                # SAP to connect to CE-15
                no shutdown
            exit
            no shutdown
        exit
        epipe 6 name "BGP-VPWS-6" customer 1 create
            description "BGP-VPWS with auto-GRE spoke-SDP_prefer provisioned SDP"
            bgp
                route-distinguisher 64500:6
                route-target export target:64500:6 import target:64500:6
                pw-template-binding 20
                exit
            exit
            bgp-vpws
                ve-name "PE-1"
                    ve-id 1
                exit
                remote-ve-name "PE-2"
                    ve-id 2
                exit
                no shutdown
            exit
            sap pxc-10.a:6 create                # SAP to connect to CE-16
                no shutdown
            exit
            no shutdown
        exit
```

The configuration of the Epipes is similar on PE-2, but the VE names and VE IDs are different, as follows:

```
# on PE-2:
configure
    service
        epipe 5 name "BGP-VPWS-5" customer 1 create
            description "BGP-VPWS with auto-GRE spoke-SDP"
            bgp
                route-distinguisher 64500:5
                route-target export target:64500:5 import target:64500:5
                pw-template-binding 10
                exit
                pw-template-binding 20
                exit
            exit
            bgp-vpws
                ve-name "PE-2"
                    ve-id 2
                exit
                remote-ve-name "PE-1"
                    ve-id 1
                exit
                no shutdown
            exit
            sap pxc-10.a:5 create                # SAP to connect to CE-25
                no shutdown
            exit
            no shutdown
        exit
        epipe 6 name "BGP-VPWS-6" customer 1 create
            description "BGP-VPWS with auto-GRE spoke-SDP_prefer provisioned SDP"
            bgp
                route-distinguisher 64500:6
                route-target export target:64500:6 import target:64500:6
                pw-template-binding 20
                exit
            exit
            bgp-vpws
                ve-name "PE-2"
                    ve-id 2
                exit
                remote-ve-name "PE-1"
                    ve-id 1
                exit
                no shutdown
            exit
            sap pxc-10.a:6 create                # SAP to connect to CE-26
                no shutdown
            exit
            no shutdown
        exit
```

PE-1 receives the following BGP-VPWS routes from PE-2:

```
*A:PE-1# show router bgp routes l2-vpn bgp-vpws
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP L2VPN-VPWS Routes
===============================================================================
Flag  RouteType                   Prefix                             MED
      RD                          SiteId                             Label
      Nexthop                     VeId                  BlockSize   LocalPref
      As-Path                     BaseOffset            vplsLabelBa
                                                        se
-------------------------------------------------------------------------------
u*>i  VPWS                        -                      -          0
      64500:5                     -                                  -
      192.0.2.2                   2                      1          100
      No As-Path                  1                      524268
u*>i  VPWS                        -                      -          0
      64500:6                     -                                  -
      192.0.2.2                   2                      1          100
      No As-Path                  1                      524267
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

The following SDP bindings are used on PE-1: the first two are used by BGP-VPLS services VPLS 1 and 2, the third and fourth are used by BGP-AD in LDP VPLS 3 and 4, and the last two are used by BGP-VPWS services Epipe 5 and 6. For the last two, SDP 32766 is auto-created, whereas SDP 12 is provisioned with BGP signaling.

```
*A:PE-1# show service sdp-using

===============================================================================
SDP Using
===============================================================================
SvcId      SdpId              Type   Far End              Opr   I.Label E.Label
```

```
                                                          State
-------------------------------------------------------------------------------
1          12:4294967293      BgpVp* 192.0.2.2            Up    524281  524280
2          12:4294967294      BgpVp* 192.0.2.2            Up    524273  524272
3          32767:4294967292   BgpAd  192.0.2.2            Up    524270  524270
4          120:4294967291     BgpAd  192.0.2.2            Up    524269  524269
5          32766:4294967290   BgpVp* 192.0.2.2            Up    524268  524268
6          12:4294967289      BgpVp* 192.0.2.2            Up    524267  524267
-------------------------------------------------------------------------------
Number of SDPs : 6
-------------------------------------------------------------------------------
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

Epipe 5 uses the following auto-GRE SDP 32766 with BGP signaling:

```
*A:PE-1# show service id 5 sdp detail

===============================================================================
Services: Service Destination Points Details
===============================================================================
-------------------------------------------------------------------------------
 Sdp Id 32766:4294967290  -(192.0.2.2)
-------------------------------------------------------------------------------
Description     : (Not Specified)
SDP Id             : 32766:4294967290         Type              : BgpVpws
PW-Template Id     : 10
VC Type            : Ether                    VC Tag            : n/a
Admin Path MTU     : 0                        Oper Path MTU     : 8954
Delivery           : GRE
Far End            : 192.0.2.2                Tunnel Far End    : n/a
Oper Tunnel Far End: 192.0.2.2
---snip---

Admin State        : Up                       Oper State        : Up
---snip---

Last Status Change : 06/23/2021 14:36:00      Signaling         : BGP
---snip---
```

PW template 20 is used in Epipe 6, so the BGP-signaled GRE SDP 12 is used, as follows:

```
*A:PE-1# show service id 6 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
12:4294967289    BgpVpws  192.0.2.2       Up      Up        524267    524267
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

Auto-GRE spoke-SDPs in dynamic MS-PW spoke-SDP FEC

Chapter Multi-Segment Pseudowire Routing describes the configuration for dynamic MS-PW spoke-SDP FEC.

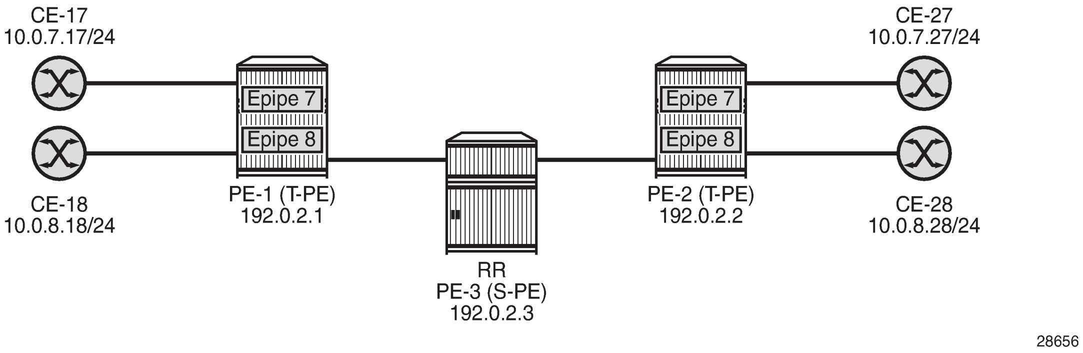
Figure 171: Dynamic MS-PW spoke-SDP FEC with auto-GRE spoke-SDPs shows the example topology with PE-1 and PE-2 as Terminating PEs (T-PEs) and PE-3 as Switching PE (S-PE). Port 1/1/1 on PE-1 (toward PE-2) and port 1/1/2 on PE-2 (toward PE-1) are disabled, as follows:

```
# on PE-1:
configure
    port 1/1/1
        shutdown
```

```
# on PE-2:
configure
    port 1/1/2
        shutdown
```

BGP-VPWS Epipes 7 and 8 are configured on PE-1 and PE-2.

Figure 171: Dynamic MS-PW spoke-SDP FEC with auto-GRE spoke-SDPs



T-LDP is configured between PE-1 and PE-3 and between PE-2 and PE-3, as follows:

```
# on PE-1, PE-2:
configure
    router Base
        ldp
            targeted-session
                peer 192.0.2.3
                    no shutdown
```

```
# on PE-3:
configure
    router Base
        ldp
            targeted-session
                peer 192.0.2.1
                    no shutdown
                exit
                peer 192.0.2.2
                    no shutdown
                exit
```

BGP configuration

BGP is configured for the MS-PW address family. On the T-PEs, an export policy is required to export MS-PW routes; in this case, a default policy matching all the MS-PW routes is configured. The configuration on PE-1 and PE-2 is as follows:

```
# on PE-1, PE-2:
configure
    router Base
        policy-options
            begin
            policy-statement "export ms-pw"
                entry 10
                    from
                        family ms-pw
                    exit
                    action accept
                        origin igp
                    exit
                exit
            exit
            commit
        exit
        bgp
            rapid-withdrawal
            group "WAN"
                family ms-pw
                type internal
                neighbor 192.0.2.3
                    export "export ms-pw"
                exit
            exit
        exit
```

S-PE (and RR) PE-3 is configured for the MS-PW address family and has next-hop-self enabled, as follows:

```
# on PE-3:
configure
    router Base
        bgp
            rapid-withdrawal
            split-horizon
            group "WAN"
                family ms-pw
                next-hop-self
                type internal
                cluster 192.0.2.3
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.2
                exit
            exit
        exit
```

Service configuration

Each T-PE and S-PE is configured with an SPE address. On S-PE PE-3, the SPE address is configured as follows:

```
# on PE-3:
configure
    service
        pw-routing
            spe-address 64500:192.0.2.3
        exit
```

The service configuration on PE-1 is as follows:

```
# on PE-1:
configure
    service
        pw-routing
            spe-address 64500:192.0.2.1
            local-prefix 64500:192.0.2.1 create
                advertise-bgp route-distinguisher 64500:7
                advertise-bgp route-distinguisher 64500:8
            exit
        exit
        pw-template 10 name "PW10-auto-GRE" auto-gre-sdp create
        exit
        pw-template 20 name "PW20-auto-GRE_prefer-prov" prefer-provisioned-sdp
                                                               auto-gre-sdp create
        exit
        epipe 7 name "Epipe-7 MS-PW" customer 1 create
            description "Epipe with dynamic MS-PW spoke-SDP FEC"
            sap pxc-10.a:7 create                # SAP to connect to CE-17
                no shutdown
            exit
            spoke-sdp-fec 7 fec 129 aii-type 2 create
                pw-template-bind 10
                saii-type2 64500:192.0.2.1:7
                taii-type2 64500:192.0.2.2:7
                no shutdown
            exit
            no shutdown
        exit
        epipe 8 name "Epipe-8 MS-PW" customer 1 create
            description "Epipe with dynamic MS-PW spoke-SDP FEC_pref-prov"
            sap pxc-10.a:8 create                # SAP to connect to CE-18
                no shutdown
            exit
            spoke-sdp-fec 8 fec 129 aii-type 2 create
                pw-template-bind 20
                saii-type2 64500:192.0.2.1:8
                taii-type2 64500:192.0.2.2:8
                no shutdown
            exit
            no shutdown
        exit
```

On PE-2, the following service configuration is similar, with different SPE address, local prefix, Source Attachment Individual Identifier (SAII), and Target Attachment Individual Identifier (TAII). The SAIIs for the

Epipes on PE-2 match the TAIIs for the matching Epipes on PE-1 and the TAIIs on PE-2 match the SAIIs on PE-1.

```
# on PE-2:
configure
    service
        pw-routing
            spe-address 64500:192.0.2.2
            local-prefix 64500:192.0.2.2 create
                advertise-bgp route-distinguisher 64500:7
                advertise-bgp route-distinguisher 64500:8
            exit
        exit
        epipe 7 name "Epipe-7 MS-PW" customer 1 create
            description "Epipe with dynamic MS-PW spoke-SDP FEC"
            sap pxc-10.a:7 create                # SAP to connect to CE-27
                no shutdown
            exit
            spoke-sdp-fec 7 fec 129 aii-type 2 create
                pw-template-bind 10
                saii-type2 64500:192.0.2.2:7
                taii-type2 64500:192.0.2.1:7
                no shutdown
            exit
            no shutdown
        exit
        epipe 8 name "Epipe-8 MS-PW" customer 1 create
            description "Epipe with dynamic MS-PW spoke-SDP FEC_pref-prov"
            sap pxc-10.a:8 create                # SAP to connect to CE-28
                no shutdown
            exit
            spoke-sdp-fec 8 fec 129 aii-type 2 create
                pw-template-bind 20
                saii-type2 64500:192.0.2.2:8
                taii-type2 64500:192.0.2.1:8
                no shutdown
            exit
            no shutdown
        exit
```

The following BGP MS-PW routes are used on PE-1:

```
*A:PE-1# show router bgp routes ms-pw
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP MSPW Routes
===============================================================================
Flag  Network                 RD
      Nexthop                 AII-Type2/Preflen
      As-Path
-------------------------------------------------------------------------------
u*>i  64500:192.0.2.2         64500:7
      192.0.2.2               64500:192.0.2.2:0/64
      No As-Path
u*>i  64500:192.0.2.2         64500:8
      192.0.2.2               64500:192.0.2.2:0/64
      No As-Path
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

The following spoke-SDP FECs are used on PE-1. Auto-GRE SDP 32766 is used in Epipe 7 and provisioned LDP-signaled SDP 120 is used in Epipe 8.

```
*A:PE-1# show service spoke-sdp-fec-using

===============================================================================
Service Spoke-SDP-Fec Information
===============================================================================
SvcId      SpokeSdpFec  Oper-SdpBind     SAII-Type2
 Path                    Retries/Secs     TAII-Type2
                        Left
-------------------------------------------------------------------------------
7          7            32766:4294967288 64500:192.0.2.1:7
 n/a                     0/0              64500:192.0.2.2:7
8          8            120:4294967287   64500:192.0.2.1:8
 n/a                     0/0              64500:192.0.2.2:8
-------------------------------------------------------------------------------
Entries found: 2
===============================================================================
```

Auto-GRE SDP 32766 with T-LDP signaling is used in Epipe 7 on PE-1, as follows:

```
*A:PE-1# show service id 7 sdp detail

===============================================================================
Services: Service Destination Points Details
===============================================================================
-------------------------------------------------------------------------------
 Sdp Id 32766:4294967288  -(192.0.2.2)
-------------------------------------------------------------------------------
Description     : (Not Specified)
SDP Id             : 32766:4294967288         Type              : MS-PW
PW-Template Id     : 10
SAII Type2         : 64500:192.0.2.1:7
TAII Type2         : 64500:192.0.2.2:7
VC Type            : Ether                    VC Tag            : n/a
Admin Path MTU     : 0                        Oper Path MTU     : 8954
Delivery           : GRE
Far End            : 192.0.2.2                Tunnel Far End    : n/a
Oper Tunnel Far End: 192.0.2.2
---snip---

Admin State        : Up                       Oper State        : Up
MinReqd SdpOperMTU : 1514
Adv Service MTU    : n/a
Acct. Pol          : None                     Collect Stats     : Disabled
Ingress Label      : 524270                   Egress Label      : 524270
---snip---

Last Status Change : 06/23/2021 14:42:31      Signaling         : TLDP
---snip---
```

The following provisioned GRE SDP with T-LDP signaling is used in Epipe 8 on PE-1:

```
*A:PE-1# show service id 8 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
120:4294967287   MS-PW    192.0.2.2       Up      Up        524269    524269
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

Conclusion

In IP-based networks, auto-GRE spoke-SDPs can be used in VPWS and VPLS services. Manually configured GRE tunnels are not an option in networks—such as LTE networks—where it is common to assign IP addresses dynamically from a pool of addresses, but auto-GRE spoke-SDPs can be applied instead.

Layer 2 Multicast Optimization for EVPN-VXLAN — Assisted Replication

This chapter provides information about Layer 2 Multicast Optimization for EVPN-VXLAN — Assisted Replication.

Topics in this chapter include:

Applicability

This chapter was initially written for SR OS Release 14.0.R4, but the CLI in the current edition is based on SR OS Release 23.3.R3. Layer 2 multicast optimization for EVPN-VXLAN - Assisted Replication (AR) is supported in SR OS Release 14.0.R4, and later.

Overview

Typically, EVPN-VXLAN can use either Ingress Replication (IR) or Protocol Independent Multicast (PIM) for Broadcast, Unknown unicast, and Multicast (BUM) traffic (although SR OS does not support PIM along with EVPN-VXLAN). PIM requires keeping multicast state awareness per subnet per tenant in the core routers, which may not scale. Not all core routers support PIM.

IR inefficiency is usually tolerable in EVPN networks for broadcast and unknown unicast traffic; however, it is not tolerable for multicast traffic:

- Broadcast traffic can be reduced by the proxy-ARP and proxy-ND capabilities supported by EVPN.
- Unknown unicast traffic is greatly reduced in virtualized Data Center (DC) networks where all MAC and IP addresses are learned in the control or management planes. In such cases, unknown MAC addresses are always outside the DC. An **unknown-mac-route** can be enabled to ensure that the unknown unicast traffic is sent only to the DC gateway, which minimizes flooding within the DC.
- Multicast traffic may be an issue for the hypervisors holding the multicast sources, because the hypervisors need to replicate the multicast traffic to the remote VXLAN Tunnel Endpoints (VTEPs). The multicast replication at the hypervisors is a software process and the throughput can be heavily impacted. This is also true when VPLS services are used in the Virtual Service Router (VSR) and many replicas must be done from the VSR. Using a dedicated service node to replicate the multicast traffic on behalf of the hypervisors can help, but the replication capabilities of such service nodes are limited too.

SR OS supports the Assisted Replication (AR) feature for IPv4 VXLAN tunnels (both replicator and leaf functions) in compliance with the non-selective mode described in *draft-ietf-bess-evpn-optimized-ir*. AR is a Layer 2 multicast optimization feature that helps software-based PEs and Network Virtualization Edge

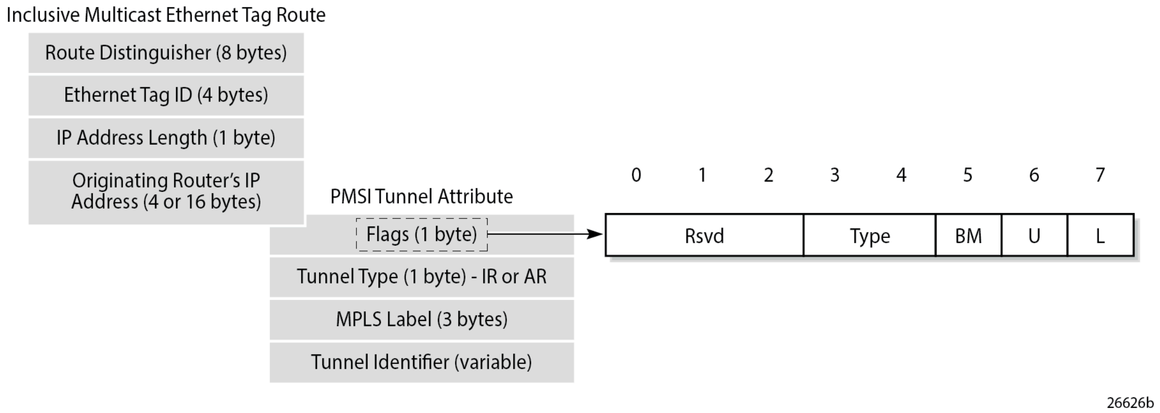
(NVE) devices with low-performance replication capabilities to deliver Broadcast and Multicast (BM) Layer 2 traffic to remote VTEPs in the VPLS.

SR OS nodes support the AR-Replicator (AR-R) and AR-Leaf (AR-L) functions, although not simultaneously on the same service. Nodes configured as AR-L select an AR-R within a service and send all BM packets to this AR-R. AR-Rs replicate traffic to all the VTEPs in the VPLS on behalf of the AR-Ls, so BM traffic is delivered to all VPLS participants without any packet loss caused by performance issues. Unknown unicast packets follow the same path as known unicast packets to avoid packet reordering. Therefore, no AR-R is used for unknown unicast traffic.

When multiple AR-Rs exist in a service, the AR-L performs per-service load-balancing of the BM traffic. The AR-L lists the candidate AR-Rs, ordered by IP address and VXLAN Network Identifier (VNI); candidate 0 having the lowest IP address and VNI. The replicator is selected using a modulo function of the service ID and the number of candidate AR-Rs. For example, assume that VPLS 1 has two candidate AR-Rs: because 1 modulo 2 equals 1, the second AR-R in the list is selected. In case of failure, a new AR-R is selected. If there are no more AR-Rs, the system falls back to IR.

Figure 172: PMSI Tunnel Attribute - Flags shows an EVPN route-type 3, an Inclusive Multicast Ethernet Tag (IMET) route containing a PMSI tunnel attribute with a flags octet. Flag L was already defined in RFC 6514. *Draft-ietf-bess-evpn-optimized-ir* defines additional flags: type, BM, and U. The BM and U flags are used for Pruned Flood Lists (PFL) signaling and they are not supported.

Figure 172: PMSI Tunnel Attribute - Flags


Inclusive Multicast Ethernet Tag Route
Route Distinguisher (8 bytes)
Ethernet Tag ID (4 bytes)
IP Address Length (1 byte)
Originating Router's IP Address (4 or 16 bytes)
PMSI Tunnel Attribute
Flags (1 byte)
Tunnel Type (1 byte) - IR or AR
MPLS Label (3 bytes)
Tunnel Identifier (variable)
0
1
2
3
4
5
6
7
Rsvd
Type
BM
U
L
26626b

The type field has two bits that define the AR role of the advertising router, as follows:

- Type 00 = Regular Network Virtualization Edge (RNVE) - indicates that AR is not supported and IR is applied instead (for backward compatibility)
- Type 01= AR-R
- Type 10 = AR-L
- Type 11 = reserved

The tunnel type in the PMSI tunnel attribute can be configured with the following options for IR and AR:

- Tunnel type 0x06 = (non-optimized) IR, sent by AR-R and AR-L if **ingress-repl-inc-mcast-advertisement** is enabled, which is the default option
- Tunnel type 0x0A = type AR, originated by AR-R

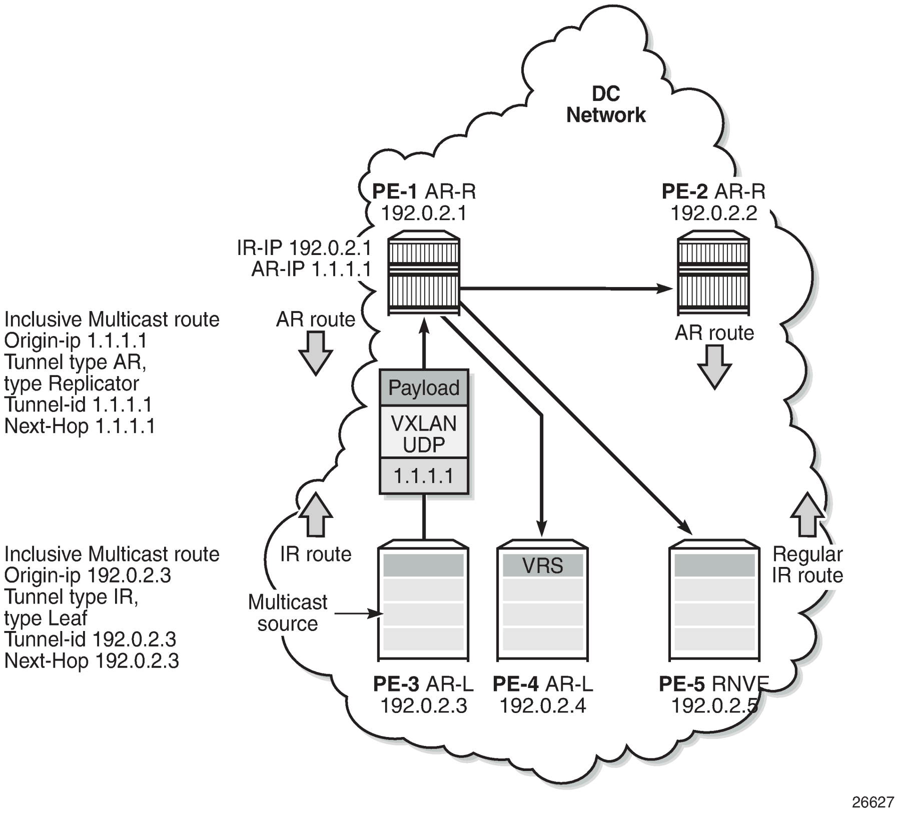
For regular IR routes, the originating router's IP address equals the system IP address. The MPLS label and tunnel identifier must be used as described in RFC 7432. The tunnel identifier is set to a routable address of the PE.

For AR routes, the originating router's IP address and the tunnel identifier are both set to the AR IP address (AR-IP) configured in the **service system vxlan** context. The AR-IP must be previously defined as a loopback interface address in the base router and must be different from the IR IP address (IR-IP).

Note:
If the AR-IP loopback interface is down, the router does not withdraw the AR route. However, the remote AR-Ls is not able to resolve the AR route's BGP next-hop if the AR-IP is no longer propagated in the IGP.

Figure 173: EVPN Assisted Replication for VXLAN shows the example topology with the multicast source connected to a hypervisor PE-3 that acts as AR-L, which sends an IR route containing the system address of PE-3. The AR-R PE-1 sends an AR route that uses AR-IPs instead of IR-IPs; for example, PE-1 has AR-IP 1.1.1.1 and IR-IP 192.0.2.1.

Figure 173: EVPN Assisted Replication for VXLAN



Hypervisor PE-3 sends the BM traffic to the AR-R, which replicates it to all the VTEPs in the VPLS, except to PE-3.

Table 9: Inclusive multicast route information sent by different AR roles shows the inclusive multicast route information sent by each role in an AR-capable service.

Table 9: Inclusive multicast route information sent by different AR roles

AR role	function	inclusive multicast route advertised
AR-R	assists AR-Ls	IR inclusive multicast route (tunnel = 0x06 = IR, IR-IP, type = 0 = none) AR inclusive multicast route (tunnel = 0x0A = AR, AR-IP, type = 1 = AR-R)
AR-L	sends BM only to AR-R	IR inclusive multicast route (tunnel = 0x06 = IR, IR- IP, type = 2 = AR-L)
RNVE	non-AR support	IR inclusive multicast route (tunnel = 0x06 = IR, IR- IP, type = 0 = none)

Unicast traffic (known or unknown) is processed as normal. For BM traffic, the AR-R uses AR or IR based on the IP destination address (DA):

- If IP DA equals the AR-IP, the AR-R replicates to the VTEPs in the VXLAN service, except for the VTEP over which the BM traffic was received.
- If IP DA equals the IR-IP, normal IR forwarding is done.

Non-optimized-IR nodes are unaware of the PMSI tunnel attribute flag definition with the additional flags for AR, so they ignore the information in the flags field.

The *draft-ietf-bess-evpn-optimized-ir* describes the following three types of IR optimizations:

- Non-selective AR - the chosen AR-R replicates the BM traffic to all NVEs in the Ethernet VPN Instance (EVI) except for the source NVE.
- Selective AR - AR-Rs replicate BM traffic to only their AR-L set and the rest of the AR-Rs. Selective AR allows a "multi-stage" AR replication, as opposed to a "single-stage" AR replication.
- Pruned Flood Lists - AR-Ls can signal PFL flags to be pruned from the flood lists for BM or for unknown unicast traffic. PFL may be used in combination with AR.

This chapter only describes non-selective AR.

Configure AR-R and AR-L

The AR-IP is configured on the AR-R, as follows:

```
*A:PE-1# configure service system vxlan assisted-replication-ip ?
  - assisted-replication-ip <ip-address>
  - no assisted-replication-ip

 <ip-address>         : a.b.c.d
```

The AR-IP is the IPv4 address of a loopback interface in the base router instance. When attempting to configure an AR-IP and the loopback address does not exist, the following error message is raised:

```
*A:PE-1# configure service system vxlan assisted-replication-ip 1.1.1.1
MINOR: SVCMGR #8110 Cannot change assisted-replicated address
 - loopback interface with address does not exist
```

The AR types replicator and leaf are configured in a VPLS with the following command:

```
*A:PE-1# configure service vpls 10 vxlan instance 1 vni 1 assisted-replication ?
  - assisted-replication {replicator|leaf} [replicator-activation-time <seconds>]
  - no assisted-replication

 <replicator|leaf>    : replicator|leaf
 <seconds>            : [1..255]
```

When attempting to configure an AR-R before the AR-IP is set, the following error is raised:

```
*A:PE-1# configure service vpls 10 name "VPLS 10" customer 1 create vxlan instance 1 vni 1
 create assisted-replication replicator
MINOR: SVCMGR #8111 Cannot change assisted-replicated role
 - assisted replicator ip not set
```

The AR type (AR-R or AR-L) cannot be changed while being used by any BGP-EVPN service. The following error is raised in such a case:

```
*A:PE-1# configure service vpls 10 vxlan instance 1 vni 1 assisted-replication leaf
MINOR: SVCMGR #8111 Cannot change assisted-replicated role - Evpn not shut
```

The assisted-replication-time can only be configured on leaf nodes. The following error is raised after an attempt to configure the assisted-replication-time on an AR-R:

```
*A:PE-1# configure service vpls 10 vxlan instance 1 vni 1 assisted-replication replicator
 replicator-activation-time 5
MINOR: SVCMGR #8112 Cannot change replicator activation time - valid only on leaf
```

The **replicator-activation-time** can optionally be activated, and works as follows. When the router creates an AR-R destination for the first time, the assisted-replication-timer must expire before this AR-R destination is eligible as candidate AR-R to forward BM traffic. Upon timer expiration, the router runs the AR-R selection (service ID modulo the number of AR-Rs provides the selected AR-R in the ordered list of candidate AR-Rs). The AR-R EVPN destination is created as "BM" and the destinations to the remaining nodes is shown as "U".

The **replicator-activation-time** allows the AR-R some time to program the leaf VTEPs in the following cases:

- Configuration of a new AR-R
- AR-R rebooting
- AR-R going operationally down and up again

If the timer is zero (default value), the AR-R may receive packets from a VTEP that has not been programmed yet, in which case the AR-R drops the packets.

With the AR-Rs and AR-Ls configured, IMET AR routes can be exchanged. IR can be enabled or disabled independently of the AR configuration. The following command is required to enable IR inclusive multicast routes, and is enabled by default:

```
*A:PE-1# configure service vpls 10 bgp-evpn ingress-repl-inc-mcast-advertisement
```

BGP-EVPN routes

By default, IR is enabled in BGP-EVPN. The following IMET IR route is sent from PE-5 (RNVE) to Route Reflector (RR) PE-1. The flags in the PMSI Tunnel Attribute (PTA) indicate that regular IR is used to forward BUM traffic (tunnel type: 0x06). The AR type is "None", because AR is disabled on PE-5. The IR-IP 192.0.2.5 is used as next-hop, originator IP address, and tunnel endpoint. The MPLS label corresponds to the VNI.

```
*A:PE-5# show debug
debug
    router "Base"
        bgp
            update
```

```
On PE-5:
12 2023/07/07 09:56:26.369 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 77
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.5
        Type: EVPN-INCL-MCAST Len: 17 RD: 192.0.2.5:1, tag: 0, orig_addr len: 32, orig_addr:
 192.0.2.5
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:1
        bgp-tunnel-encap:VXLAN
    Flag: 0xc0 Type: 22 Len: 9 PMSI:
        Tunnel-type Ingress Replication (6)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf: not required]
        MPLS Label 1
        Tunnel-Endpoint 192.0.2.5
"
```

A similar IMET IR route is sent from AR-L PE-3 toward RR PE-1, as follows. The difference is that the flags indicate that PE-3 is configured as an AR-L for the VPLS. The IR-IP 192.0.2.3 is used as next-hop, originator address, and tunnel endpoint.

```
On PE-3:
8 2023/07/07 09:55:54.883 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 77
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-INCL-MCAST Len: 17 RD: 192.0.2.3:1, tag: 0, orig_addr len: 32, orig_addr:
 192.0.2.3
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:1
        bgp-tunnel-encap:VXLAN
    Flag: 0xc0 Type: 22 Len: 9 PMSI:
        Tunnel-type Ingress Replication (6)
        Flags: (0x10)[Type: AR Leaf BM: 0 U: 0 Leaf: not required]
        MPLS Label 1
        Tunnel-Endpoint 192.0.2.3
"
```

The IMET IR routes contain the system IP addresses of the nodes, not the AR-IPs.

The following AR route is advertised from AR-R PE-1. The tunnel type is AR and the flags indicate that PE-1 is configured as AR-R. The AR-IP 1.1.1.1 is the next-hop address, the originator address, and the tunnel endpoint.

```
On PE-1:
4 2023/07/07 09:55:29.613 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 77
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 1.1.1.1
        Type: EVPN-INCL-MCAST Len: 17 RD: 192.0.2.1:1, tag: 0, orig_addr len: 32, orig_addr:
 1.1.1.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:1
        bgp-tunnel-encap:VXLAN
    Flag: 0xc0 Type: 22 Len: 9 PMSI:
        Tunnel-type Assisted Replication (10)
        Flags: (0x8)[Type: AR Replicator BM: 0 U: 0 Leaf: not required]
        MPLS Label 1
        Tunnel-Endpoint 1.1.1.1
"
```

Besides IMET AR routes, PE-1 may also advertise IMET IR routes to the other nodes using IR-IP 192.0.2.1 (system IP address). By default, BGP-EVPN has IR enabled. For example, the following IMET IR route is advertised to PE-4:

```
On PE-1:
3 2023/07/07 09:55:29.613 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 77
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.1
        Type: EVPN-INCL-MCAST Len: 17 RD: 192.0.2.1:1, tag: 0, orig_addr len: 32, orig_addr:
 192.0.2.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:1
        bgp-tunnel-encap:VXLAN
    Flag: 0xc0 Type: 22 Len: 9 PMSI:
        Tunnel-type Ingress Replication (6)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf: not required]
        MPLS Label 1
        Tunnel-Endpoint 192.0.2.1
"
```

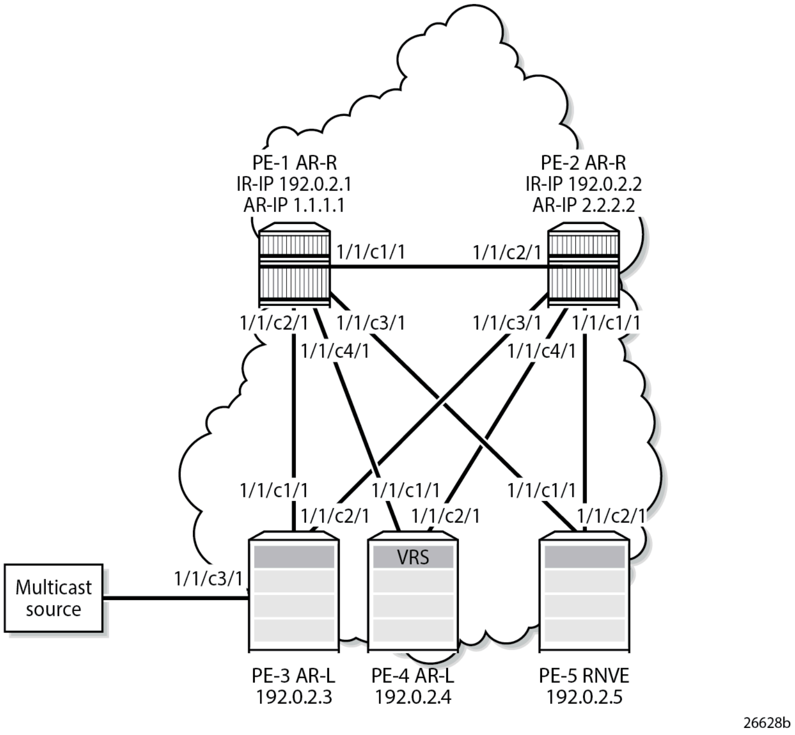
The following IMET routes have been received by PE-4:

```
*A:PE-4# show router bgp routes evpn incl-mcast
===============================================================================
 BGP Router ID:192.0.2.4        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Flag  Route Dist.         OrigAddr
      Tag                 NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.1:1         1.1.1.1
      0                   1.1.1.1

u*>i  192.0.2.1:1         192.0.2.1
      0                   192.0.2.1

u*>i  192.0.2.2:1         2.2.2.2
      0                   2.2.2.2

u*>i  192.0.2.2:1         192.0.2.2
      0                   192.0.2.2

u*>i  192.0.2.3:1         192.0.2.3
      0                   192.0.2.3

u*>i  192.0.2.5:1         192.0.2.5
      0                   192.0.2.5

-------------------------------------------------------------------------------
Routes : 6
===============================================================================
```

Configuration

Figure 174: Example topology shows the example topology with PE-1 and PE-2 as AR-R nodes, PE-3 and PE-4 as AR-L nodes, and PE-5 as RNVE node. The multicast source is connected to PE-3, which is a low-performance node. PE-1 acts as an RR for all nodes.

Figure 174: Example topology



The initial configuration on the nodes includes:

- Cards, MDAs, ports
- Router interfaces between the nodes
- IS-IS as IGP (alternatively, OSPF can be used)

BGP is configured for address family EVPN with RR PE-1. The BGP configuration on PE-1 is as follows:

```
On PE-1:
configure
    router
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "DC"
                family evpn
                cluster 192.0.2.1
                peer-as 64500
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
        exit
```

The BGP configuration on the other nodes is as follows:

```
On the other PEs:
configure
    router
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "DC"
                family evpn
                peer-as 64500
                neighbor 192.0.2.1
                exit
            exit
        exit
```

VPLS 10 is configured on all nodes. PE-1 is configured as AR-R with AR-IP 1.1.1.1, which must be configured as loopback IPv4 address in the base router and as AR-IP that can be shared between services. When attempting to configure an AR-IP with an IP address that does not exist in the base router, the following error is raised:

```
*A:PE-1# configure service system vxlan assisted-replication-ip 1.1.1.1
MINOR: SVCMGR #8110 Cannot change assisted-replicated address
 - loopback interface with address does not exist
```

First, a loopback interface is configured in the base router. The IP address needs to be routable and, in this example, an export policy exporting this IP address is configured in IS-IS. Alternatively, a static route can be configured or an additional IS-IS passive interface can be configured for the loopback interface. The IP address is then configured as AR-IP in the **service system vxlan** context. PE-1 is configured as AR-R for VPLS 10, as follows:

```
On PE-1:
configure
    router
        interface "AR-IP"
            address 1.1.1.1/32
            loopback
        exit
        policy-options
            begin
            prefix-list "AR-IP"
                prefix 1.1.1.1/32 exact
            exit
            policy-statement "export_AR-IP"
                entry 10
                    from
                        prefix-list "AR-IP"
                    exit
                    action accept
                    exit
                exit
            exit
            commit
        exit
        isis
            export "export_AR-IP"
        exit
    exit
    service
        system
            vxlan
                assisted-replication-ip 1.1.1.1
            exit
        exit
        vpls 10 name "VPLS 10" customer 1 create
            vxlan instance 1 vni 1 create
                assisted-replication replicator
            exit
            bgp
            exit
            bgp-evpn
                evi 1
                vxlan
                    no shutdown
                exit
            exit
            no shutdown
        exit
    exit
```

The configuration is similar on PE-2, but with AR-IP 2.2.2.2 instead of 1.1.1.1.

PE-3 and PE-4 are configured as AR-L nodes for VPLS 10. No AR-IP needs to be configured. The configuration of VPLS 10 on PE-3 is as follows:

```
On PE-3:
configure
    service
        vpls 10 name "VPLS 10" customer 1 create
            vxlan instance 1 vni 1 create
                assisted-replication leaf
            exit
            bgp
            exit
            bgp-evpn
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            sap 1/1/c3/1 create    # sap for ingress traffic from STC
            exit
            sap 1/2/c1/1:1 create    # sap for egress traffic to VPLS 10
            exit
            no shutdown
```

Multicast traffic enters SAP 1/1/c3/1, whereas receiving hosts can be connected to other SAPs, such as SAP 1/2/c1/1:1. The configuration of VPLS 10 on PE-4 is similar, but no multicast source is connected. When a node is configured as AR-L, optionally the **replicator-activation-time** can be configured to define the waiting time before the leaf can begin sending multicast traffic to a new replicator or a replicator that was rebooted. The default is zero seconds, in which case the AR-L starts sending packets to the AR-R without delay. Nokia recommends configuring a **replicator-activation-time** value different from zero.

```
*A:PE-3# configure service vpls 10 vxlan instance 1 vni 1 assisted-replication leaf ?
  - assisted-replication {replicator|leaf} [replicator-activation-time <seconds>]
  - no assisted-replication

 <replicator|leaf>    : replicator|leaf
 <seconds>            : [1..255]
```

PE-5 is configured as an RNVE node for VPLS 10, as follows:

```
On PE-5:
configure
    service
        vpls 10 name "VPLS 10" customer 1 create
            vxlan instance 1 vni 1 create
            exit
            bgp
            exit
            bgp-evpn
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            sap 1/2/c1/1:1 create    # sap for egress traffic to VPLS 10
            exit
            no shutdown
```

BGP-EVPN IMET routes are exchanged between the nodes. The following IMET routes are used on AR-L PE-3, with two routes from each AR-R: one IR route with BGP next-hop 192.0.2.x and one AR route with BGP next-hop x.x.x.x (with x equal to 1 or 2).

```
*A:PE-3# show router bgp routes evpn incl-mcast
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Flag  Route Dist.         OrigAddr
      Tag                 NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.1:1         1.1.1.1
      0                   1.1.1.1

u*>i  192.0.2.1:1         192.0.2.1
      0                   192.0.2.1

u*>i  192.0.2.2:1         2.2.2.2
      0                   2.2.2.2

u*>i  192.0.2.2:1         192.0.2.2
      0                   192.0.2.2
```

```
u*>i  192.0.2.4:1         192.0.2.4
      0                   192.0.2.4

u*>i  192.0.2.5:1         192.0.2.5
      0                   192.0.2.5

-------------------------------------------------------------------------------
Routes : 6
===============================================================================
```

When the AR-R has no local attachment circuits, such as SAPs or SDP-bindings, it should not generate regular IR routes. This can be controlled by disabling **ingress-repl-inc-mcast-advertisement** on PE-1 and PE-2, as follows:

```
On PE-1 and PE-2:
configure
    service
        vpls 10
            bgp-evpn
                vxlan bgp 1 vxlan-instance 1 shutdown
                no ingress-repl-inc-mcast-advertisement
                vxlan bgp 1 vxlan-instance 1 no shutdown
```

When IR is disabled on the AR-Rs, no IR routes are sent to the other nodes and PE-3 only sees the AR routes from PE-1 and PE-2, as follows:

```
*A:PE-3# show router bgp routes evpn incl-mcast
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Flag  Route Dist.         OrigAddr
      Tag                 NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.1:1         1.1.1.1
      0                   1.1.1.1

u*>i  192.0.2.2:1         2.2.2.2
      0                   2.2.2.2

u*>i  192.0.2.4:1         192.0.2.4
      0                   192.0.2.4

u*>i  192.0.2.5:1         192.0.2.5
      0                   192.0.2.5

-------------------------------------------------------------------------------
Routes : 4
===============================================================================
```

The detailed information about the AR route sent by AR-R PE-1 can be shown with the following command. The AR tunnel has endpoint 1.1.1.1.

```
*A:PE-3# show router bgp routes evpn incl-mcast rd 192.0.2.1:1 hunt
```

```
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network        : n/a
Nexthop        : 1.1.1.1
Path Id        : None
From           : 192.0.2.1
---snip---
Community      : target:64500:1 bgp-tunnel-encap:VXLAN
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.1
Flags          : Used Valid Best IGP
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : INCL-MCAST
Tag            : 0
Originator IP  : 1.1.1.1
Route Dist.    : 192.0.2.1:1
Route Tag      : 0
---snip---
-------------------------------------------------------------------------------
PMSI Tunnel Attributes :
Tunnel-type    : Assisted Replication
Flags          : Type: AR-Replicator(1) BM: 0 U: 0 Leaf: not required
MPLS Label     : VNI 1
Tunnel-Endpoint: 1.1.1.1
-------------------------------------------------------------------------------

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

The following command shows the VXLAN destinations for VPLS 10 on PE-3:

```
*A:PE-3# show service id 10 vxlan destinations

===============================================================================
Egress VTEP, VNI (Instance 1)
===============================================================================
VTEP Address                                      Egress VNI Oper  Mcast Num
                                                             State       MACs
-------------------------------------------------------------------------------
1.1.1.1                                           1          Up    BM    0
2.2.2.2                                           1          Up    -     0
192.0.2.4                                         1          Up    U     1
192.0.2.5                                         1          Up    U     1
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 4
-------------------------------------------------------------------------------
```

```
---snip---
===============================================================================
```

PE-3 is configured as AR-L and no **replicator-activation-time** is defined (default). Four egress VTEPs are listed: the system IP addresses are used for IR routes and the AR-IPs are used for AR routes. All BM traffic is forwarded to AR-IP 1.1.1.1 on PE-1. The AR-R in use is selected by the modulo operation on the service ID (10). In this example, two AR-Rs are available, and the service ID modulo 2 equals zero: 10 mod 2 = 0. This is the lowest possible outcome, so the first AR-R in the ordered candidate list is used. The AR-Rs are ordered by IP and VNI, with candidate 0 the lowest IP and VNI.

```
*A:PE-3# show service id 10 vxlan assisted-replication replicator

===============================================================================
Vxlan AR Replicator Candidates
===============================================================================
Inst  VTEP Address           Egr VNI  In Use  In Candidate List Pending Time
-------------------------------------------------------------------------------
1     1.1.1.1                1        yes     yes               0
1     2.2.2.2                1        no      yes               0
-------------------------------------------------------------------------------
Number of entries : 2
-------------------------------------------------------------------------------
===============================================================================
```

Within a service, no load-sharing is done between the AR-Rs. However, different AR-Rs can be used for different services.

- If PE-3 were configured as AR-L in VPLS 11, the calculation would be as follows: 11 mod 2 = 1; therefore, the second AR-R in the list would be selected.
- When three AR-Rs were available for VPLS 11, the calculation would be: 11 mod 3 = 2, so the third AR-R in the list would be used.

In case different VNIs are configured for the AR-Rs, the lowest IP address is always higher in the list, even when the VNI is higher. This can be shown when the VPLS VXLAN configuration on PE-1 is modified with VNI 99 instead of VNI 1, as follows:

```
On PE-1:
configure service vpls 10 bgp-evpn vxlan bgp 1 vxlan-instance 1 shutdown
configure service vpls 10 bgp-evpn no vxlan
configure service vpls 10 no vxlan instance 1 vni 1
configure service vpls 10 vxlan instance 1 vni 99 create assisted-replication replicator
configure service vpls 10 bgp-evpn vxlan bgp 1 vxlan-instance 1 no shutdown
```

The list of AR-Rs on PE-3 shows that the first entry is the VTEP with the lowest IP address (1.1.1.1), even though the VNI 99 is higher than 1:

```
*A:PE-3# show service id 10 vxlan assisted-replication replicator

===============================================================================
Vxlan AR Replicator Candidates
===============================================================================
Inst  VTEP Address           Egr VNI  In Use  In Candidate List Pending Time
-------------------------------------------------------------------------------
1     1.1.1.1                99       yes     yes               0
1     2.2.2.2                1        no      yes               0
-------------------------------------------------------------------------------
Number of entries : 2
-------------------------------------------------------------------------------
```

```
===============================================================================
```

Note:
If the AR-IP loopback interface is down, BGP does not withdraw the AR route. When the route to the AR-IP is signaled using IGP, the route is removed from the routing table and the AR-L selects another AR-R. However, when a static route is defined for the AR-IP, a black-hole exists when the AR-IP interface is down.

PE-5 is configured as an RNVE node that signals regular IMET IR routes and is unaware of the AR-R and AR-L roles in the EVI. RNVE nodes ignore IMET AR routes. In the example, only PE-3, PE-4, and PE-5 send IMET IR updates, so the list of VTEP addresses on PE-5 only contains PE-3 and PE-4, as follows:

```
*A:PE-5# show service id 10 vxlan destinations

===============================================================================
Egress VTEP, VNI (Instance 1)
===============================================================================
VTEP Address                                        Egress VNI Oper  Mcast Num
                                                               State       MACs
-------------------------------------------------------------------------------
192.0.2.3                                           1          Up    BUM   0
192.0.2.4                                           1          Up    BUM   0
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 2
-------------------------------------------------------------------------------
---snip---
===============================================================================
```

The RNVE is unaware of AR-Rs; therefore, the list of AR-Rs is empty on PE-5:

```
*A:PE-5# show service id 10 vxlan assisted-replication replicator

===============================================================================
Vxlan AR Replicator Candidates
===============================================================================
Inst  VTEP Address           Egr VNI  In Use  In Candidate List Pending Time
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

Verification of multicast traffic

The multicast source connected to PE-3 generates multicast traffic. PE-3 acts as AR-L and forwards the multicast packets to AR-R PE-1. In this example topology, multicast traffic enters port 1/1/c3/1 on PE-3 and is forwarded to egress port 1/1/c1/1 toward PE-1. Port statistics are cleared and traffic is generated, then the port statistics are verified.

```
*A:PE-3# show port 1/1/c1/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c1/1                                 67                            7397
```

```
                                      48901                        75700070
===============================================================================
*A:PE-3# show port 1/1/c2/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c2/1                                 56                            6460
                                         57                            6587
===============================================================================
*A:PE-3# show port 1/1/c3/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c3/1                              48834                        73251000
                                          0                               0
===============================================================================
```

Besides the multicast traffic, IGP signaling is sent and received on the network interfaces. This explains why the counters on the network interface 1/1/c1/1 toward PE-1 show a slightly higher value than on the interface 1/1/c3/1 toward the multicast source. No multicast traffic is forwarded to PE-2, which is an AR-R candidate, but not used. AR-L PE-3 selected PE-1 for VPLS 10.

When the AR-R PE-1 receives the multicast traffic from PE-3, it forwards the traffic to PE-4 and PE-5 within the VXLAN service. The VXLAN information for VPLS 10 on PE-1 shows that PE-2 is not in the list of egress VTEPs. The reason is that PE-2 does not have any SAPs or SDP-bindings and no IMET IR route is sent by PE-2 because **ingress-repl-inc-mcast-advertisement** is disabled.

```
*A:PE-1# show service id 10 vxlan destinations

===============================================================================
Egress VTEP, VNI (Instance 1)
===============================================================================
VTEP Address                                        Egress VNI Oper  Mcast Num
                                                               State       MACs
-------------------------------------------------------------------------------
192.0.2.3                                           1          Up    BUM   1
192.0.2.4                                           1          Up    BUM   1
192.0.2.5                                           1          Up    BUM   1
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 3
-------------------------------------------------------------------------------
---snip---
===============================================================================
```

AR-R PE-1 receives the multicast traffic from PE-3 on port 1/1/c2/1 and forwards it to the egress ports 1/1/c3/1 toward PE-5 and 1/1/c4/1 toward PE-4, as follows. No multicast traffic needs to be forwarded to egress port 1/1/c1/1 toward PE-2. Source squelching ensures that the traffic is not sent back to the originator AR-L PE-3. PE-1 has no local SAPs or SDP-bindings.

```
*A:PE-1# show port 1/1/c1/1 statistics

===============================================================================
Port Statistics on Slot 1
```

```
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c1/1                                 66                            7252
                                         70                            7779
===============================================================================
*A:PE-1# show port 1/1/c2/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c2/1                              48902                        75700143
                                         66                            7261
===============================================================================
*A:PE-1# show port 1/1/c3/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c3/1                                 69                            7434
                                      48902                        75700238
===============================================================================
*A:PE-1# show port 1/1/c4/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c4/1                                 68                            7420
                                      48904                        75700388
===============================================================================
```

An egress AR-L or RNVE node performs regular egress BUM forwarding procedures. Packets are replicated to local SAPs or SDP-bindings, but not to VXLAN-bindings.

AR-R failure scenarios

When the AR-IP interface on the used AR-R is down for any kind of reason, the route to this AR-IP is removed from the routing table on AR-L PE-3, and PE-3 selects AR-R PE-2. To simulate an AR-R failure, the AR-IP interface on PE-1 is disabled, as follows:

```
*A:PE-1# configure router interface "AR-IP" shutdown
```

After a while, the routing table on PE-3 does not contain an entry for prefix 1.1.1.1/32 anymore, as follows:

```
*A:PE-3# show router route-table 1.1.1.1/32

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
```

```
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
No. of Routes: 0
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

AR-R PE-1 is not eligible anymore when the AR-IP is not reachable. PE-2 is now selected as AR-R, so BM traffic is forwarded to PE-2. Log 99 on PE-3 shows the change in AR-R from PE-1 to PE-2, as follows:

```
On PE-3:
117 2023/07/07 10:26:01.965 UTC MINOR: SVCMGR #2090 Base
"Assisted replicator in service 10 changed to VTEP 2.2.2.2, Egress VNI 1 vxlan-instance 1."
```

The VXLAN destinations for VPLS 10 on PE-3 do not include VTEP 1.1.1.1 anymore, as follows:

```
*A:PE-3# show service id 10 vxlan destinations

===============================================================================
Egress VTEP, VNI (Instance 1)
===============================================================================
VTEP Address                                        Egress VNI Oper  Mcast Num
                                                               State       MACs
-------------------------------------------------------------------------------
2.2.2.2                                             1          Up    BM    0
192.0.2.4                                           1          Up    U     1
192.0.2.5                                           1          Up    U     0
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 3
-------------------------------------------------------------------------------
---snip---
===============================================================================
```

Only PE-2 is listed as AR-R for VPLS 10 on PE-3, and PE-2 is the selected AR-R for VPLS 10, as follows:

```
*A:PE-3# show service id 10 vxlan assisted-replication replicator

===============================================================================
Vxlan AR Replicator Candidates
===============================================================================
Inst  VTEP Address           Egr VNI  In Use  In Candidate List Pending Time
-------------------------------------------------------------------------------
1     2.2.2.2                1        yes     yes               0
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

Incoming multicast traffic on port 1/1/c3/1 on PE-3 is now forwarded to port 1/1/c2/1 toward PE-2, as follows:

```
*A:PE-3# show port 1/1/c1/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
```

```
Id                            Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c1/1                                 69                            7705
                                         70                            7880
===============================================================================
*A:PE-3# show port 1/1/c2/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c2/1                                 59                            6986
                                      48178                        74592682
===============================================================================
*A:PE-3# show port 1/1/c3/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c3/1                              48120                        72180000
                                          0                               0
===============================================================================
```

When the AR-IP interface on AR-R PE-2 is also disabled, no AR-R is available anymore and PE-3 reverts to IR instead.

```
*A:PE-2# configure router interface "AR-IP" shutdown
```

The following log 99 message on AR-L PE-3 indicates that there is no AR-R anymore (VTEP 0.0.0.0, Egress VNI 0).

```
On PE-3:
125 2023/07/07 10:29:38.545 UTC MINOR: SVCMGR #2090 Base
"Assisted replicator in service 10 changed to VTEP 0.0.0.0, Egress VNI 0 vxlan-instance 1."
```

The list of VXLAN destinations for VPLS 10 on PE-3 does not include any AR-R (VTEP 1.1.1.1 or 2.2.2.2) anymore, as follows:

```
*A:PE-3# show service id 10 vxlan destinations

===============================================================================
Egress VTEP, VNI (Instance 1)
===============================================================================
VTEP Address                                        Egress VNI Oper  Mcast Num
                                                               State       MACs
-------------------------------------------------------------------------------
192.0.2.4                                           1          Up    BUM   0
192.0.2.5                                           1          Up    BUM   0
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 2
-------------------------------------------------------------------------------
---snip---
===============================================================================
```

```
*A:PE-3# show service id 10 vxlan assisted-replication replicator
```

```
===============================================================================
Vxlan AR Replicator Candidates
===============================================================================
Inst  VTEP Address           Egr VNI  In Use  In Candidate List Pending Time
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

In this case, IR is done for all BUM traffic toward PE-4 and PE-5.

Conclusion

AR uses replicators to forward broadcast and multicast traffic on behalf of less-performing nodes that are configured as AR-Ls. AR is primarily used for L2 multicast optimization in data centers, but may also be used in any network using overlay EVPN-VXLAN tunnels.

LDP VPLS Using BGP Auto-Discovery

This chapter provides information about LDP VPLS using BGP Auto-Discovery.

Topics in this chapter include:

Applicability

This chapter was initially written for SR OS Release 9.0.R3. The CLI in this edition is based on SR OS Release 20.10.R2. There are no prerequisites for this configuration.

Knowledge of BGP-auto-discovery RFC 6074 architecture and functionality, RFC 4447 Pseudo-wire set-up using label distribution protocol is assumed throughout this chapter, as well as knowledge of Multi-Protocol BGP (MP-BGP).

Overview

MPLS-based Virtual Private LAN Services (VPLS) may have many different provisioning models to allow the signaling of pseudowires between Provider Edge (PE) routers containing VPLS instances.

Network Management System (NMS) provisioning using Label Distribution Protocol (LDP) signaling is a well understood method of provisioning of Layer 2 VPLS services as described in RFC 4762. This relies on the provisioning of pseudowires between VPLS instances using LDP signaling with a common Virtual Circuit (VC) identifier within the label mapping message to instantiate pseudowires.

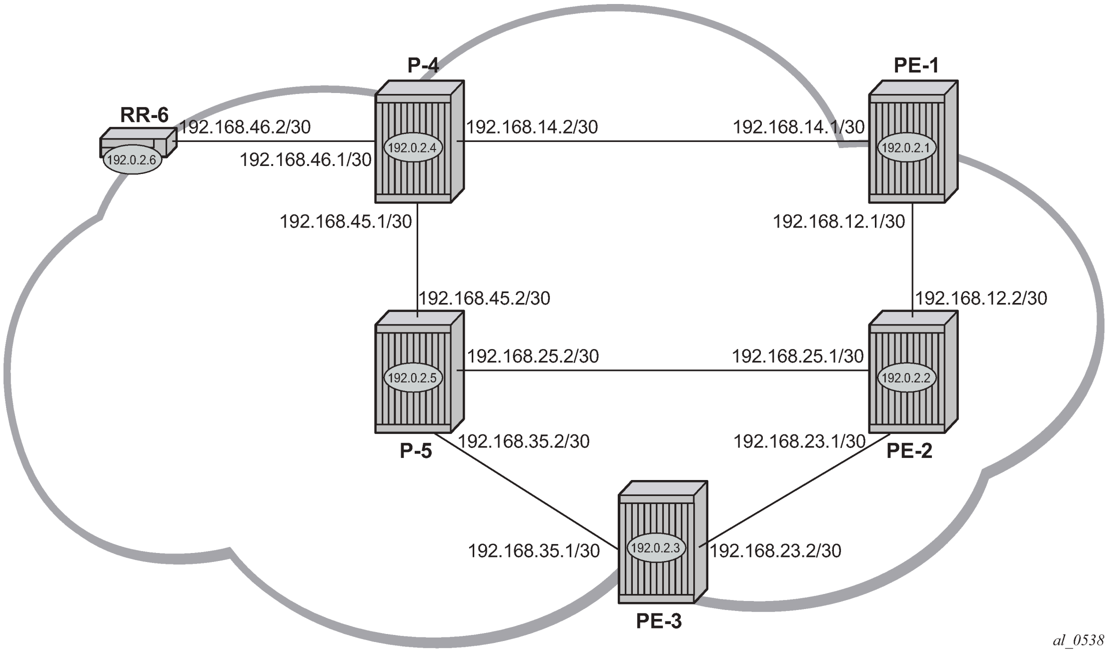
Border Gateway Protocol (BGP) Auto-Discovery (RFC 6074) is an alternative method of provisioning of Layer 2 PE routers containing VPLS service instances to those described above where PEs in a common VPLS instance are automatically discovered using BGP Auto-Discovery (BGP-AD) techniques.

Each PE router advertises the presence of VPLS instances to other PE routers using defined parameters within a BGP update message.

LDP is used as the pseudowire signaling protocol and relies on the auto-discovery of VPLS endpoints to instantiate pseudowires instead of manually provisioning virtual circuits. Locally configured parameters, along with BGP learned parameters, are used to determine local and remote VPLS endpoints, which are used by LDP to signal service labels to peer routers.

Figure 175: Example topology shows the example topology with six SR OS nodes located in the same autonomous system (AS).There are three PEs and RR-6 will act as a route reflector for the AS. The PE routers are all VPLS-aware. The provider (P) routers are VPLS-unaware and do not take part in the BGP process. A full mesh VPLS between PE-1, PE-2, and PE-3 is described.

Figure 175: Example topology



The following configuration tasks are completed as a prerequisite:

- IS-IS or OSPF is enabled on all network interfaces between each of the PE/P routers and route reflector RR-6.
- MPLS is configured on all interfaces between PE and P routers; MPLS is not required between P-4 and RR-6.
- LDP is configured on interfaces between PE and P routers; LDP is not required between P-4 and RR-6.
- The RSVP protocol must be enabled.

BGP-AD

In this architecture, a VPLS service is a collection of local VPLS instances present on a number of PEs in a provider network. In this context, VPLS-aware devices are PE routers. Each VPLS instance has a unique identifier known as the VPLS identifier (VPLS-ID). All PEs that have this VPLS instance present will have a common VPLS-ID configured.

Each VPLS instance within a PE contains a Virtual Switching Instance (VSI). The VPLS attachment circuits and pseudowires are associated with the VSI. Each VSI within a VPLS has a unique identifier called the VSI identifier (VSI-ID) and is a concatenation of the VPLS-ID plus an IP address, usually the system IP address.

The PEs communicate with each other at the control plane level by means of BGP updates containing BGP Layer 2 Network Layer Reachability Information (NLRI). Each update contains enough information for a PE to determine the presence of other local VPLS instances on peering PEs. In turn, this allows peer PE routers to set up pseudowire connectivity using LDP signaling for data flow between peers containing a local VPLS within the same VPLS instances.

Each update contains parameters usually associated with Multi-Protocol BGP updates:

- NLRI encoded as route target (RT)—usually the VPLS-ID—and PE system address.
- Next-Hop — The system IP address of the sending PE router.
- Extended communities — Contains the RT extended community and the VPLS-ID as community values.

Each VPLS instance is configured with import and export RT extended communities to create the required pseudowire topology by controlling the distribution of each NLRI.

This chapter describes the provisioning of a VPLS instance across three PE routers. A full mesh of pseudowires interconnects the VSI of each PE within the VPLS instance. A single attachment circuit is also configured on each VSI.

Configuration

The first step is to configure an MP-iBGP session using the L2-VPN address family between each of the PEs and the RR.

The configuration for the PEs is as follows:

```
# on PE-1, PE-2, and PE-3:
configure
     router Base
        autonomous-system 65536
        bgp
            group "internal"
                family l2-vpn
                peer-as 65536
                neighbor 192.0.2.6
                exit
            exit
            no shutdown
        exit
    exit
```

The IP addresses can be derived from Figure 175: Example topology.

The configuration for RR-6 is as follows:

```
# on RR-6:
configure
     router Base
     autonomous-system 65536
        bgp
            group "rr-internal"
                family l2-vpn
                cluster 1.1.1.1
                peer-as 65536
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
            no shutdown
        exit
    exit
```

On PE-1, the BGP session with RR-6 is established with address family L2-VPN capability negotiated, as follows:

```
*A:PE-1# show router bgp neighbor 192.0.2.6

===============================================================================
BGP Neighbor
===============================================================================
-------------------------------------------------------------------------------
Peer                 : 192.0.2.6
Description          : (Not Specified)
Group                : internal
-------------------------------------------------------------------------------
Peer AS              : 65536            Peer Port            : 50757
Peer Address         : 192.0.2.6
Local AS             : 65536            Local Port           : 179
Local Address        : 192.0.2.1
Peer Type            : Internal         Dynamic Peer         : No
State                : Established      Last State           : Established
Last Event           : recvOpen
Last Error           : Cease (Connection Collision Resolution)
Local Family         : L2-VPN
Remote Family        : L2-VPN
---snip---
```

On RR-6, the following BGP sessions are established with each PE for the L2-VPN address family:

```
*A:RR-6# show router bgp summary all

===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
Description
ServiceId          AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
192.0.2.1
Def. Instance  65536        6    0 00h01m35s 0/0/0 (L2VPN)
                            6    0
192.0.2.2
Def. Instance  65536        6    0 00h01m35s 0/0/0 (L2VPN)
                            6    0
192.0.2.3
Def. Instance  65536        6    0 00h01m35s 0/0/0 (L2VPN)
                            6    0

-------------------------------------------------------------------------------
```

A full mesh of RSVP Label Switched Paths (LSPs) is configured between the PE routers. For reference, the MPLS interface configuration and LSPs for PE-1 to PE-2 and PE-3 is as follows:

```
# on PE-1:
configure
    router Base
        mpls
            interface "int-PE-1-PE-2"
                no shutdown
            exit
            interface "int-PE-1-P-4"
```

```
                no shutdown
            exit
            path "loose"
                no shutdown
            exit
            lsp "LSP-PE-1-PE-2"
                to 192.0.2.2
                primary "loose"
                exit
                no shutdown
            exit
            lsp "LSP-PE-1-PE-3"
                to 192.0.2.3
                primary "loose"
                exit
                no shutdown
            exit
            no shutdown
```

VPLS PE configuration

Pseudowire templates

Pseudowire templates are used by BGP to dynamically instantiate service destination point (SDP) bindings. For a given service, pseudowire templates signal the egress service de-multiplexer labels used by remote PEs to reach the local PE.

The template determines the signaling parameters of the pseudowire, control word presence, plus other usage characteristics such as Split Horizon Groups (SHGs), MAC-pinning, filters, and so on.

The MPLS transport tunnel between PE routers can be signaled using either LDP or RSVP.

LDP-based pseudowires can be automatically instantiated; RSVP-based SDPs have to be pre-provisioned.

Pseudowire templates for auto-SDP creation using LDP

In order to use an LDP transport tunnel for data flow between PEs, it is necessary for link layer LDP to be configured between all PEs/Ps so that a transport label for each PE system interface address is available. Using this mechanism, SDPs can be auto-instantiated with SDP-IDs starting at 32767. Any subsequent SDPs created use SDP-IDs decrementing from this value.

A pseudowire template is required which may contain an SHG. Each SDP created with this template is contained within the configured SHG so that traffic cannot be forwarded between them.

```
# on PE-1, PE-2, PE-3:
configure
    service
        pw-template 1 name "PW1" create
            split-horizon-group "vpls-shg"
            exit
        exit
```

A pseudowire template can also be created that does not contain a split horizon group. The split horizon group can then be specified when the pw-template is included within the service.

```
# on PE-1, PE-2, PE-3:
configure
    service
        pw-template 2 name "PW2" create
        exit
```

Pseudowire templates for provisioned SDPs using RSVP

To use an RSVP tunnel as transport between PEs, it is necessary to bind the RSVP LSPs to the SDPs between each PE.

On PE-1, SDP 12 from PE-1 to PE-2 is configured as follows:

```
# on PE-1:
configure
    service
        sdp 12 mpls create
            description "RSVP-based SDP from PE-1 to PE-2"
            far-end 192.0.2.2
            lsp "LSP-PE-1-PE-2"
            no shutdown
        exit
```

To create an SDP within a service that uses the RSVP transport tunnel, a pseudowire template is required that has the **use-provisioned-sdp** parameter.

```
# on PE-1, PE-2, PE-3:
configure
    service
        pw-template 3 name "PW3" use-provisioned-sdp create
            exit
        exit
```

Alternatively, the **prefer-provisioned-sdp** parameter can be used, see chapter LDP VPLS Using BGP Auto-Discovery — Prefer Provisioned SDP.

VPLS BGP-AD using auto-provisioned SDPs

Figure 176: VPLS instance with auto-provisioned SDPs

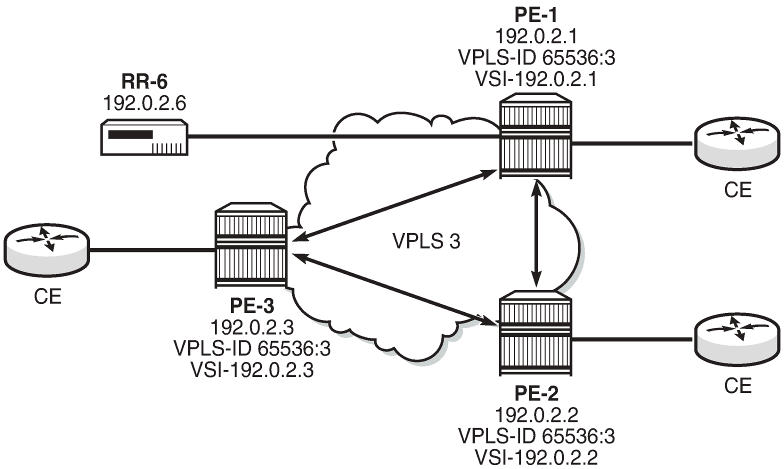


Figure 176: VPLS instance with auto-provisioned SDPs shows a schematic of a VPLS instance where the SDPs are auto-provisioned. SDPs are instantiated by a PE router using LDP signaling upon receipt of BGP Auto-Discovery (BGP-AD) updates from peer PE routers.

PE-1 configuration:

The following output shows the configuration required for a VPLS service using a pseudowire template configured for auto-provisioning of SDPs.

```
# on PE-1:
configure
    service
        vpls 3 name "VPLS3" customer 1 create
            bgp
                route-distinguisher 65536:3
                route-target export target:65536:3 import target:65536:3
                pw-template-binding 2 split-horizon-group "vpls-shg"
                    import-rt "target:65536:3"
                exit
            exit
            bgp-ad
                vpls-id 65536:3
                vsi-id
                    prefix 192.0.2.1
                exit
                no shutdown
            exit
            sap 1/1/4:3.0 create
            exit
            no shutdown
        exit
```

Within the **bgp** context, the pseudowire template is referenced which can be linked to an SHG and an import RT, if required.

Within the **bgp-ad** context, the signaling parameters are configured. These are two parameters used by each PE to determine the presence of a VPLS instance on a PE router. In turn, these are translated into endpoint identifiers for LDP signaling of pseudowires. As previously discussed, these parameters are:

- VPLS-ID — a unique identifier of the VPLS instance. Each PE that is a member of a VPLS must share the same VPLS-ID. This is inserted as an extended community value in the format AS:n. In this case, the VPLS-ID for VPLS 3 is 65536:3. This is a mandatory parameter and if it is not configured, it is not possible to enable BGP-AD using no shutdown.
- Virtual Switching Instance (VSI) prefix — This identifies a specific instance of the VPLS. This must be unique within the VPLS instance, and is encoded using the 4 byte dotted decimal notation. Generally, the system address is used as the VSI prefix. If this parameter is not configured, then the system address is used automatically.

The VPLS-ID and VSI prefix for VPLS 3 on each PE is shown in Figure 176: VPLS instance with auto-provisioned SDPs.

The VPLS-ID and VSI prefix are concatenated to form a unique VSI-ID. In this case, PE-1 has a VSI-ID of 65536:3:192.0.2.1. This uniquely identifies the VPLS instance on each individual PE and is advertised as an L2-VPN BGP update.

A BGP-AD update is transmitted to all other PEs via the RR, as follows:

```
*A:PE-1# show router bgp routes l2-vpn rd 65536:3 hunt
===============================================================================
 BGP Router ID:192.0.2.1        AS:65536       Local AS:65536
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP L2VPN Routes
===============================================================================
---snip---

-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
Route Type     : AutoDiscovery
Route Dist.    : 65536:3
Prefix         : 192.0.2.1
Nexthop        : 192.0.2.1
To             : 192.0.2.6
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : 0
AIGP Metric    : None                   IGP Cost       : n/a
Connector      : None
Community      : target:65536:3 l2-vpn/vrf-imp:65536:3
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.6
Origin         : IGP
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : n/a
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0

-------------------------------------------------------------------------------
```

```
Routes : 4
===============================================================================
```

The preceding BGP update is transmitted by PE-1 and has route type auto-discovery.

In this L2-VPN update, the VPLS-ID is encoded as the L2-VPN extended community 65536:3.

The VSI is seen as the prefix 192.0.2.1. The combination of the VPLS-ID and the VSI forms the VSI-ID and uniquely identifies the VPLS instance within this PE router.

The next-hop is also encoded as the local system IP address 192.0.2.1, which allows remote PEs to identify a suitable transport tunnel to PE-1 and for the targeted-LDP peer for instantiating the SDP.

As can be seen within the update, the VPLS-ID 65536:3 is also used to determine the RT extended community and the route distinguisher (RD).

PE-2 configuration

On PE-2, VPLS 3 is created using pseudowire template 1, with VPLS-ID 65536:3 and VSI-ID prefix 192.0.2.2 (system IP address), as follows”

```
# on PE-2:
configure
    service
        vpls 3 name "VPLS3" customer 1 create
            bgp
                route-distinguisher 65536:3
                route-target export target:65536:3 import target:65536:3
                pw-template-binding 2 split-horizon-group "vpls-shg"
                    import-rt "target:65536:3"
                exit
            exit
            bgp-ad
                vpls-id 65536:3
                vsi-id
                    prefix 192.0.2.2
                exit
                no shutdown
            exit
            sap 1/1/4:3.0 create
            exit
            no shutdown
        exit
```

PE-3 configuration

On PE-3, VPLS 3 is created using pseudowire template 2, with VPLS-ID 65536:3—identical to the VPLS-ID of PE-1 and PE-2—and VSI-ID 192.0.2.3 (system IP address), as follows:

```
# on PE-3:
configure
    service
        vpls 3 name "VPLS3" customer 1 create
            bgp
                route-distinguisher 65536:3
                route-target export target:65536:3 import target:65536:3
                pw-template-binding 2 split-horizon-group "vpls-shg"
                    import-rt "target:65536:3"
                exit
            exit
            bgp-ad
                vpls-id 65536:3
                vsi-id
```

```
                    prefix 192.0.2.3
                exit
                no shutdown
            exit
            sap 1/1/4:3.0 create
            exit
            no shutdown
        exit
```

PE-1 service operation verification

The following output shows that the service is operationally up on PE-1:

```
*A:PE-1# show service id 3 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 3                   Vpn Id            : 0
Service Type      : VPLS
---snip---

Admin State       : Up                  Oper State        : Up
MTU               : 1514
SAP Count         : 1                   SDP Bind Count    : 2
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:3.0                            qinq         1522    1522    Up   Up
sdp:32766:4294967294 SB(192.0.2.2)       BgpAd        0       1556    Up   Up
sdp:32767:4294967295 SB(192.0.2.3)       BgpAd        0       1556    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

As seen from the output, the service is operationally up, with the SAPs and SDPs also up. The *SB* flag indicates that the SDP is of type spoke-SDP (S flag) BGP (B flag).

BGP is used to discover the VPLS endpoints and exchange network reachability information. LDP is used to signal the pseudowires between the PEs.

LDP signaling occurs when each PE has discovered the endpoints of the VPLS instance. This compares with the use of the provisioned virtual circuit IDs used in an NMS provisioned VPLS instances as per RFC 4762.

The ability of PE-1 to reach the other PE routers with VSIs within the VPLS instance is verified from the following L2-route table:

```
*A:PE-1# show service l2-route-table bgp-ad

========================================================================
Services: L2 Route Information - Summary
========================================================================
Svc Id     L2-Routes (RD-Prefix)                 Next Hop        Origin
               Sdp Bind Id                           PW Temp Id
------------------------------------------------------------------------
3          *65536:3-192.0.2.2                    192.0.2.2       BGP-L2
               32766:4294967294                      2
3          *65536:3-192.0.2.3                    192.0.2.3       BGP-L2
```

```
               32767:4294967295                      2
------------------------------------------------------------------------
No. of L2 Route Entries: 2
========================================================================
```

This output shows the presence of the signaled pseudowire SDPs. SDPs from PE-1 to PE-2 and PE-3 are signaled using LDP Forwarding Equivalence Class (FEC) Element 129.

Each PE router uses targeted LDP to signal the local and remote endpoints. If there is an endpoint match, then SDPs are instantiated. This compares with the use of LDP for NMS provisioned SDPs, which uses virtual circuit IDs to signal pseudowires using LDP FEC Element 128.

In order to signal the SDPs, the following parameters are required:

1. Attachment Group Identifier (AGI): this is used to carry the VPLS-ID of the local PE router VPLS instance. The VPLS-ID must be identical for all PEs in the same VPLS instance.
2. Source Attachment Individual Identifier (SAII) and Target Attachment Individual Identifier (TAII): these use AII type 1 (RFC 4446) and are used to carry the NLRI (VSI-ID minus the RD) of the remote PE router VPLS instance.

The AGI for each PE must be identical. SAII and TAII must be different.

The following shows the service LDP bindings for VPLS 3 on PE-1:

```
*A:PE-1# show router ldp bindings services service-id 3

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        S - Status Signaled Up,  D - Status Signaled Down, e - Label ELC
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
Service Type:
        E - Epipe Service, V - VPLS Service, M - Mirror Service
        A - Apipe Service, F - Fpipe Service, I - IES Service, R - VPRN service
        P - Ipipe Service, C - Cpipe Service
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
===============================================================================
LDP Service FEC 128 Bindings
===============================================================================
Type                                          VCId       SDPId             LMTU
Peer                                          SvcId      IngLbl            RMTU
                                                         EgrLbl
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================

===============================================================================
LDP Service FEC 129 Bindings
===============================================================================
SAII                                            AGII        IngLbl         LMTU
TAII                                            Type        EgrLbl         RMTU
Peer                                            SvcId       SDPId
-------------------------------------------------------------------------------
192.0.2.1                                       1,8:020A00* 524279U        1500
192.0.2.2                                       V-Eth       524278S        1500
192.0.2.2:0                                     3           32766
```

```
192.0.2.1                                       1,8:020A00* 524280U        1500
192.0.2.3                                       V-Eth       524280S        1500
192.0.2.3:0                                     3           32767

-------------------------------------------------------------------------------
No. of FEC 129s: 2
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

This shows the two T-LDP bindings for PE-1 toward PE-2 and PE-3 for VPLS 3. The label bindings from this LDP output is identical to the SDP bindings output that follows. The following command can be used to list the SDP IDs and the SDP label bindings:

```
*A:PE-1# show service id 3 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
32766:4294967294 BgpAd    192.0.2.2       Up      Up        524279    524278
32767:4294967295 BgpAd    192.0.2.3       Up      Up        524280    524280
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
```

The SDP ID for the auto-provisioned SDP toward PE-2 is 32766, the SDP ID toward PE-3 is 32767.

The actual AGI, SAII, and TAII values are seen in the following detailed SDP output.

- AGI — 65536:3
- SAII — Local system IP address 192.0.2.1
- TAII — Remote system IP address 192.0.2.2 or 192.0.2.3

```

*A:PE-1# show service id 3 sdp 32767:4294967295 detail

===============================================================================
Service Destination Point (Sdp Id : 32767:4294967295) Details
===============================================================================
-------------------------------------------------------------------------------
 Sdp Id 32767:4294967295  -(192.0.2.3)
-------------------------------------------------------------------------------
Description     : (Not Specified)
SDP Id             : 32767:4294967295         Type             : BgpAd
PW-Template Id     : 2
AGI                : 65536:3                  SDP Bind Source  : bgp-l2vpn
Local AII          : 192.0.2.1
Remote AII         : 192.0.2.3
Split Horiz Grp    : vpls-shg
Etree Root Leaf Tag: Disabled                 Etree Leaf AC    : Disabled
VC Type            : Ether                    VC Tag           : n/a
Admin Path MTU     : 0                        Oper Path MTU    : 1556
Delivery           : MPLS
Far End            : 192.0.2.3                Tunnel Far End   : n/a
Oper Tunnel Far End: 192.0.2.3
LSP Types          : LDP/BGP
---snip---
```

PE-2 service operation verification

For completeness, the following shows that the VPLS service is operationally up on PE-2.

```
*A:PE-2# show service id 3 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 3                   Vpn Id            : 0
Service Type      : VPLS
---snip---

Admin State       : Up                  Oper State        : Up
MTU               : 1514
SAP Count         : 1                   SDP Bind Count    : 2
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:3.0                            qinq         1522    1522    Up   Up
sdp:32766:4294967294 SB(192.0.2.3)       BgpAd        0       1556    Up   Up
sdp:32767:4294967295 SB(192.0.2.1)       BgpAd        0       1556    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

```
*A:PE-2# show service l2-route-table bgp-ad

========================================================================
Services: L2 Route Information - Summary
========================================================================
Svc Id     L2-Routes (RD-Prefix)                Next Hop        Origin
               Sdp Bind Id                          PW Temp Id
------------------------------------------------------------------------
3          *65536:3-192.0.2.1                   192.0.2.1       BGP-L2
               32767:4294967295                     2
3          *65536:3-192.0.2.3                   192.0.2.3       BGP-L2
               32766:4294967294                     2
------------------------------------------------------------------------
No. of L2 Route Entries: 2
========================================================================
```

```
*A:PE-2# show router ldp bindings services service-id 3

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        S - Status Signaled Up,  D - Status Signaled Down, e - Label ELC
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
Service Type:
        E - Epipe Service, V - VPLS Service, M - Mirror Service
        A - Apipe Service, F - Fpipe Service, I - IES Service, R - VPRN service
        P - Ipipe Service, C - Cpipe Service
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
===============================================================================
LDP Service FEC 128 Bindings
```

```
===============================================================================
Type                                          VCId       SDPId             LMTU
Peer                                          SvcId      IngLbl            RMTU
                                                         EgrLbl
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================

===============================================================================
LDP Service FEC 129 Bindings
===============================================================================
SAII                                            AGII        IngLbl         LMTU
TAII                                            Type        EgrLbl         RMTU
Peer                                            SvcId       SDPId
-------------------------------------------------------------------------------
192.0.2.2                                       1,8:020A00* 524278U        1500
192.0.2.1                                       V-Eth       524279S        1500
192.0.2.1:0                                     3           32767

192.0.2.2                                       1,8:020A00* 524277U        1500
192.0.2.3                                       V-Eth       524279S        1500
192.0.2.3:0                                     3           32766

-------------------------------------------------------------------------------
No. of FEC 129s: 2
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

```
*A:PE-2# show service id 3 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
32766:4294967294 BgpAd    192.0.2.3       Up      Up        524277    524279
32767:4294967295 BgpAd    192.0.2.1       Up      Up        524278    524279
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
```

PE-3 service operation verification

On PE-3, the VPLS service is operationally up with the following BGP-AD SDPs:

```
*A:PE-3# show service id 3 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 3                   Vpn Id            : 0
Service Type      : VPLS
---snip---

Admin State       : Up                  Oper State        : Up
MTU               : 1514
SAP Count         : 1                   SDP Bind Count    : 2
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
```

```
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:3.0                            qinq         1522    1522    Up   Up
sdp:32766:4294967294 SB(192.0.2.2)       BgpAd        0       1556    Up   Up
sdp:32767:4294967295 SB(192.0.2.1)       BgpAd        0       1556    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

```
*A:PE-3# show service l2-route-table bgp-ad

========================================================================
Services: L2 Route Information - Summary
========================================================================
Svc Id     L2-Routes (RD-Prefix)                 Next Hop        Origin
               Sdp Bind Id                           PW Temp Id
------------------------------------------------------------------------
3          *65536:3-192.0.2.1                    192.0.2.1       BGP-L2
               32767:4294967295                      2
3          *65536:3-192.0.2.2                    192.0.2.2       BGP-L2
               32766:4294967294                      2
------------------------------------------------------------------------
No. of L2 Route Entries: 2
========================================================================
```

```
*A:PE-3# show router ldp bindings services service-id 3

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        S - Status Signaled Up,  D - Status Signaled Down, e - Label ELC
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
Service Type:
        E - Epipe Service, V - VPLS Service, M - Mirror Service
        A - Apipe Service, F - Fpipe Service, I - IES Service, R - VPRN service
        P - Ipipe Service, C - Cpipe Service
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
===============================================================================
LDP Service FEC 128 Bindings
===============================================================================
Type                                          VCId       SDPId             LMTU
Peer                                          SvcId      IngLbl            RMTU
                                                         EgrLbl
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================

===============================================================================
LDP Service FEC 129 Bindings
===============================================================================
SAII                                            AGII        IngLbl         LMTU
TAII                                            Type        EgrLbl         RMTU
Peer                                            SvcId       SDPId
-------------------------------------------------------------------------------
192.0.2.3                                       1,8:020A00* 524280U        1500
192.0.2.1                                       V-Eth       524280S        1500
192.0.2.1:0                                     3           32767
```

```
192.0.2.3                                       1,8:020A00* 524279U        1500
192.0.2.2                                       V-Eth       524277S        1500
192.0.2.2:0                                     3           32766

-------------------------------------------------------------------------------
No. of FEC 129s: 2
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

```
*A:PE-3# show service id 3 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
32766:4294967294 BgpAd    192.0.2.2       Up      Up        524279    524277
32767:4294967295 BgpAd    192.0.2.1       Up      Up        524280    524280
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
```

BGP AD using pre-provisioned SDPs

It is possible to configure BGP-AD instances that use RSVP transport tunnels. In this case, the LSPs and SDPs must be manually created.

Figure 177: VPLS instance using pre-provisioned SDPs

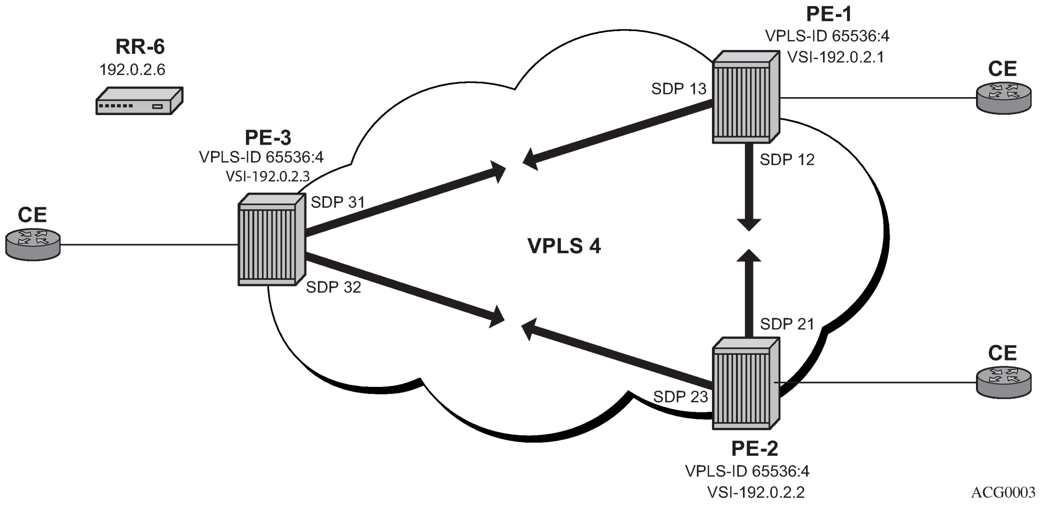


Figure 177: VPLS instance using pre-provisioned SDPs shows a VPLS instance configured across three PE routers as before.

The SDP configurations for the three PEs are as follows:

```
# on PE-1:
configure
```

```
    service
        sdp 12 mpls create
            description "RSVP-based SDP from PE-1 to PE-2"
            far-end 192.0.2.2
            lsp "LSP-PE-1-PE-2"
            no shutdown
        exit
        sdp 13 mpls create
            description "RSVP-based SDP from PE-1 to PE-3"
            far-end 192.0.2.3
            lsp "LSP-PE-1-PE-3"
            no shutdown
        exit
```

```
# on PE-2:
configure
    service
        sdp 21 mpls create
            description "RSVP-based SDP from PE-2 to PE-1"
            far-end 192.0.2.1
            lsp "LSP-PE-2-PE-1"
            no shutdown
        exit
        sdp 23 mpls create
            description "RSVP-based SDP from PE-2 to PE-3"
            far-end 192.0.2.3
            lsp "LSP-PE-2-PE-3"
            no shutdown
        exit
```

```
# on PE-3:
configure
    service
        sdp 31 mpls create
            description "RSVP-based SDP from PE-3 to PE-1"
            far-end 192.0.2.1
            lsp "LSP-PE-3-PE-1"
            no shutdown
        exit
        sdp 32 mpls create
            description "RSVP-based SDP from PE-3 to PE-2"
            far-end 192.0.2.2
            lsp "LSP-PE-3-PE-2"
            no shutdown
        exit
```

The pw-template that is to be used within each VPLS instance must be provisioned on all PEs and must use the keyword **use-provisioned-sdp**. The pw-template is configured on all PEs with the following command:

```
# on PE-1, PE-2, PE-3:
configure
    service
        pw-template 3 name "PW3" use-provisioned-sdp create
        exit
```

The following output shows the configuration required for a VPLS service using a pseudowire template configured for pre-provisioned RSVP SDPs.

```
# on PE-1:
```

```
configure
    service
        vpls 4 name "VPLS4" customer 1 create
            bgp
                route-distinguisher 65536:4
                route-target export target:65536:4 import target:65536:4
                pw-template-binding 3 split-horizon-group "vpls-shg"
                    import-rt "target:65536:4"
                exit
            exit
            bgp-ad
                vpls-id 65536:4
                vsi-id
                    prefix 192.0.2.1
                exit
                no shutdown
            exit
            sap 1/1/4:4.0 create
            exit
            no shutdown
        exit
```

Similarly, on PE-2 the configuration is as follows:

```
# on PE-2:
configure
    service
        vpls 4 name "VPLS4" customer 1 create
            bgp
                route-distinguisher 65536:4
                route-target export target:65536:4 import target:65536:4
                pw-template-binding 3 split-horizon-group "vpls-shg"
                    import-rt "target:65536:4"
                exit
            exit
            bgp-ad
                vpls-id 65536:4
                vsi-id
                    prefix 192.0.2.2
                exit
                no shutdown
            exit
            sap 1/1/4:4.0 create
            exit
            no shutdown
        exit
```

On PE-3, VPLS 4 is configured as follows:

```
# on PE-3:
configure
    service
        vpls 4 name "VPLS4" customer 1 create
            bgp
                route-distinguisher 65536:4
                route-target export target:65536:4 import target:65536:4
                pw-template-binding 3 split-horizon-group "vpls-shg"
                    import-rt "target:65536:4"
                exit
            exit
            bgp-ad
                vpls-id 65536:4
                vsi-id
```

```
                    prefix 192.0.2.3
                exit
                no shutdown
            exit
            sap 1/1/4:4.0 create
            exit
            no shutdown
        exit
```

The following output shows that the service is operationally up on PE-1.

```
*A:PE-1# show service id 4 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 4                   Vpn Id            : 0
Service Type      : VPLS
---snip---

Admin State       : Up                  Oper State        : Up
MTU               : 1514
SAP Count         : 1                   SDP Bind Count    : 2
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:4.0                            qinq         1522    1522    Up   Up
sdp:12:4294967293 S(192.0.2.2)           BgpAd        0       1556    Up   Up
sdp:13:4294967292 S(192.0.2.3)           BgpAd        0       1556    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

The SDP identifiers are the pre-provisioned SDPs: SDP 12 and 13.

The following command shows that the service is operationally up on PE-2.

```
*A:PE-2# show service id 4 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 4                   Vpn Id            : 0
Service Type      : VPLS
---snip---

Admin State       : Up                  Oper State        : Up
MTU               : 1514
SAP Count         : 1                   SDP Bind Count    : 2
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:4.0                            qinq         1522    1522    Up   Up
sdp:21:4294967293 S(192.0.2.1)           BgpAd        0       1556    Up   Up
sdp:23:4294967292 S(192.0.2.3)           BgpAd        0       1556    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

The following command shows that the service is operationally up on PE-3.

```
*A:PE-3# show service id 4 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 4                   Vpn Id            : 0
Service Type      : VPLS
---snip---

Admin State       : Up                  Oper State        : Up
MTU               : 1514
SAP Count         : 1                   SDP Bind Count    : 2
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:4.0                            qinq         1522    1522    Up   Up
sdp:31:4294967293 S(192.0.2.1)           BgpAd        0       1556    Up   Up
sdp:32:4294967292 S(192.0.2.2)           BgpAd        0       1556    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

Conclusion

BGP-AD coupled with LDP pseudowire signaling allows the delivery of L2-VPN services to customers where BGP is commonly used. This example shows the configuration of BGP-AD together with the associated show outputs which can be used for verification and troubleshooting.

LDP VPLS Using BGP Auto-Discovery — Prefer Provisioned SDP

This chapter provides information about LDP VPLS using BGP auto-discovery — prefer provisioned SDP.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS Release 14.0.R6, but the CLI in the current edition is based on SR OS Release 21.2.R1. BGP Auto-Discovery (BGP-AD) based on RFC 6074 is supported in SR OS Release 6.0, and later. The **prefer-provisioned-sdp** option is supported in SR OS Release 14.0.R1, and later.

Overview

As described in chapter LDP VPLS Using BGP Auto-Discovery, BGP-AD based on RFC 6074 can auto-create SDP bindings, but an operator can force the system to use a provisioned SDP by specifying the **use-provisioned-sdp** option. This chapter compares the **use-provisioned-sdp** option with the **prefer-provisioned-sdp** option. The chapter describes a migration scenario for a VPLS service with a pseudowire (PW) template binding, restricted to using provisioned SDPs toward a PW template binding preferring to use provisioned SDPs, but auto-creating SDPs in case there is no suitable manually created SDP available.

PW templates

PW templates can be configured with the following command:

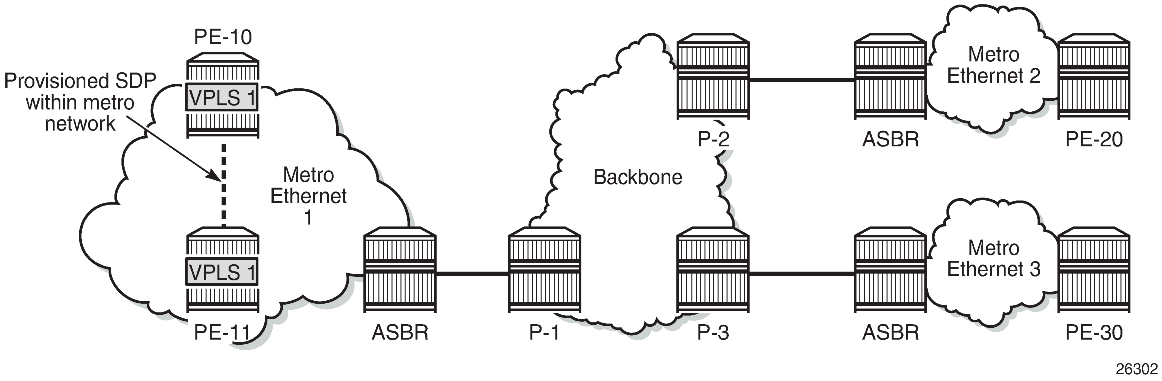
```
*A:PE-1>config>service# pw-template ?
  - pw-template <policy-id> [create] [prefer-provisioned-sdp] [name <name>]
                                                                      [auto-gre-sdp]
  - no pw-template <policy-id>
  - pw-template <policy-id> use-provisioned-sdp [create] [name <name>]

 <policy-id>           : [1..2147483647]
 <use-provisioned-s*> : keyword
 <create>              : keyword - mandatory while creating an entry.
 <prefer-provisione*> : keyword
 <name>                : [64 chars max]
 <auto-gre-sdp>        : keyword
```

```
---snip---
```

- When the **use-provisioned-sdp** keyword is added at creation time, the tunnel manager is forced to look for a provisioned and active SDP to the far-end PE. The far-end PE is auto-discovered from the BGP next hop. If multiple SDPs are active to this far-end PE, the tunnel manager chooses the SDP template with the best metric. If there is a tie, the SDP ID is used as a tie-breaker and the highest SDP ID wins. However, if no provisioned SDP exists, the SDP binding will not be instantiated.
- When the **prefer-provisioned-sdp** keyword is added at creation time, the behavior is the same as when a provisioned SDP exists. When the tunnel manager finds an existing matching SDP, it will use it even if it is operationally down. Only when no provisioned SDP exists, will the SDP binding be auto-created.
- When a PW template is created without the **use-provisioned-sdp** or **prefer-provisioned-sdp** keyword, the SDP bindings will be auto-created.

Figure 178: LDP VPLS using BGP-AD with use-provisioned-sdp option shows the following use case: the metro Ethernet networks were initially built with provisioned SDPs. Intra-metro services are provisioned using provisioned SDPs; for example, customer X has a VPLS service defined in the metro Ethernet networks, using BGP-AD with a PW template to use the provisioned SDPs in the metro Ethernet networks.

Figure 178: LDP VPLS using BGP-AD with use-provisioned-sdp option

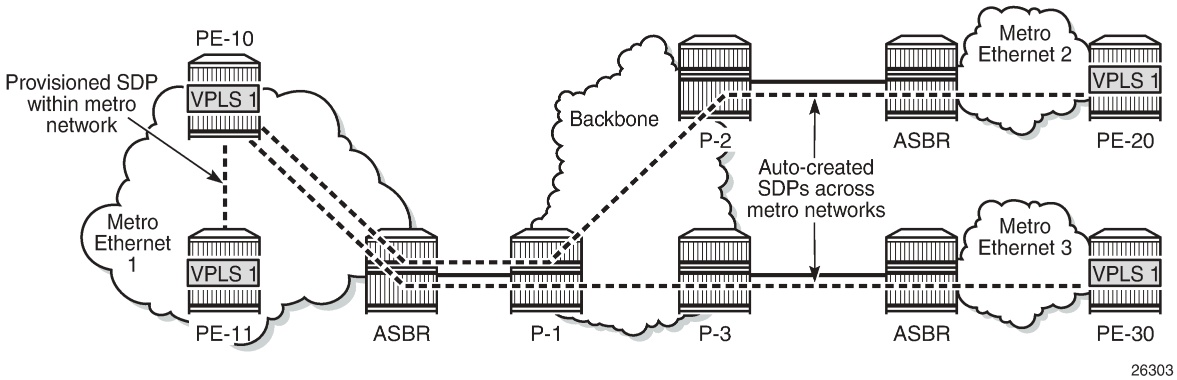


The service provider initially started with PE-10 and PE-11 in metro Ethernet 1, but now wants to add PE-20 and PE-30 as new sites to the VPLS service. Therefore, the BGP-AD routes should propagate beyond the boundaries of the metro Ethernet network. The backbone network may be in a different AS, but in this example, all networks are in the same AS. VPLS 1 of customer X can have sites added to the service on PEs in different metro Ethernet networks. A new PW template is created with the **prefer-provisioned-sdp** option and applied to the VPLS service.

- When a new site within the metro Ethernet network is added, an SDP is already provisioned to this site and this SDP is used for the SDP binding in the VPLS.
- When a new site in a different metro Ethernet network is added, no SDP is available to the site in the remote metro Ethernet network and the SDP binding is auto-created.

Figure 179: LDP VPLS using BGP-AD with prefer-provisioned-sdp option shows the SDP bindings in VPLS 1 between PE-10 and the other PEs. For simplicity, the SDP bindings between the other PEs are not shown.

Figure 179: LDP VPLS using BGP-AD with prefer-provisioned-sdp option

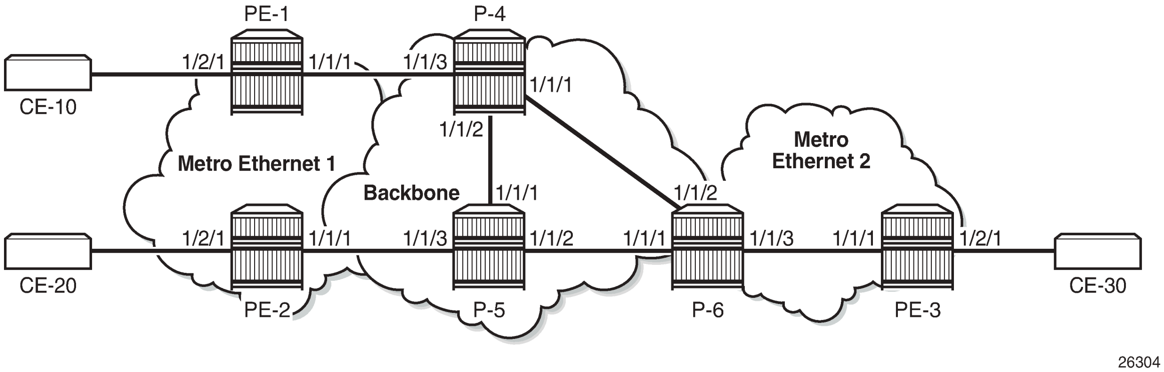


The **prefer-provisioned-sdp** and **use-provisioned-sdp** options can only be defined at creation time, implying that existing PW templates cannot be changed from prefer-provisioned-sdp to use-provisioned-sdp and vice versa. To support migration from one PW template to another with minimal service impact, two PW templates can be applied in parallel, as shown in the Configuration section.

Configuration

Figure 180: Example topology shows the example topology. For simplicity, all nodes are in the same AS.

Figure 180: Example topology



The initial configuration includes the following:

- Cards, MDAs, ports
- Router interfaces
- IS-IS as IGP (or OSPF) on all interfaces
- MPLS and RSVP on all interfaces, except "int-P-4-P-6" and "int-P-5-P-6".
- LDP on all interfaces

BGP is configured on all PE routers for address family l2-vpn, as follows:

```
# on PE-1, PE-2, PE-3:
configure
    router Base
        autonomous-system 64496
        bgp
            group "internal"
                family l2-vpn
                peer-as 64496
                neighbor 192.0.2.6
                exit
            exit
```

The BGP configuration on the route reflector (RR) P-6 is as follows:

```
# on P-6:
configure
    router Base
        autonomous-system 64496
        bgp
            group "rr-internal"
                family l2-vpn
                cluster 1.1.1.1
                peer-as 64496
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
```

On PE-1 and PE-2 in metro Ethernet network 1, an RSVP LSP is created that is used in a manually created SDP. The LSP configuration on PE-1 is as follows:

```
# on PE-1:
configure
    router Base
        mpls
            path "loose"
                no shutdown
            exit
            lsp "LSP-PE-1-PE-2"
                to 192.0.2.2
                primary "loose"
                exit
                no shutdown
            exit
            no shutdown
```

On PE-1, SDP 12 is configured as follows:

```
# on PE-1:
configure
    service
        sdp 12 mpls create
            description "SDP12 to 192.0.2.2"
            far-end 192.0.2.2
            lsp "LSP-PE-1-PE-2"
            keep-alive
```

```
                shutdown
            exit
            no shutdown
        exit
```

The configuration on PE-2 is similar.

LDP VPLS using AD without prefer-provisioned-sdp option

Initially, the following two PW templates are created on all PEs: PW template 1 has the **use-provisioned-sdp** option and PW template 2 is created without any option; therefore, SDP bindings will be auto-created.

```
# on PE-1, PE-2, PE-3:
configure
    service
        pw-template 1 name "PW 1" use-provisioned-sdp create
        exit
        pw-template 2 name "PW 2" create
        exit
```

The following lists the PW templates configured on PE-1:

```
*A:PE-1# show service pw-template

===============================================================================
PW Template information
===============================================================================
PW Template Id      SDP                           Last Update
-------------------------------------------------------------------------------
1                   Use-provisioned               04/01/2021 08:47:29
2                   Auto-mpls                     04/01/2021 08:07:06
===============================================================================
```

On all PEs, two VPLS services are created: VPLS 1 with BGP-AD PW template 1 and VPLS 2 with PW template 2, as follows:

```
# on PE-1, PE-2, PE-3:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            bgp
                route-distinguisher 64496:1
                route-target export target:64496:1 import target:64496:1
                pw-template-binding 1
                exit
            exit
            bgp-ad
                vpls-id 64496:1
                no shutdown
            exit
            stp
                shutdown
            exit
            sap 1/2/1:1 create
                no shutdown
            exit
            no shutdown
        exit
        vpls 2 name "VPLS 2" customer 1 create
```

```
            bgp
                route-distinguisher 64496:2
                route-target export target:64496:2 import target:64496:2
                pw-template-binding 2 import-rt "target:64496:2"
                exit
            exit
            bgp-ad
                vpls-id 64496:2
                no shutdown
            exit
            stp
                shutdown
            exit
            sap 1/2/1:2 create
                no shutdown
            exit
            no shutdown
        exit
```

On PE-1, the following SDP bindings have been created:

```
*A:PE-1# show service sdp-using

===============================================================================
SDP Using
===============================================================================
SvcId      SdpId              Type   Far End              Opr   I.Label E.Label
                                                          State
-------------------------------------------------------------------------------
1          12:4294967295      BgpAd  192.0.2.2            Up    524280  524280
2          32766:4294967293   BgpAd  192.0.2.3            Up    524278  524280
2          32767:4294967294   BgpAd  192.0.2.2            Up    524279  524279
-------------------------------------------------------------------------------
Number of SDPs : 3
-------------------------------------------------------------------------------
===============================================================================
```

The first SDP binding is created by BGP-AD in VPLS 1 and uses the configured SDP 12 with far-end PE-2; the other two SDP bindings have been auto-created by BGP-AD in VPLS 2 and have far-end PE-2 and PE-3.

The list of SDP bindings on PE-2 looks similar:

```
*A:PE-2# show service sdp-using

===============================================================================
SDP Using
===============================================================================
SvcId      SdpId              Type   Far End              Opr   I.Label E.Label
                                                          State
-------------------------------------------------------------------------------
1          21:4294967295      BgpAd  192.0.2.1            Up    524280  524280
2          32766:4294967293   BgpAd  192.0.2.3            Up    524278  524281
2          32767:4294967294   BgpAd  192.0.2.1            Up    524279  524279
-------------------------------------------------------------------------------
Number of SDPs : 3
-------------------------------------------------------------------------------
===============================================================================
```

On PE-3, there are only two SDP bindings, both in VPLS 2:

```
*A:PE-3# show service sdp-using
```

```
===============================================================================
SDP Using
===============================================================================
SvcId      SdpId              Type   Far End              Opr   I.Label E.Label
                                                          State
-------------------------------------------------------------------------------
2          32766:4294967294   BgpAd  192.0.2.1            Up    524280  524278
2          32767:4294967295   BgpAd  192.0.2.2            Up    524281  524278
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
```

Log "99" on PE-3 shows that the system failed to create a dynamic BGP-L2VPN SDP binding because no provisioned SDP was found, as follows:

```
77 2021/04/01 08:49:00.672 UTC MAJOR: SVCMGR #2322 Base
"The system failed to create a dynamic bgp-l2vpn SDP Bind  in service 1 with SDP
pw-template policy 1 for the following reason: suitable manual SDP not found."
```

Figure 181: SDP bindings in VPLS 1 with use-provisioned-sdp option shows the SDPs used in VPLS 1. PE-1 and PE-2 both used the provisioned SDP. PE-3 has no SDP bindings in VPLS 1.

Figure 181: SDP bindings in VPLS 1 with use-provisioned-sdp option

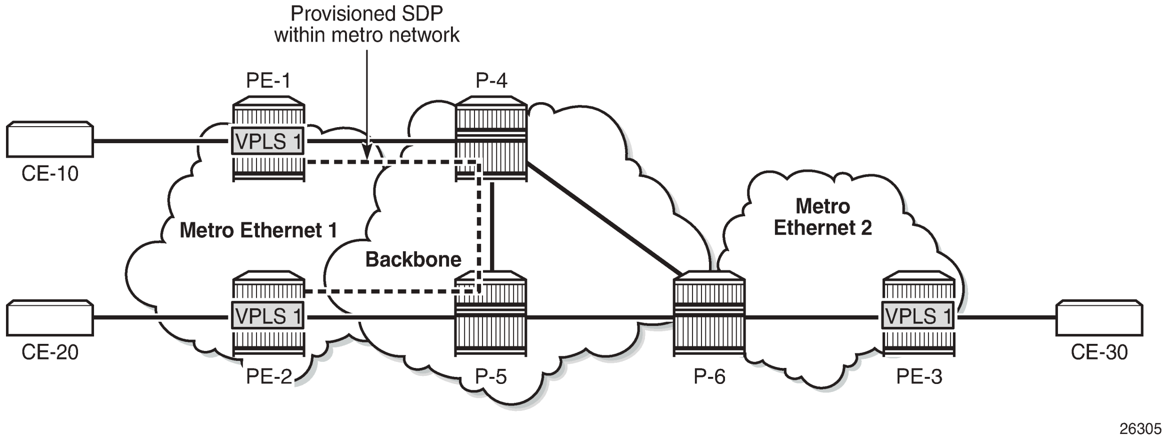
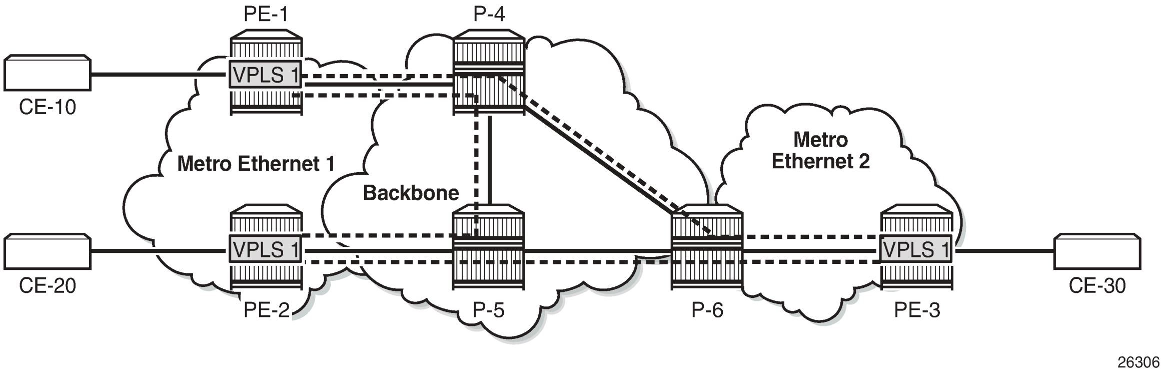


Figure 182: Auto-created SDP bindings in VPLS 2 shows the auto-created SDP bindings in VPLS 2. Each PE has two auto-created SDP bindings to each other PE.

Figure 182: Auto-created SDP bindings in VPLS 2



Migrate VPLS 1 to prefer-provisioned-sdp option

VPLS 1 uses PW template 1 with the **use-provisioned-sdp** option. This option is defined at creation time and cannot be modified afterward, as follows:

```
*A:PE-1>config>service# pw-template 1 prefer-provisioned-sdp
MINOR: CLI The prefer-provisioned-sdp option cannot be modified after creation.
```

The following steps are needed to migrate to another PW template with the **prefer-provisioned-sdp** option without service outage:

1. Create new PW template with **prefer-provisioned-sdp** option.
2. Add new PW template binding to VPLS and verify which PW template is used.
3. Modify old PW template binding to make it not usable.
4. Launch tools command to re-evaluate old PW template in the VPLS.
5. When the old PW template is not used anymore, remove PW template binding from the VPLS configuration.

A new PW template with the prefer-provisioned-sdp option is configured on all PEs, as follows:

```
# on PE-1, PE-2, PE-3:
configure
    service
        pw-template 10 name "PW 10" prefer-provisioned-sdp create
        exit
```

An additional PW template binding is configured in VPLS 1 on all PEs, as follows:

```
# on PE-1, PE-2, PE-3:
configure
    service
        vpls "VPLS 1"
            bgp
                pw-template-binding 10
                exit
```

The configuration of VPLS 1 includes two PW template bindings, as follows:

```
*A:PE-1>config>service>vpls# info
----------------------------------------------
            bgp
                route-distinguisher 64496:1
                route-target export target:64496:1 import target:64496:1
                pw-template-binding 1
                exit
                pw-template-binding 10
                exit
            exit
            bgp-ad
                vpls-id 64496:1
                no shutdown
            exit
            stp
                shutdown
            exit
            sap 1/2/1:1 create
                no shutdown
            exit
            no shutdown
```

The following shows that no additional SDP bindings have been created. The only SDP binding in VPLS 1 on PE-1 uses the provisioned SDP 12.

```
*A:PE-1# show service id 1 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
12:4294967295    BgpAd    192.0.2.2       Up      Up        524280    524280
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

The following shows that PW template 1 was used for the creation of the SDP binding:

```
*A:PE-1# show service id 1 sdp detail | match "SDP Id|PW-Template Id" expression
SDP Id             : 12:4294967295            Type              : BgpAd
PW-Template Id     : 1
```

The PW template 10 has a higher ID than PW template 1 and is not used. Re-evaluating the PW template binding for PW template 1 in VPLS 1 will make no difference if both PW templates are usable. However, PW template 1 can be made unusable by adding a dummy **import-rt** not matching any route in the VPLS, as follows:

```
# on PE-1, PE-2, PE-3:
configure
    service
        vpls "VPLS 1"
            bgp
                pw-template-binding 1 import-rt "target:111:111"
                exit
            exit
```

As a result, PW template 10 with the **prefer-provisioned-sdp** option is used for the automatic creation of SDP bindings where no provisioned SDP is available, as follows:

```
*A:PE-1# show service id 1 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
12:4294967295    BgpAd    192.0.2.2       Up      Up        524280    524280
32766:4294967292 BgpAd    192.0.2.3       Up      Up        524277    524278
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
```

For the first SDP binding, PW template 1 is used, and for the second SDP binding, PW template 10 is used, as follows:

```
*A:PE-1# show service id 1 sdp detail | match "SDP Id|PW-Template Id" expression
SDP Id             : 12:4294967295            Type              : BgpAd
PW-Template Id     : 1
SDP Id             : 32766:4294967292         Type              : BgpAd
PW-Template Id     : 10
```

The following command forces the system to re-evaluate PW template 1 in VPLS 1:

```
*A:PE-1# tools perform service id 1 eval-pw-template 1 allow-service-impact
eval-pw-template succeeded for Svc 1 12:4294967295 Policy 1
```

As a result, only PW template 10 is used for the creation of SDP bindings in VPLS 1, as follows:

```
*A:PE-1# show service id 1 sdp detail | match "SDP Id|PW-Template Id" expression
SDP Id             : 12:4294967291            Type              : BgpAd
PW-Template Id     : 10
SDP Id             : 32766:4294967292         Type              : BgpAd
PW-Template Id     : 10
```

PW template 1 is not used anymore and can be removed from the VPLS configuration, as follows:

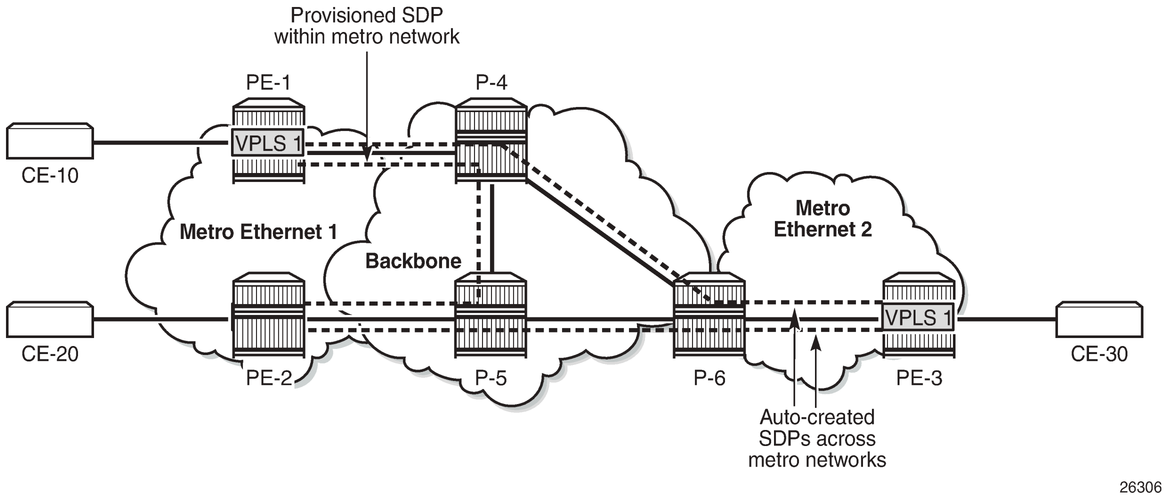
```
# on PE-1, PE-2, PE-3:
configure
    service
        vpls "VPLS 1"
            bgp
                no pw-template-binding 1
            exit
```

The configuration of VPLS 1 on PE-1 contains only a PW template binding for PW template 10, as follows:

```
*A:PE-1>config>service>vpls# info
----------------------------------------------
            bgp
                route-distinguisher 64496:1
                route-target export target:64496:1 import target:64496:1
                pw-template-binding 10
                exit
            exit
            bgp-ad
                vpls-id 64496:1
                no shutdown
            exit
            stp
                shutdown
            exit
            sap 1/2/1:1 create
                no shutdown
            exit
            no shutdown
----------------------------------------------
```

Figure 183: SDP bindings in VPLS 1 with prefer-provisioned-sdp option shows the SDP bindings in VPLS 1 with the **prefer-provisioned-sdp** option. Within metro Ethernet network 1, the provisioned SDP is used, and between metro Ethernet networks, auto-created SDP bindings are used.

Figure 183: SDP bindings in VPLS 1 with prefer-provisioned-sdp option



Conclusion

LDP VPLS using BGP-AD allows the creation of SDP bindings that are either auto-created or that use provisioned SDPs. When the **prefer-provisioned-sdp** option is used, the tunnel manager will look for a provisioned and active SDP to the far end and use it, if available, even if it is down. When no provisioned SDP is available, the system will auto-create an SDP binding.

Mobility for EVPN Hosts Within an R-VPLS

This chapter provides information about Mobility for EVPN Hosts Within an R-VPLS.

Topics in this chapter include:

Applicability

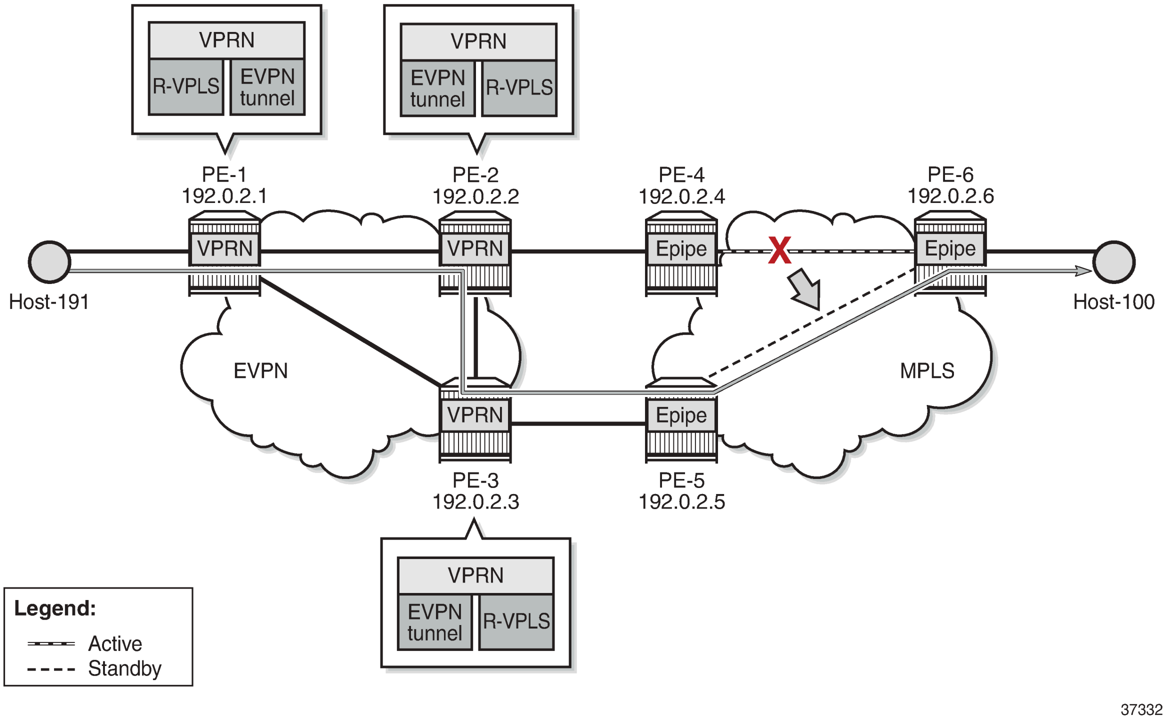
The information and configuration in this chapter are based on SR OS Release 21.10.R2. Efficient EVPN host mobility without tromboning or hairpinning in an R-VPLS is supported for IPv4 in SR OS Release 19.10.R3 and later and is supported for IPv6 in SR OS Release 20.5.R1 and later.

Overview

SR OS can populate a VPRN route table with host routes learned from the IPv4 Address Resolution Protocol (ARP) messages or IPv6 Neighbor Discovery (ND) protocol messages. The host routes can be advertised in the VPRN context as IP-VPN or EVPN route type 5 (RT5), to be used by an IP-VPN or EVPN core network for inter-subnet forwarding. SR OS supports *draft-ietf-bess-evpn-inter-subnet-forwarding* for a dynamic and efficient routing between remote hosts, avoiding hairpinning.

In SR OS Releases earlier than Release 19.10.R3, inefficient hairpinning situations may occur when the VPRN is configured to advertise IPv4 host routes as IP-VPN or EVPN RT5 routes. Figure 184: Hairpinning in a broadcast domain after switchover for SR OS Releases earlier than Release 19.10.R3 shows hairpinning in an EVPN broadcast domain with PE-1, PE-2, and PE-3.

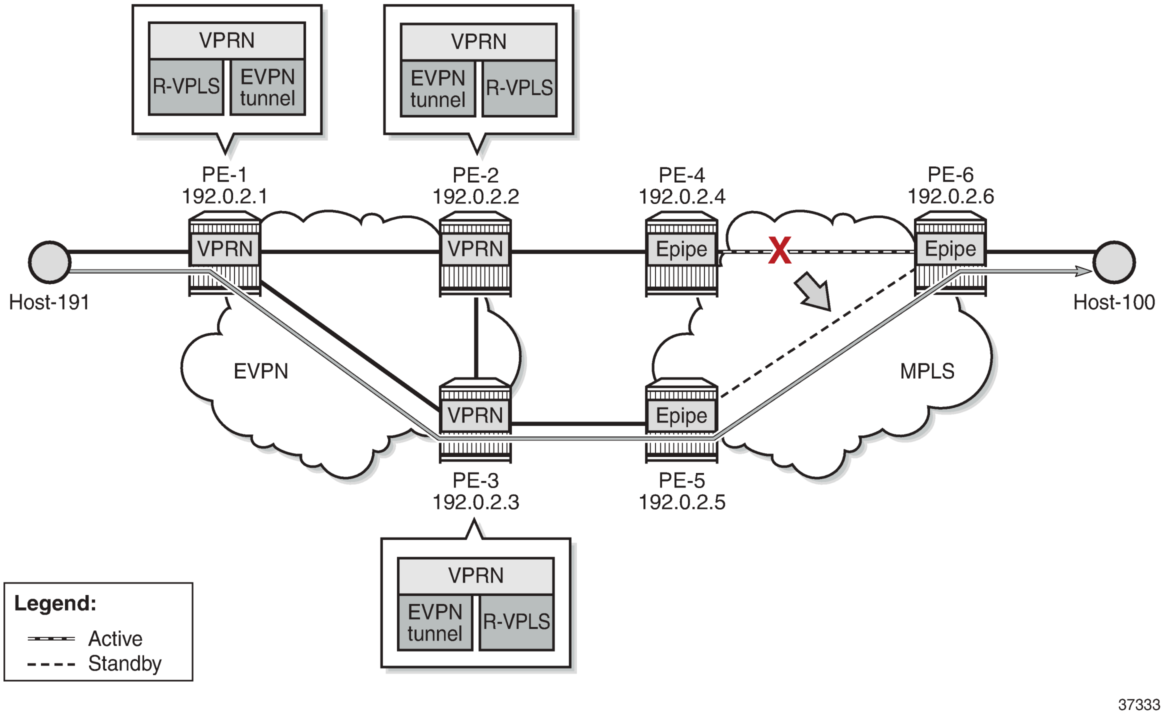
Figure 184: Hairpinning in a broadcast domain after switchover for SR OS Releases earlier than Release 19.10.R3



When host-100 comes up, it sends a Gratuitous Address Resolution Protocol (GARP) message that is then learned on PE-2 and PE-3. PE-2 and PE-3 are configured to advertise host routes, so they generate an RT5 host route for prefix 10.0.0.100/32 of host-100. PE-3 selects its best RT5 for 10.0.0.100/32 and traffic from host-191 to host-100 uses the path via PE-1, PE-2, PE-4, and PE-6. However, when the active path between PE-4 and PE-6 fails, the standby path between PE-5 and PE-6 takes over and hairpinning occurs when PE-1 continues selecting PE-2 as the next hop, while a more efficient path is possible via next-hop PE-3. Traffic from host-191 to host-100 uses the path via PE-1, PE-2, PE-3, PE-5, and PE-6.

In SR OS Release 19.10.R3 and later, the more efficient path from host-191 via PE-1, PE-3, PE-5, and PE-6 to host-100 is used, as shown in Figure 185: Forwarding in a broadcast domain after switchover for SR OS Release 19.10.R3 and later.

Figure 185: Forwarding in a broadcast domain after switchover for SR OS Release 19.10.R3 and later



In SR OS Release 19.10.R3 and later, EVPN host mobility is supported for IPv4 as described in section "Symmetric and Asymmetric IRB" of *draft-ietf-bess-evpn-inter-subnet-forwarding*. When a host moves from a source PE to a target PE in the same broadcast domain, the behavior for IPv4 hosts is one of the following.

1. The host initiates an ARP request or GARP.
2. The host sends a data packet without first initiating an ARP request or GARP.
3. The host does not send any traffic and the source PE generates an ARP request when the MAC address of the host expires and the EVPN-MAC is withdrawn.

All three scenarios are described in more detail later, where the move of host-100 from source PE-2 to target PE-3 is simulated.

For the first of these scenarios, the VPRN configuration on PE-2 is as follows:

```
# on PE-2:
configure
    service
        vprn 16 name "ip-vrf-16" customer 1 create
            interface "evi-15" create
                mac 00:00:00:00:00:02
                vpls "sbd-15"
                    evpn-tunnel
                exit
            exit
            interface "evi-17" create
                address 10.0.0.2/24
                mac 00:00:00:00:2e:17
                arp-host-route
```

```
                    populate dynamic
                exit
                arp-timeout 300
                arp-learn-unsolicited
                arp-proactive-refresh
                vrrp 1 passive
                    backup 10.0.0.254
                    ping-reply
                    traceroute-reply
                exit
                local-proxy-arp
                vpls "evi-17"
                    evpn
                        arp
                            no learn-dynamic
                            no flood-garp-and-unknown-req
                            advertise dynamic
                        exit
                    exit
                exit
            exit
            no shutdown
        exit
```

The behavior is controlled by the following commands.

- **arp-host-route>populate [dynamic | evpn | static]** configures PE-2 to advertise host routes. The type of ARP entry that can create a host route can be dynamic, EVPN, static, or a combination of these.
- **arp-learn-unsolicited** triggers the learning of an ARP entry upon receiving an ARP or GARP message that was not requested by the router.
- **arp-proactive-refresh** triggers the refresh of the ARP entry 30 seconds before aging out.
- **local-proxy-arp** ensures that PE-2 replies to any received ARP request on behalf of the other hosts in the R-VPLS broadcast domain.
- **vpls>evpn>arp>[no] learn-dynamic** controls whether data path ARP messages received on EVPN connections can populate the ARP tables.
- **vpls>evpn>arp>[no] flood-garp-and-unknown-req** controls the flooding of Control Processing Module (CPM)-generated ARP requests to EVPN destinations.
- **vpls>evpn>arp>advertise [dynamic | static]** enables PE-2 to advertise MAC and IP in EVPN-MAC routes for ARP entries of the dynamic or static type.

For IPv6, the corresponding commands are as follows:

- **ipv6>nd-host-route>populate [dynamic | evpn | static]**
- **ipv6>nd-learn-unsolicited [global | link-local | both]**
- **ipv6>nd-proactive-refresh [global | link-local | both]** triggers the refresh of the ND entry upon aging out.
- **ipv6>local-proxy-nd**
- **vpls>evpn>nd>[no] learn-dynamic**
- **vpls>evpn>nd>advertise [dynamic | static]**

For IPv6, CPM-generated Neighbor Solicitation (NS) messages are always flooded to EVPN destinations. This is not configurable in the **vpls>evpn>nd** context of the VPRN service, in contrast to the **[no] flood-garp-and-unknown-req** command in the **vpls>evpn>arp** context for IPv4.

The behavior for IPv6 hosts when moving from a source PE to a target PE is one of the following.

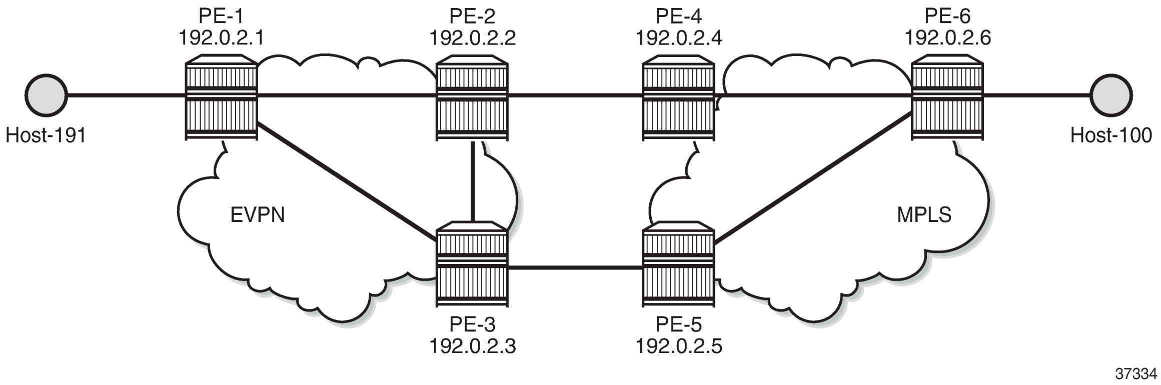
1. The host initiates an unsolicited Neighbor Advertisement (NA).
2. The host sends traffic, without first initiating NA or NS messages.
3. The host does not send any traffic, and the source PE generates an NS message when the MAC address of the host expires and the EVPN-MAC is withdrawn.

All three scenarios are described in more detail later, where the move of host-66 from source PE-2 to target PE-3 is simulated.

Configuration

Figure 186: Example topology with system IP addresses shows the example topology with PE-1, PE-2, and PE-3 in an EVPN-MPLS network and PE-4, PE-5, and PE-6 in an MPLS network.

Figure 186: Example topology with system IP addresses



The initial configuration includes:

- Cards, MDAs, ports
- Router interfaces
- IS-IS between PE-1, PE-2, PE-3 and between PE-4, PE-5, and PE-6
- LDP between PE-1, PE-2, PE-3 and between PE-4, PE-5, and PE-6
- BGP configured for the EVPN address family on PE-1, PE-2, and PE-3

On PE-1, BGP is configured as follows:

```
# on PE-1:
configure
    router Base
        autonomous-system 64500
        bgp
            family evpn
            vpn-apply-import
            vpn-apply-export
            rapid-withdrawal
            rapid-update evpn
            group "dc"
                type internal
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
        exit
```

The BGP configuration is similar on PE-2 and PE-3.

IPv4 host mobility

The following use cases for IPv4 host mobility are described:

1. Host initiates ARP request or GARP after moving
2. Host initiates non-ARP traffic after moving
3. Host does not send any traffic after moving

IPv4 host mobility case 1: host initiates ARP request or GARP after moving

The service configuration on PE-1 is as follows:

```
# on PE-1:
configure
    service
        vpls 15 name "sbd-15" customer 1 create
            description "R-VPLS 15"
            allow-ip-int-bind
            exit
            bgp
                route-distinguisher 192.0.2.1:15
            exit
            bgp-evpn
                ip-route-advertisement
                evi 15
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
        vprn 16 name "ip-vrf-16" customer 1 create
            interface "evi-15" create
                mac 00:00:00:00:00:01
                vpls "sbd-15"
                    evpn-tunnel
                exit
            exit
            interface "evi-20" create
                address 10.0.20.1/24
                mac 00:00:00:00:1e:20
                vpls "evi-20"
                exit
            exit
            no shutdown
        exit
        vpls 20 name "evi-20" customer 1 create
            description "R-VPLS 20"
            allow-ip-int-bind
            exit
            sap pxc-10.a:20 create
                no shutdown
            exit
            no shutdown
        exit
```

VPRN "ip-vrf-16" has two interfaces: interface "evi-15" toward R-VPLS "sbd-15" and interface "evi-20" toward R-VPLS "evi-20". Host-191 is connected to interface "evi-20" of R-VPLS "evi-20".

PE-2 and PE-3 are configured with an anycast gateway, that is, a VRRP passive instance with the same backup IP address 10.0.0.254 on interface "evi-17" in VPRN "ip-vrf-16". The MAC address under VRRP is by default derived from the Virtual Router ID (VRID), so both PE-2 and PE-3 get MAC address 00:00:5E:00:01:01. The service configuration on PE-2 and PE-3 is similar.

```
# on PE-2:
configure
    service
         vpls 15 name "sbd-15" customer 1 create
            description "R-VPLS 15"
            allow-ip-int-bind
            exit
            bgp
                route-distinguisher 192.0.2.2:15                  # on PE-3: 192.0.2.3:15
            exit
            bgp-evpn
                ip-route-advertisement
                evi 15
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
        vprn 16 name "ip-vrf-16" customer 1 create
            interface "evi-15" create
                mac 00:00:00:00:00:02                             # on PE-3: 00:00:00:00:00:03
                vpls "sbd-15"
                    evpn-tunnel
                exit
            exit
            interface "evi-17" create
                address 10.0.0.2/24                               # on PE-3: 10.0.0.3/24
                mac 00:00:00:00:2f:17                             # on PE-3: 00:00:00:00:3f:17
                arp-host-route
                    populate dynamic
                exit
                arp-timeout 300
                arp-learn-unsolicited
                arp-proactive-refresh
                vrrp 1 passive
```

```
                    backup 10.0.0.254                             # anycast IP address on PE-2,
 PE-3
                    ping-reply
                    traceroute-reply
                exit
                local-proxy-arp
                vpls "evi-17"
                    evpn
                        arp
                            no learn-dynamic
                            no flood-garp-and-unknown-req
                            advertise dynamic
                        exit
                    exit
                exit
            exit
            no shutdown
        exit
        vpls 17 name "evi-17" customer 1 create
            description "R-VPLS 17"
            allow-ip-int-bind
            exit
            bgp
                route-distinguisher 192.0.2.2:17                  # on PE-3: 192.0.2.3:17
            exit
            bgp-evpn
                evi 17
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/1/1:17 create                                   # on PE-3: sap 1/1/2:17
                no shutdown
            exit
            no shutdown
        exit
```

The **arp-host-route>populate dynamic** ensures that route-table ARP-ND host routes are created for dynamic entries, not for static or EVPN entries. The **no learn-dynamic** command prevents PE-2 and PE-3 from learning ARP entries from ARP messages received on an EVPN destination. The **no flood-garp-and-unknown-req** command suppresses CPM-generated ARP to reduce unnecessary ARP flooding.

In this sample topology, an Epipe is used where a failover from the primary to the secondary path simulates a move of host-100 from PE-2 to PE-3. SAP 1/1/1:17 in R-VPLS "evi-17" on PE-2 is connected to a SAP of Epipe 17 on PE-4; SAP 1/1/2:17 in R-VPLS "evi-17" on PE-3 to a SAP of Epipe 17 on PE-5. The service configuration on PE-4 is as follows. The configuration on PE-5 is similar.

```
# on PE-4:
configure
    service
        oper-group "op-grp-1" create
        exit
        sdp 46 mpls create                                        # on PE-5: sdp 56
            far-end 192.0.2.6
            ldp
            keep-alive
```

```
                shutdown
            exit
            no shutdown
        exit
        epipe 17 name "Epipe 17" customer 1 create
            sap 1/1/2:17 create                                   # on PE-5: sap 1/1/1:17
                description "SAP connected to SAP 1/1/1:17 on PE-2"
                monitor-oper-group "op-grp-1"
                no shutdown
            exit
            spoke-sdp 46:17 create                                # on PE-5: spoke-sdp 56:17
                oper-group "op-grp-1"
                no shutdown
            exit
            no shutdown
        exit
```

On PE-6, the service configuration is as follows:

```
# on PE-6:
configure
    service
        sdp 64 mpls create
            far-end 192.0.2.4
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        sdp 65 mpls create
            far-end 192.0.2.5
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        epipe 17 name "Epipe 17" customer 1 create
            endpoint "EP17" create
            exit
            sap 1/2/1:17 create                                   # toward host-100
                no shutdown
            exit
            spoke-sdp 64:17 endpoint "EP17" create
                precedence primary
                no shutdown
            exit
            spoke-sdp 65:17 endpoint "EP17" create
                no shutdown
            exit
            no shutdown
        exit
```

Host-100 is connected to SAP 1/2/1:17 in Epipe 17.

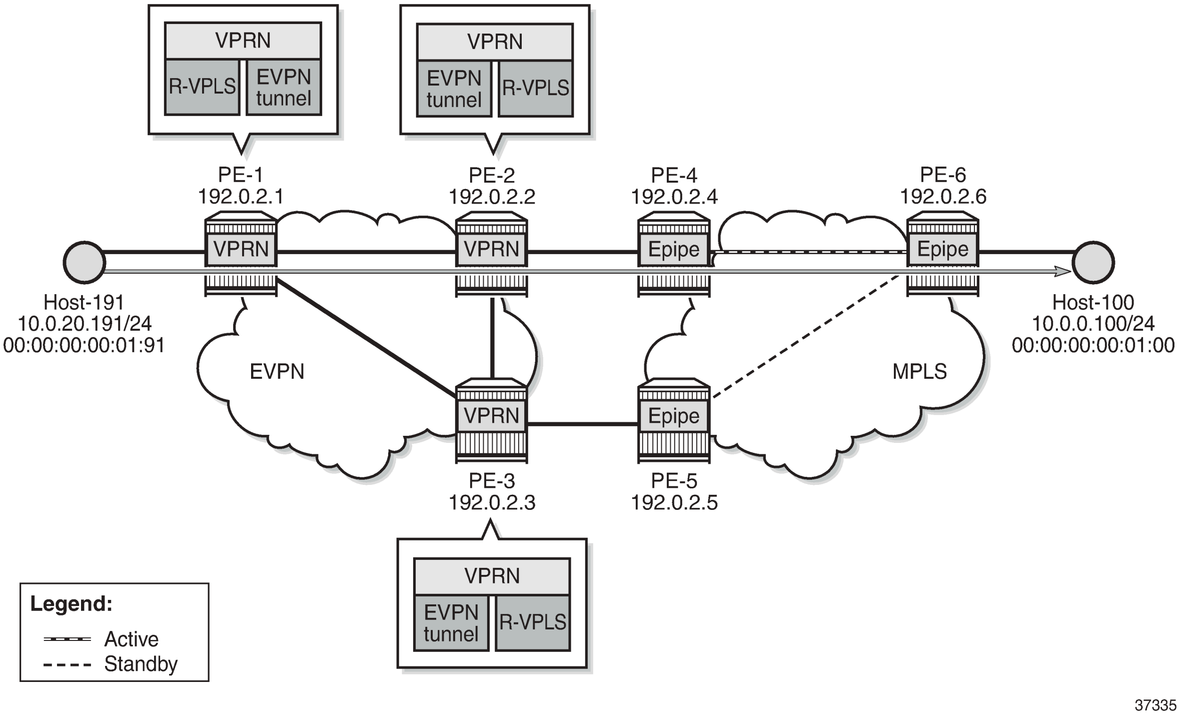
On PE-2 and PE-3, debugging is enabled:

```
# on PE-2, PE-3:
debug
    router "Base"
        bgp
            update
        exit
    exit
    router service-name "ip-vrf-16"
        ip
            arp
            route-table
        exit
    exit
exit
```

Initial phase

Figure 187: Initial situation with forwarding path via PE-2 shows that traffic from host-191 to host-100 is forwarded via PE-1, PE-2, PE-4, and PE-6.

Figure 187: Initial situation with forwarding path via PE-2



Host-191 sends a traceroute to host-100 via PE-2 (10.0.0.2):

```
*A:PE-1# traceroute router 19 10.0.0.100 source 10.0.20.191
traceroute to 10.0.0.100 from 10.0.20.191, 30 hops max, 40 byte packets
  1  10.0.20.1 (10.0.20.1)    2.34 ms  2.35 ms  2.27 ms
  2  10.0.0.2 (10.0.0.2)    3.37 ms  3.29 ms  3.14 ms
  3  10.0.0.100 (10.0.0.100)    5.90 ms  5.28 ms  5.46 ms
```

The ARP table for VPRN "ip-vrf-16" on PE-2 shows that IP address 10.0.0.100 corresponds to MAC address 00:00:00:00:01:00 and is learned dynamically:

```
*A:PE-2# show router 16 arp 10.0.0.100
```

```
===============================================================================
ARP Table (Service: 16)
===============================================================================
IP Address      MAC Address       Expiry    Type   Interface
-------------------------------------------------------------------------------
10.0.0.100      00:00:00:00:01:00 00h04m49s Dyn[I] evi-17
===============================================================================
```

The ARP table for VPRN "ip-vrf-16" on PE-3 shows that IP address 10.0.0.100 is advertised through EVPN:

```
*A:PE-3# show router 16 arp 10.0.0.100

===============================================================================
ARP Table (Service: 16)
===============================================================================
IP Address      MAC Address       Expiry    Type   Interface
-------------------------------------------------------------------------------
10.0.0.100      00:00:00:00:01:00 00h00m00s Evp[I] evi-17
===============================================================================
```

On PE-2, MAC address 00:00:00:00:01:00 is learned on SAP 1/1/1:17 in R-VPLS "evi-17":

```
*A:PE-2# show service id 17 fdb detail

===============================================================================
Forwarding Database, Service 17
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
17         00:00:00:00:01:00 sap:1/1/1:17            L/30     01/28/22 12:45:29
17         00:00:00:00:2e:17 cpm                     Intf     01/28/22 12:43:20
17         00:00:00:00:3f:17 mpls-1:                 EvpnS:P  01/28/22 12:43:29
                             192.0.2.3:524283
           ldp:65538
17         00:00:5e:00:01:01 cpm                     Intf     01/28/22 12:43:20
-------------------------------------------------------------------------------
No. of MAC Entries: 4
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

On PE-3, the FDB for R-VPLS "evi-17" shows that MAC address 00:00:00:00:01:00 is advertised through EVPN:

```
*A:PE-3# show service id 17 fdb mac 00:00:00:00:01:00

===============================================================================
Forwarding Database, Service 17
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
17         00:00:00:00:01:00 mpls-1:                 Evpn     01/28/22 12:45:29
                             192.0.2.2:524283
           ldp:65537
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The route table for VPRN "ip-vrf-16" on PE-2 shows an ARP-ND host route with preference 1 for prefix 10.0.0.100/32:

```
*A:PE-2# show router 16 route-table 10.0.0.100

===============================================================================
Route Table (Service: 16)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.0.100/32                                 Remote  ARP-ND    00h02m44s  1
       10.0.0.100                                                   0
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The route table for VPRN "ip-vrf-16" on PE-3 shows an EVPN host route for prefix 10.0.0.100/32:

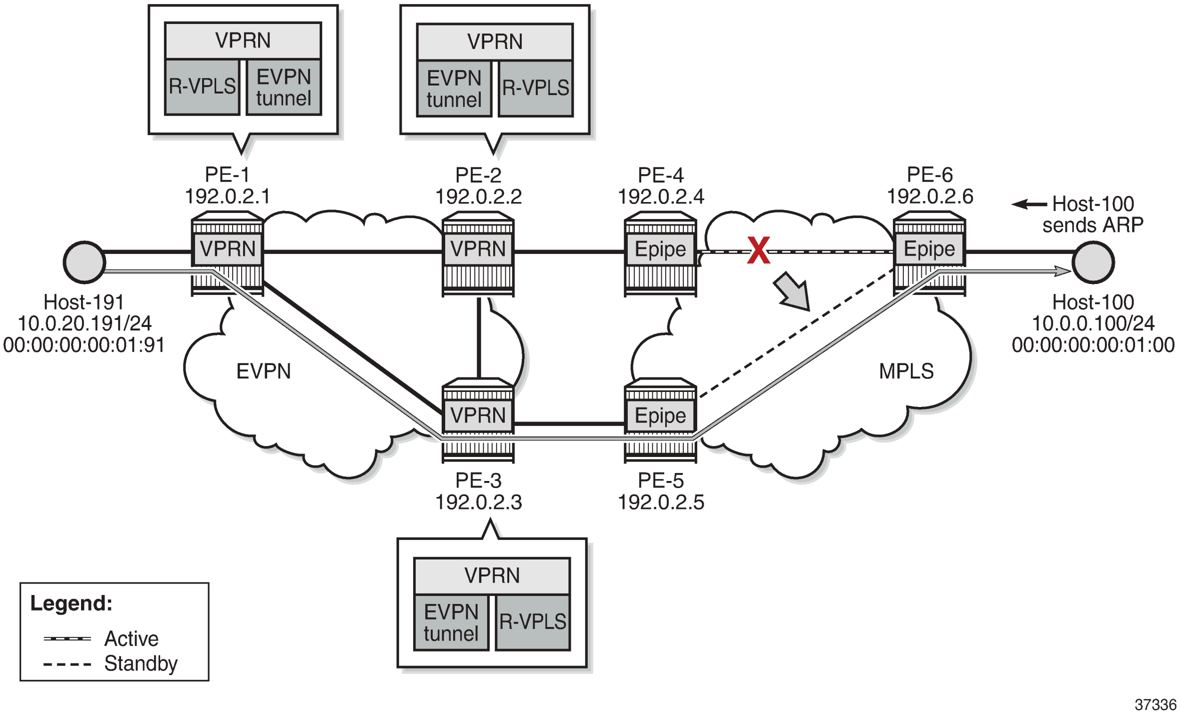
```
*A:PE-3# show router 16 route-table 10.0.0.100

===============================================================================
Route Table (Service: 16)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.0.100/32                                 Remote  EVPN-IFF  00h02m43s  169
       evi-15 (ET-00:00:00:00:00:02)                                0
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

PE-3 receives the IP and MAC addresses of host-100 as EVPN type. PE-3 must not learn these IP and MAC addresses as dynamic because PE-3 must be prevented from advertising an RT5 route. If PE-3 advertised prefix 10.0.0.100, then PE-1 could select PE-3 as next hop to reach host-100, causing an undesired hairpinning forwarding behavior.

Host-100 sends an ARP request or GARP after moving

Figure 188: Host-100 sends an ARP request or GARP after switchover shows a switchover from the active to the standby path where host-100 sends an ARP request or GARP and its IP and MAC addresses are learned on PE-3 instead of PE-2. The failure is simulated by disabling the SDP from PE-4 to PE-6.

Figure 188: Host-100 sends an ARP request or GARP after switchover



Due to the SDP failure on PE-4, the initial path can no longer be used. Host-100 sends an ARP request or GARP with its IP and MAC addresses. In the following example, PE-3 receives the following ARP request and replies to it:

```
1 2022/01/28 12:49:45.684 UTC MINOR: DEBUG #2001 vprn16 PIP
"PIP: ARP
instance 2 (16), interface index 6 (evi-17),
ARP ingressing on evi-17
   Who has 10.0.0.254 ? Tell 10.0.0.100
"
2 2022/01/28 12:49:45.684 UTC MINOR: DEBUG #2001 vprn16 PIP
"PIP: ARP
instance 2 (16), interface index 6 (evi-17),
ARP egressing on evi-17
   10.0.0.254 is at 00:00:5e:00:01:01
"
```

The Route Table Manager (RTM) for prefix 10.0.0.100 in VPRN "ip-vrf-16" is modified with preference 1 and owner ARP-ND. This behavior is due to the **arp-host-route populate dynamic** command.

```
3 2022/01/28 12:49:45.684 UTC MINOR: DEBUG #2001 vprn16 PIP
"PIP: ROUTE
instance 2 (16), RTM MODIFY event
   New Route Info
      prefix: 10.0.0.100/32 (0x119549018)  preference: 1   metric: 0
                             backup metric: 0  owner: ARP-ND ownerId: 0
      1 ecmp hops  0 backup hops:
         hop 0:  10.0.0.100 @ if 6, weight 0
"
```

PE-3 sends an RT5 for prefix 10.0.0.100/32 to PE-1 and PE-2:

```
4 2022/01/28 12:49:45.685 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 90
    Flag: 0x90 Type: 14 Len: 45 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-IP-PREFIX Len: 34 RD: 192.0.2.3:15, tag: 0,
                  ip_prefix: 10.0.0.100/32 gw_ip 0.0.0.0
                  Label: 8388544 (Raw Label: 0x7fffc0)
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:15
        mac-nh:00:00:00:00:00:03
        bgp-tunnel-encap:MPLS
"
```

PE-3 sends EVPN-MAC routes for MAC 00:00:00:00:01:00 with an increased sequence number for MAC mobility: one EVPN-MAC route with MAC address 00:00:00:00:01:00 and IP address 10.0.0.100 and another EVPN-MAC route with MAC address 00:00:00:00:01:00 only and a null IP address.

```
5 2022/01/28 12:49:45.685 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 128
    Flag: 0x90 Type: 14 Len: 83 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-MAC Len: 37 RD: 192.0.2.3:17 ESI: ESI-0, tag: 0, mac len: 48
                  mac: 00:00:00:00:01:00, IP len: 4, IP: 10.0.0.100, label1: 8388528
        Type: EVPN-MAC Len: 33 RD: 192.0.2.3:17 ESI: ESI-0, tag: 0, mac len: 48
                  mac: 00:00:00:00:01:00, IP len: 0, IP: NULL, label1: 8388528
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:17
        bgp-tunnel-encap:MPLS
        mac-mobility:Seq:1
"
```

The FDB for R-VPLS "evi-17" shows that MAC address 00:00:00:00:01:00 is dynamically learned on SAP 1/1/2:17 on PE-3:

```
*A:PE-3# show service id 17 fdb mac 00:00:00:00:01:00

===============================================================================
Forwarding Database, Service 17
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
17         00:00:00:00:01:00 sap:1/1/2:17            L/14     01/28/22 12:49:46
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
```

```
===============================================================================
```

On PE-2, the FDB for R-VPLS "evi-17" is updated and PE-2 withdraws its EVPN-MAC route based on the higher sequence number of the received EVPN-MAC route for MAC address 00:00:00:00:01:00 with next hop 192.0.2.3:

```
*A:PE-2# show service id 17 fdb mac 00:00:00:00:01:00

===============================================================================
Forwarding Database, Service 17
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
17         00:00:00:00:01:00 mpls-1:                 Evpn     01/28/22 12:49:46
                             192.0.2.3:524283
           ldp:65538
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

On PE-3, the ARP table for VPRN "ip-vrf-16" shows that IP address 10.0.0.100 is learned dynamically on interface "evi-17":

```
*A:PE-3# show router 16 arp 10.0.0.100

===============================================================================
ARP Table (Service: 16)
===============================================================================
IP Address      MAC Address       Expiry    Type   Interface
-------------------------------------------------------------------------------
10.0.0.100      00:00:00:00:01:00 00h04m40s Dyn[I] evi-17
===============================================================================
```

On PE-2, the ARP table for VPRN "ip-vrf-16" shows that the entry for IP address 10.0.0.100 is updated from dynamic to type EVPN:

```
*A:PE-2# show router 16 arp 10.0.0.100

===============================================================================
ARP Table (Service: 16)
===============================================================================
IP Address      MAC Address       Expiry    Type   Interface
-------------------------------------------------------------------------------
10.0.0.100      00:00:00:00:01:00 00h00m00s Evp[I] evi-17
===============================================================================
```

An ARP entry's change from dynamic to EVPN triggers a CPM-generated ARP request from PE-2, but the configured **no flood-garp-and-unknown-req** command prevents PE-2 from flooding the ARP request to EVPN destinations such as PE-3.

On PE-3, the route table for VPRN "ip-vrf-16" shows an ARP-ND host route for prefix 10.0.0.100:

```
*A:PE-3# show router 16 route-table 10.0.0.100

===============================================================================
Route Table (Service: 16)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
```

```
-------------------------------------------------------------------------------
10.0.0.100/32                                 Remote  ARP-ND    00h01m32s  1
       10.0.0.100                                                   0
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The route table for VPRN "ip-vrf-16" for prefix 10.0.0.100 shows that PE-2 removed its ARP-ND host route and the received EVPN route from PE-3 is used instead:

```
*A:PE-2# show router 16 route-table 10.0.0.100

===============================================================================
Route Table (Service: 16)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.0.100/32                                 Remote  EVPN-IFF  00h01m31s  169
       evi-15 (ET-00:00:00:00:00:03)                                0
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

IPv4 host mobility case 2: host sends traffic without first initiating an ARP request or GARP after moving

In use cases 2 and 3, the configuration of VPRN "ip-vrf-16" is modified on PE-2 and PE-3. The only difference from case 1 is that the **flood-garp-and-unknown-req** is configured, which is the default setting. The VPRN configuration on PE-2 is as follows:

```
# on PE-2:
configure
    service
        vprn 16 name "ip-vrf-16" customer 1 create
            interface "evi-15" create
                mac 00:00:00:00:00:02                            # on PE-3: 00:00:00:00:00:03
                vpls "sbd-15"
                    evpn-tunnel
                exit
            exit
            interface "evi-17" create
                address 10.0.0.2/24
                mac 00:00:00:00:2f:17
                arp-host-route
                    populate dynamic
                exit
                arp-timeout 300
                arp-learn-unsolicited
                arp-proactive-refresh
                vrrp 1 passive
```

```
                    backup 10.0.0.254
                    ping-reply
                    traceroute-reply
                exit
                local-proxy-arp
                vpls "evi-17"
                    evpn
                        arp
                            no learn-dynamic
                            flood-garp-and-unknown-req            # default
                            advertise dynamic
                        exit
                    exit
                exit
            exit
            no shutdown
```

Initial forwarding path

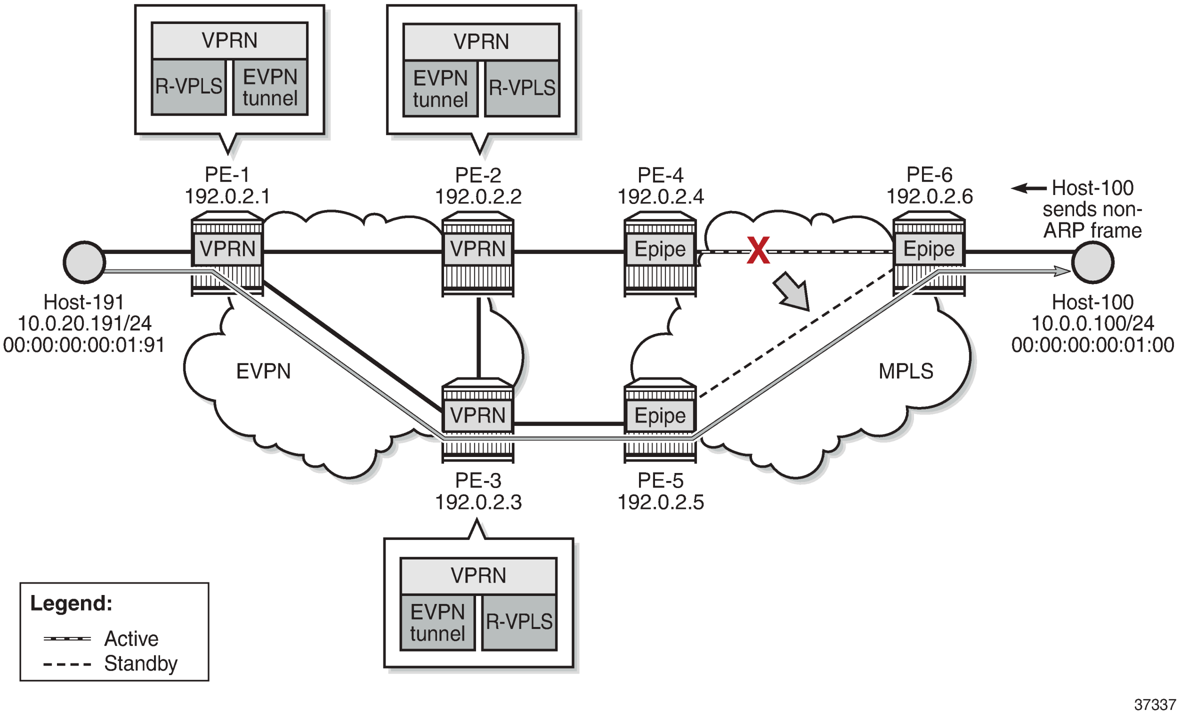
The initial forwarding path via PE-2 is restored by enabling the SDP from PE-4 to PE-6. The route table for VPRN "ip-vrf-16" on PE-2 shows the following ARP-ND host route for prefix 10.0.0.100/32:

```
*A:PE-2# show router 16 route-table 10.0.0.100

===============================================================================
Route Table (Service: 16)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.0.100/32                                 Remote  ARP-ND    00h03m46s  1
       10.0.0.100                                                   0
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

Host-100 generates non-ARP traffic after moving

On PE-4, the SDP from PE-4 to PE-6 is disabled, causing a switchover to the standby path. Figure 189: Host sends non-ARP frame after switchover shows the path after switchover. Host-100 generates non-ARP traffic after moving.

Figure 189: Host sends non-ARP frame after switchover



Host-100 sends a non-ARP frame with MAC source address 00:00:00:00:01:00 to host-191. The following steps occur:

1. PE-3 receives this frame with MAC 00:00:00:00:01:00 and updates its FDB.
2. PE-3 advertises an EVPN-MAC route for MAC 00:00:00:00:01:00 (with a null IP address) with a higher sequence number.
3. PE-2 receives this EVPN MAC route, updates its FDB and withdraws its EVPN-MAC routes for MAC 00:00:00:00:01:00.
4. The FDB update for MAC 00:00:00:00:01:00 triggers PE-2 to send an ARP request for MAC 00:00:00:00:01:00.
5. PE-2 is configured with **flood-garp-and-unknown-req**, so the ARP request is flooded to the EVPN destinations PE-1 and PE-3. PE-3 floods this ARP request to its SAPs and SDP-bindings; in this case, to SAP 1/1/2:17.
6. When the ARP request reaches host-100, it sends an ARP reply to the anycast IP address 10.0.0.254. This ARP reply is received by PE-3.
7. When PE-3 receives the ARP reply, it updates the ARP entry for 10.0.0.100 to type dynamic instead of type EVPN.
8. PE-3 is configured with **populate dynamic**, so it advertises an RT5 for prefix 10.0.0.100/32. Also, MAC 00:00:00:00:01:00 is now learned in ARP as local, so PE-3 sends an EVPN-MAC route with MAC 00:00:00:00:01:00 and IP prefix 10.0.0.100.

9. PE-2 receives the EVPN routes and updates the ARP entry for prefix 10.0.0.100 from type dynamic to type EVPN. PE-2 also removes its ARP-ND host route from the route table and withdraws its RT5 for prefix 10.0.0.100/32.

On PE-3, the route for prefix 10.0.0.100/32 is an ARP-ND host route:

```
*A:PE-3# show router 16 route-table 10.0.0.100

================================================================================
Route Table (Service: 16)
================================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
--------------------------------------------------------------------------------
10.0.0.100/32                                 Remote  ARP-ND    00h01m05s  1
       10.0.0.100                                                   0
--------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
================================================================================
```

IPv4 host mobility case 3: host does not send any traffic after moving

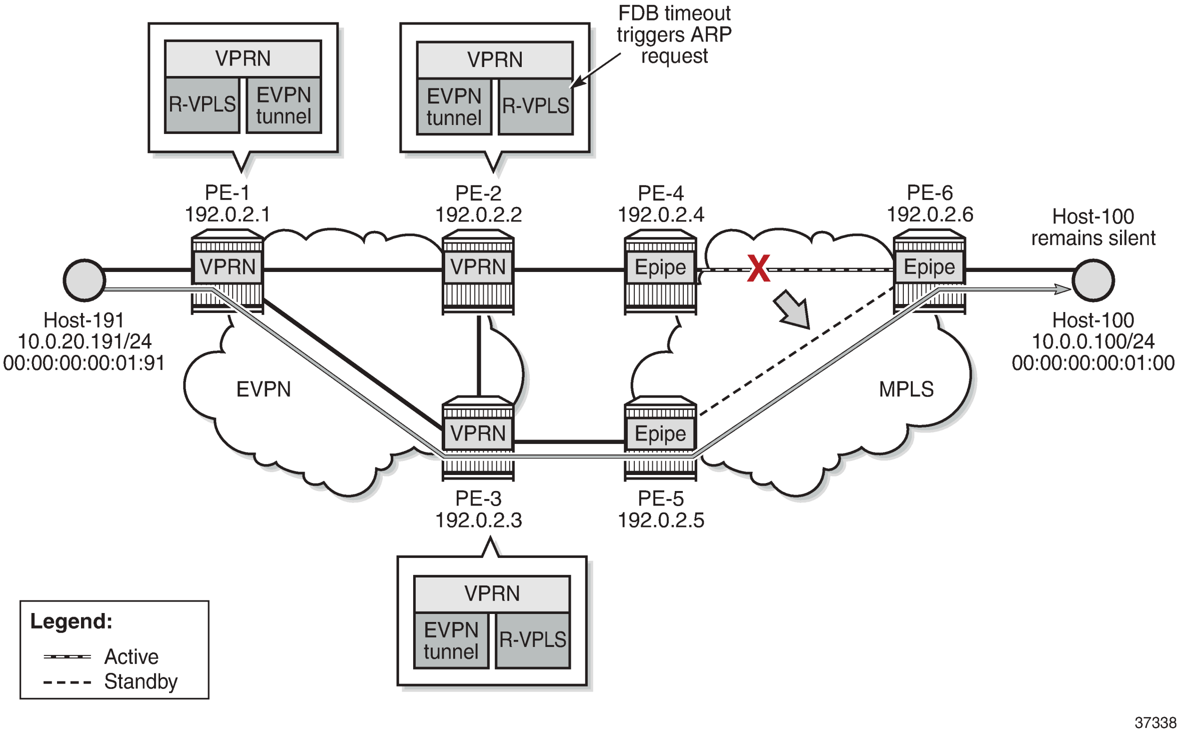
The service configuration on PE-2 and PE-3 remains the same as in use case 2.

The forwarding path is restored by enabling the SDP from PE-4 to PE-6, so the initial situation is the same as in the preceding cases. PE-2 learns MAC address 00:00:00:00:01:00 on its local SAP 1/1/1:17, as follows:

```
*A:PE-2# show service id 17 fdb detail

================================================================================
Forwarding Database, Service 17
================================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
--------------------------------------------------------------------------------
17         00:00:00:00:01:00 sap:1/1/1:17            L/0      01/28/22 13:10:34
17         00:00:00:00:2f:17 cpm                     Intf     01/28/22 12:53:11
17         00:00:00:00:3f:17 mpls-1:                 EvpnS:P  01/28/22 12:43:29
                             192.0.2.3:524283
           ldp:65538
17         00:00:5e:00:01:01 cpm                     Intf     01/28/22 12:43:20
--------------------------------------------------------------------------------
No. of MAC Entries: 4
--------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
================================================================================
```

The SDP from PE-4 to PE-6 is disabled and host-100 does not send any traffic, as shown in Figure 190: Host does not send any traffic after switchover.

Figure 190: Host does not send any traffic after switchover



The following steps occur:

1. When MAC 00:00:00:00:01:00 ages out in the FDB of R-VPLS 17 on PE-2, PE-2 withdraws the EVPN-MAC routes for MAC 00:00:00:00:01:00. The update for MAC 00:00:00:00:01:00 triggers PE-2 to send an ARP request for 10.0.0.100.

```
# on PE-2:
99 2022/01/28 13:12:04.028 UTC MINOR: DEBUG #2001 vprn16 PIP
"PIP: ARP
instance 2 (16), interface index 6 (evi-17),
ARP egressing on evi-17
   Who has 10.0.0.100 ? Tell 10.0.0.254
"
```

```
101 2022/01/28 13:12:04.028 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 42
    Flag: 0x90 Type: 15 Len: 38 Multiprotocol Unreachable NLRI:
        Address Family EVPN
        Type: EVPN-MAC Len: 33 RD: 192.0.2.2:17 ESI: ESI-0, tag: 0, mac len: 48
                       mac: 00:00:00:00:01:00, IP len: 0, IP: NULL, label1: 0
"
```

2. PE-2 is configured with **flood-garp-and-unknown-req**. PE-2 floods the CPM-generated ARP request to PE-3. PE-3 forwards the ARP request to host-100.
3. Host-100 sends an ARP reply that is received by PE-3. PE-3 updates its FDB and ARP tables.

4. The FDB update on PE-3 makes PE-3 advertise an EVPN-MAC route for MAC 00:00:00:00:01:00 (with a null IP address). The ARP update makes PE-3 advertise an EVPN-MAC route with MAC 00:00:00:00:01:00 and IP prefix 10.0.0.100. PE-2 receives two EVPN-MAC routes from PE-3:

```
103 2022/01/28 13:12:04.033 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 128
    Flag: 0x90 Type: 14 Len: 83 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-MAC Len: 37 RD: 192.0.2.3:17 ESI: ESI-0, tag: 0, mac len: 48
                  mac: 00:00:00:00:01:00, IP len: 4, IP: 10.0.0.100, label1: 8388528
        Type: EVPN-MAC Len: 33 RD: 192.0.2.3:17 ESI: ESI-0, tag: 0, mac len: 48
                  mac: 00:00:00:00:01:00, IP len: 0, IP: NULL, label1: 8388528
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:17
        bgp-tunnel-encap:MPLS
        mac-mobility:Seq:5
"
```

5. PE-3 is configured with **populate dynamic**, so it advertises an RT5 for prefix 10.0.0.100/32. In the route table for VPRN "ip-vrf-16", the route for IP prefix 10.0.0.100/32 is ARP-ND host route. PE-2 receives the following RT5 route from PE-3:

```
102 2022/01/28 13:12:04.033 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 90
    Flag: 0x90 Type: 14 Len: 45 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-IP-PREFIX Len: 34 RD: 192.0.2.3:15, tag: 0,
         ip_prefix: 10.0.0.100/32 gw_ip 0.0.0.0 Label: 8388544 (Raw Label: 0x7fffc0)
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:15
        mac-nh:00:00:00:00:00:03
        bgp-tunnel-encap:MPLS
"
```

6. PE-2 receives the EVPN routes and updates its FDB and ARP tables. When the ARP entry changes its type from dynamic to EVPN, PE-2 withdraws its RT5 route.

On PE-2, the FDB for R-VPLS 17 shows an EVPN route for MAC 00:00:00:00:01:00:

```
*A:PE-2# show service id 17 fdb detail

===============================================================================
Forwarding Database, Service 17
===============================================================================
ServId     MAC               Source-Identifier    Type     Last Change
            Transport:Tnl-Id                      Age
-------------------------------------------------------------------------------
```

```
17         00:00:00:00:01:00 mpls-1:                Evpn     01/28/22 13:12:04
                             192.0.2.3:524283
           ldp:65538
17         00:00:00:00:2f:17 cpm                    Intf     01/28/22 12:53:11
17         00:00:00:00:3f:17 mpls-1:                EvpnS:P  01/28/22 12:43:29
                             192.0.2.3:524283
           ldp:65538
17         00:00:5e:00:01:01 cpm                    Intf     01/28/22 12:43:20
-------------------------------------------------------------------------------
No. of MAC Entries: 4
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

IPv6 host mobility

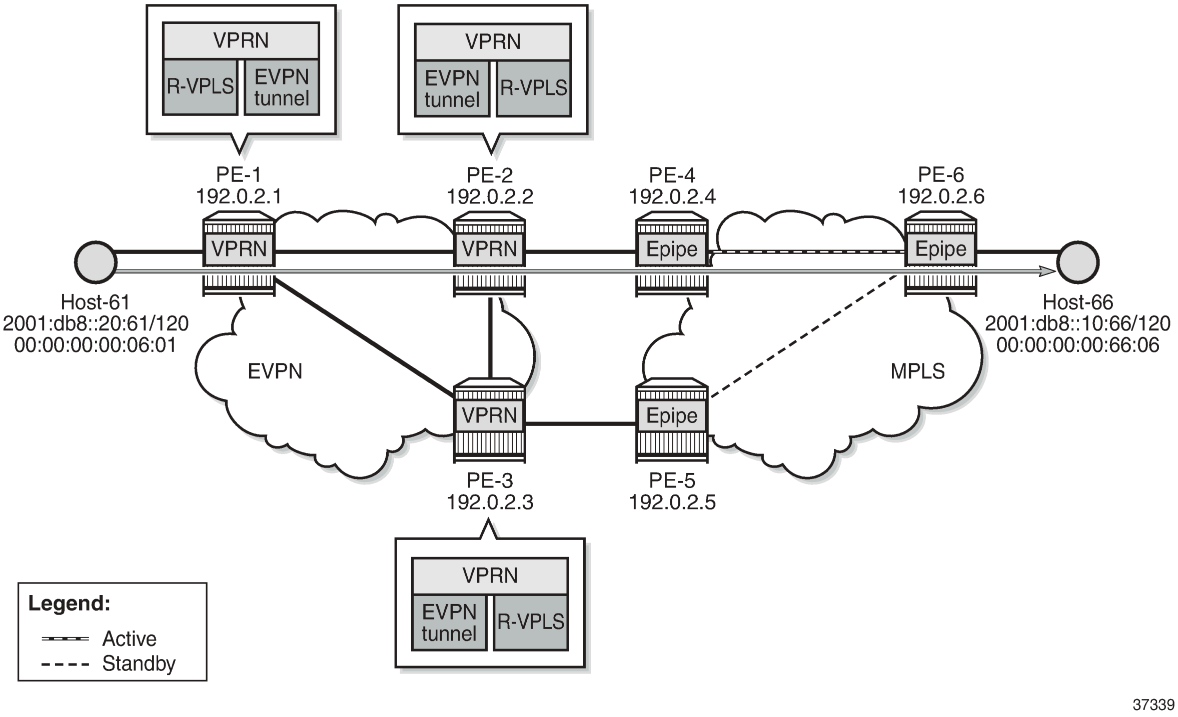
The following use cases for IPv6 host mobility are described:

1. Host initiates an unsolicited NA message after moving
2. Host sends non-ND traffic after moving
3. Host does not send any traffic after moving

The configuration is identical in these use cases.

Figure 191: Example topology for initial forwarding path via PE-2 with IPv6 addresses shows the topology with IPv6 addresses for host-61 and host-66.

Figure 191: Example topology for initial forwarding path via PE-2 with IPv6 addresses



The services are the following:

- R-VPLS "sbd-5" on PE-1, PE-2, and PE-3
- VPRN "ip-vrf-6" on PE-1, PE-2, and PE-3
- R-VPLS "evi-10" on PE-1; R-VPLS "evi-7" on PE-2 and PE-3
- Epipe "Epipe 7" on PE-4, PE-5, and PE-6
- Host-61 is connected to R-VPLS "evi-10" on PE-1
- Host-66 is connected to Epipe "Epipe 7" on PE-6

The service configuration on PE-1 is as follows:

```
# on PE-1:
configure
    service
        vpls 5 name "sbd-5" customer 1 create
            description "R-VPLS 5"
            allow-ip-int-bind
            exit
            bgp
                route-distinguisher 192.0.2.1:5
            exit
            bgp-evpn
                ip-route-advertisement
                evi 5
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
        vprn 6 name "ip-vrf-6" customer 1 create
            interface "evi-5" create
                mac 00:00:00:00:06:01
                ipv6
                exit
                vpls "sbd-5"
                    evpn-tunnel
                exit
            exit
            interface "evi-10" create
                mac 00:00:00:06:1e:20
                ipv6
                    address 2001:db8::20:1/120
                exit
                vpls "evi-10"
                exit
            exit
            no shutdown
        exit
        vpls 10 name "evi-10" customer 1 create
            description "R-VPLS 10"
            allow-ip-int-bind
            exit
            stp
                shutdown
            exit
            sap pxc-10.a:10 create
                no shutdown
            exit
            no shutdown
        exit
```

The service configuration on PE-2 is as follows. The service configuration on PE-3 is similar.

```
# on PE-2:
configure
    service
        vpls 5 name "sbd-5" customer 1 create
            description "R-VPLS 5"
            allow-ip-int-bind
            exit
            bgp
                route-distinguisher 192.0.2.2:5                  # on PE-3: 192.0.2.3:5
            exit
            bgp-evpn
                ip-route-advertisement
                evi 5
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
        vprn 6 name "ip-vrf-6" customer 1 create
            interface "evi-5" create
                mac 00:00:00:00:06:02                            # on PE-3: 00:00:00:00:06:03
                ipv6
                exit
                vpls "sbd-5"
                    evpn-tunnel
                exit
            exit
            interface "evi-7" create
                mac 00:00:00:00:2f:07                            # on PE-3: 00:00:00:00:3f:07
                ipv6
                    address 2001:db8::10:2/120                   # on PE-3: 2001:db8::10:3/120
                    link-local-address fe80::10:2 dad-disable    # on PE-3: fe80::10:3
                    nd-learn-unsolicited both
                    nd-proactive-refresh both
                    nd-host-route
                        populate dynamic
                    exit
                    local-proxy-nd
                    vrrp 1 passive
                        backup fe80::10:fe
                        ping-reply
                        traceroute-reply
                    exit
                exit
                vpls "evi-7"
                    evpn
                        nd
                            no learn-dynamic
                            advertise dynamic
                        exit
                    exit
                exit
            exit
            no shutdown
        exit
        vpls 7 name "evi-7" customer 1 create
            description "R-VPLS 7"
            allow-ip-int-bind
            exit
            bgp
                route-distinguisher 192.0.2.2:7                  # on PE-3: 192.0.2.3:7
            exit
            bgp-evpn
                evi 7
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/1/1:7 create                                   # on PE-3: sap 1/1/2:7
                no shutdown
            exit
            no shutdown
        exit
```

Debugging is enabled on PE-2 and PE-3:

```
# on PE-2, PE-3:
debug
    router "Base"
        bgp
            update
        exit
    exit
    router service-name "ip-vrf-6"
        ip
            route-table
            icmp6 "evi-7"
            neighbor "evi-7"
        exit
    exit
exit
```

Initially, the traceroute from host-66 to host-61 is via PE-2 (2001:db8::10:2):

```
*A:PE-6# traceroute router 8 2001:db8::20:61 source 2001:db8::10:66
traceroute to 2001:db8::20:61 from 2001:db8::10:66, 30 hops max, 60 byte packets
  1  2001:db8::10:2 (2001:db8::10:2)    8.86 ms  3.79 ms  4.00 ms
  2  ::  * * *
  3  2001:db8::20:61 (2001:db8::20:61)    11.5 ms  5.45 ms  5.52 ms
```

The following route table on PE-2 shows an ARP-ND host route for prefix 2001:db8::10:66/128:

```
*A:PE-2# show router 6 route-table 2001:db8::10:66
```

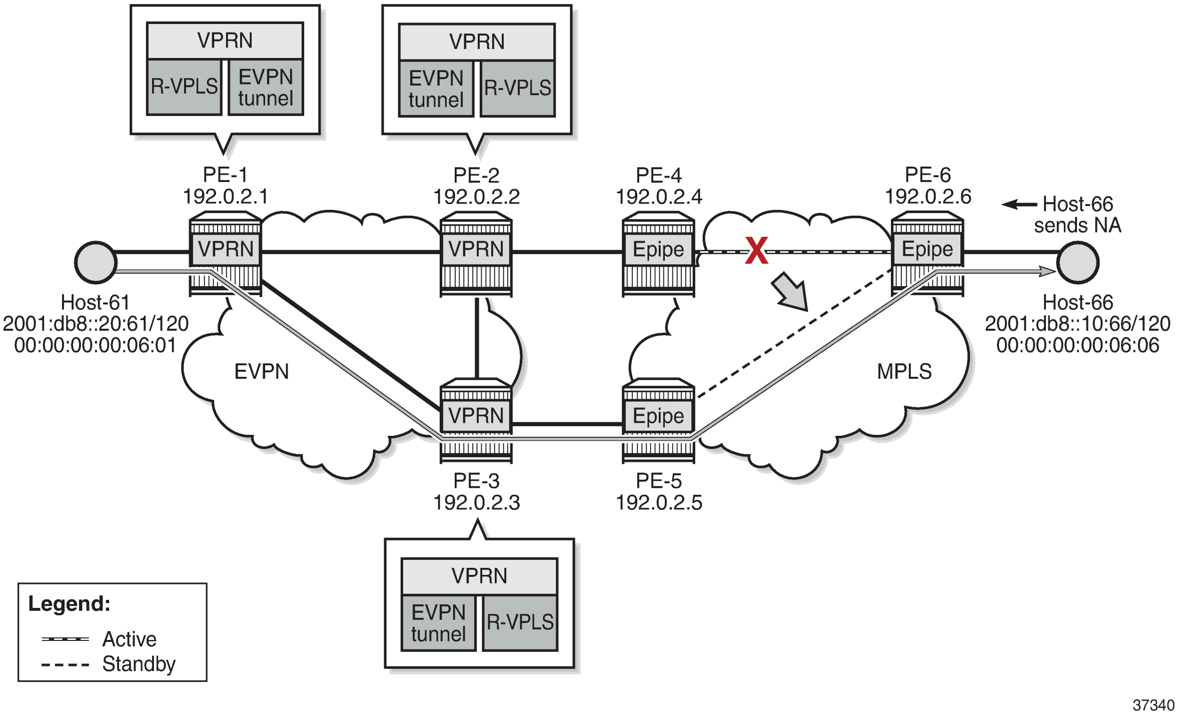
```
===============================================================================
IPv6 Route Table (Service: 6)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
2001:db8::10:66/128                           Remote  ARP-ND    00h00m33s  1
       2001:db8::10:66                                              0
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

IPv6 host mobility case 1: host initiates an unsolicited NA message after moving

On PE-2 and PE-3, the **nd-learn-unsolicited** command is configured on interface "evi-7" in VPRN "ip-vrf-6". When an unsolicited NA message is received, a stale neighbor is created. If **nd-host-route>populate dynamic** is enabled, a confirmation message is sent for all the neighbor entries created as stale, and if confirmed, the corresponding ARP-ND routes are added to the route table.

Disabling SDP 46 on PE-4 causes a failover from the primary path via PE-4 to the secondary path via PE-5, simulating host-66 moving from PE-2 to PE-3. To trigger an unsolicited NA message from host-66, its MAC address 00:00:00:00:66:06 is replaced by MAC address 00:00:00:00:06:06. Figure 192: Host-66 sends unsolicited NA message after switchover shows that host-66 sends an unsolicited NA message.

Figure 192: Host-66 sends unsolicited NA message after switchover



Host-66 advertises its new MAC address in unsolicited NA messages. PE-3 receives the following NA messages from host-66. PE-2 also receives the NA messages, but it rejects NA messages received on interface "evi-7" when **no learn-dynamic** is configured:

```
3 2022/01/28 13:19:55.747 UTC MINOR: DEBUG #2001 vprn6 TIP
"TIP: ICMP6_PKT
ICMP6 ingressing on evi-7 (vprn6):
   fe80::10:6 -> ff02::1
   Type: Neighbor Advertisement (136)
   Code: No Code (0)
      Tgt Addr: 2001:db8::10:66
      Flags   : Router Override
      Option  : Tgt Link Layer Addr 00:00:00:00:06:06
"
```

```
1 2022/01/28 13:19:55.747 UTC MINOR: DEBUG #2001 vprn6 TIP
"TIP: ICMP6_PKT
ICMP6 ingressing on evi-7 (vprn6):
   fe80::10:6 -> ff02::1
   Type: Neighbor Advertisement (136)
   Code: No Code (0)
      Tgt Addr: fe80::10:6
      Flags   : Router Override
      Option  : Tgt Link Layer Addr 00:00:00:00:06:06
"
```

PE-3 learns the MAC address dynamically:

```
*A:PE-3# show service id 7 fdb mac 00:00:00:00:06:06

================================================================================
Forwarding Database, Service 7
================================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                          Age
--------------------------------------------------------------------------------
7          00:00:00:00:06:06 sap:1/1/2:7             L/90     01/28/22 13:19:56
--------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
================================================================================
```

PE-3 sends CPM-generated NS messages that are also flooded to the EVPN destinations. The **no learn-dynamic** command prevents PE-2 from learning MAC addresses dynamically on an EVPN connection.

PE-3 sends an EVPN-MAC update to PE-2 and MAC address 00:00:00:00:06:06 appears in the FDB on PE-2 as an EVPN entry:

```
*A:PE-2# show service id 7 fdb mac 00:00:00:00:06:06

================================================================================
Forwarding Database, Service 7
================================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                          Age
--------------------------------------------------------------------------------
7          00:00:00:00:06:06 mpls-1:                 Evpn     01/28/22 13:19:56
                             192.0.2.3:524281
           ldp:65538
--------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
```

```
===============================================================================
```

The route table for VPRN "ip-vrf-6" on PE-3 shows an ARP-ND entry for destination prefix 2001:db8::10:66/128, as follows:

```
*A:PE-3# show router 6 route-table 2001:db8::10:66

===============================================================================
IPv6 Route Table (Service: 6)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
2001:db8::10:66/128                           Remote  ARP-ND    00h02m37s  1
       2001:db8::10:66                                              0
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

On PE-2, the route table for VPRN "ip-vrf-6" shows an EVPN entry for prefix 2001:db8::10:66/128:

```
*A:PE-2# show router 6 route-table 2001:db8::10:66

===============================================================================
IPv6 Route Table (Service: 6)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
2001:db8::10:66/128                           Remote  EVPN-IFF  00h02m39s  169
       fe80::7:b0d1:3fa3:2f60-"evi-5"                               0
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

IPv6 host mobility case 2: host sends non-ND traffic after moving,

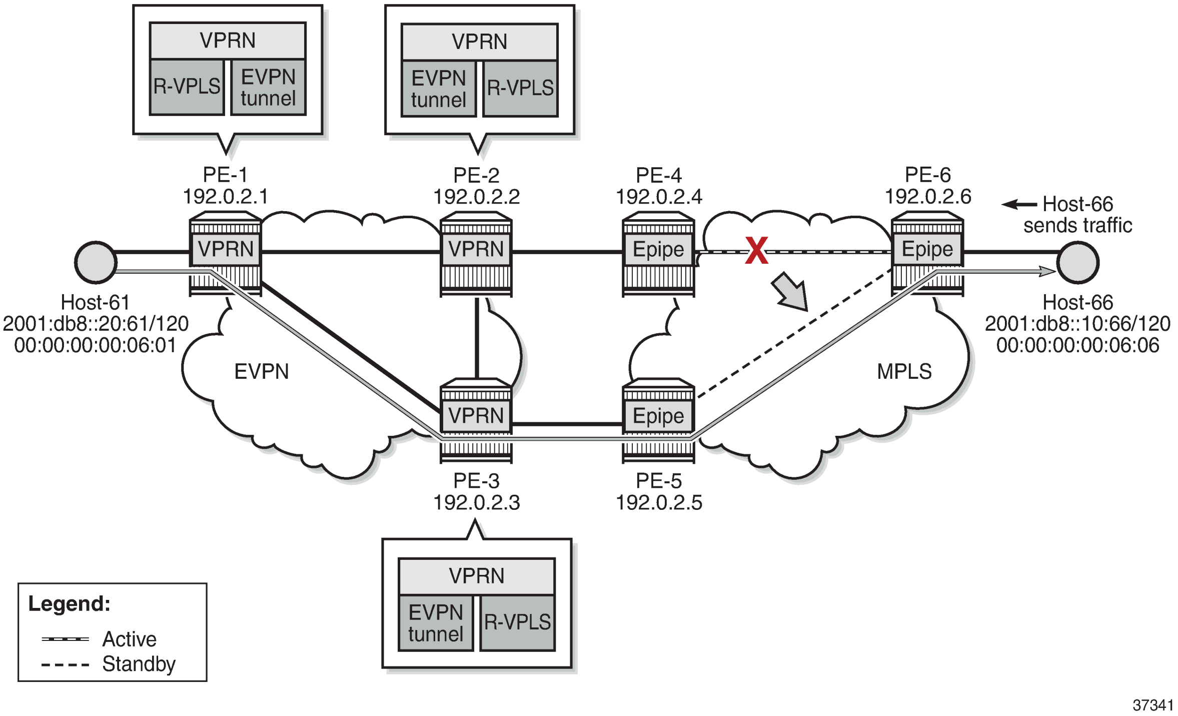
The service configuration is the same as in the use case 1. The only difference from use case 1 is the type of message that is sent by host-66 after moving.

Initially, the traceroute from host-66 to host-61 is via PE-2 (2001:db8::10:2):

```
*A:PE-6# traceroute router 8 2001:db8::20:61 source 2001:db8::10:66
traceroute to 2001:db8::20:61 from 2001:db8::10:66, 30 hops max, 60 byte packets
  1  2001:db8::10:2 (2001:db8::10:2)    7.88 ms  3.85 ms  3.67 ms
  2  ::  * * *
  3  2001:db8::20:61 (2001:db8::20:61)    11.0 ms  5.19 ms  5.30 ms
```

A switchover from the primary path to the secondary path takes place, so host-66 moves from PE-2 to PE-3. Figure 193: Host generates non-ND traffic after switchover shows that host-66 sends non-ND traffic after moving.

Figure 193: Host generates non-ND traffic after switchover



The traceroute from host-66 to host-61 is via PE-3 (2001:db8::10:3) instead of PE-2, as follows:

```
*A:PE-6# traceroute router 8 2001:db8::20:61 source 2001:db8::10:66
traceroute to 2001:db8::20:61 from 2001:db8::10:66, 30 hops max, 60 byte packets
  1  2001:db8::10:3 (2001:db8::10:3)    4.16 ms  4.11 ms  4.01 ms
  2  ::  * * *
  3  2001:db8::20:61 (2001:db8::20:61)    5.56 ms  5.58 ms  5.53 ms
```

On PE-3, MAC address 00:00:00:00:06:06 from host-66 is learned on the local SAP 1/1/2:17:

```
*A:PE-3# show service id 7 fdb mac 00:00:00:00:06:06

===============================================================================
Forwarding Database, Service 7
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
7          00:00:00:00:06:06 sap:1/1/2:7             L/0      01/28/22 13:29:59
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

PE-3 advertises MAC address 00:00:00:00:06:06 from host-66 in three EVPN-MAC routes: one with the global IP address 2001:db8::10:66, one with the link local IP address fe80::200:ff:fe00:606, and one with a null IP address. PE-2 receives the following EVPN-MAC routes from PE-3:

```
59 2022/01/28 13:29:59.437 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 191
    Flag: 0x90 Type: 14 Len: 146 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-MAC Len: 49 RD: 192.0.2.3:7 ESI: ESI-0, tag: 0, mac len: 48
            mac: 00:00:00:00:06:06, IP len: 16, IP: fe80::10:6, label1: 8388496
        Type: EVPN-MAC Len: 49 RD: 192.0.2.3:7 ESI: ESI-0, tag: 0, mac len: 48
            mac: 00:00:00:00:06:06, IP len: 16, IP: 2001:db8::10:66, label1: 8388496
        Type: EVPN-MAC Len: 33 RD: 192.0.2.3:7 ESI: ESI-0, tag: 0, mac len: 48
            mac: 00:00:00:00:06:06, IP len: 0, IP: NULL, label1: 8388496
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:7
        bgp-tunnel-encap:MPLS
        mac-mobility:Seq:2
"
```

On PE-2, the following EVPN entry for MAC 00:00:00:00:06:06 is added to the FDB:

```
*A:PE-2# show service id 7 fdb mac 00:00:00:00:06:06

===============================================================================
Forwarding Database, Service 7
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
7          00:00:00:00:06:06 mpls-1:                 Evpn     01/28/22 13:29:59
                             192.0.2.3:524281
           ldp:65538
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The route table on PE-3 shows an ARP-ND host route for prefix 2001:db8::10:66/128:

```
*A:PE-3# show router 6 route-table 2001:db8::10:66

===============================================================================
IPv6 Route Table (Service: 6)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
2001:db8::10:66/128                           Remote  ARP-ND    00h01m38s  1
       2001:db8::10:66                                              0
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
```

```
===============================================================================
```

PE-2 receives the following RT5 route from PE-3 for prefix 2001:db8::10:66/128:

```
95 2022/01/28 13:30:00.438 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 106
    Flag: 0x90 Type: 14 Len: 69 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-IP-PREFIX Len: 58 RD: 192.0.2.3:5, tag: 0,
                  ip_prefix: 2001:db8::10:66/128 gw_ip fe80::7:b0d1:3fa3:2f60
                  Label: 8388512 (Raw Label: 0x7fffa0)
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:5
        bgp-tunnel-encap:MPLS
"
```

In the route table on PE-2, the route for prefix 2001:db8::10:66/128 is an EVPN route:

```
*A:PE-2# show router 6 route-table 2001:db8::10:66

===============================================================================
IPv6 Route Table (Service: 6)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
2001:db8::10:66/128                           Remote  EVPN-IFF  00h01m37s  169
       fe80::7:b0d1:3fa3:2f60-"evi-5"                               0
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

IPv6 host mobility case 3: host does not send any traffic after moving

The service configuration is the same as use cases 1 and 2. SDP 46 is enabled on PE-4, so host-66 moves back to PE-2. The following traceroute shows that the forwarding path from host-66 to host-61 is via PE-2:

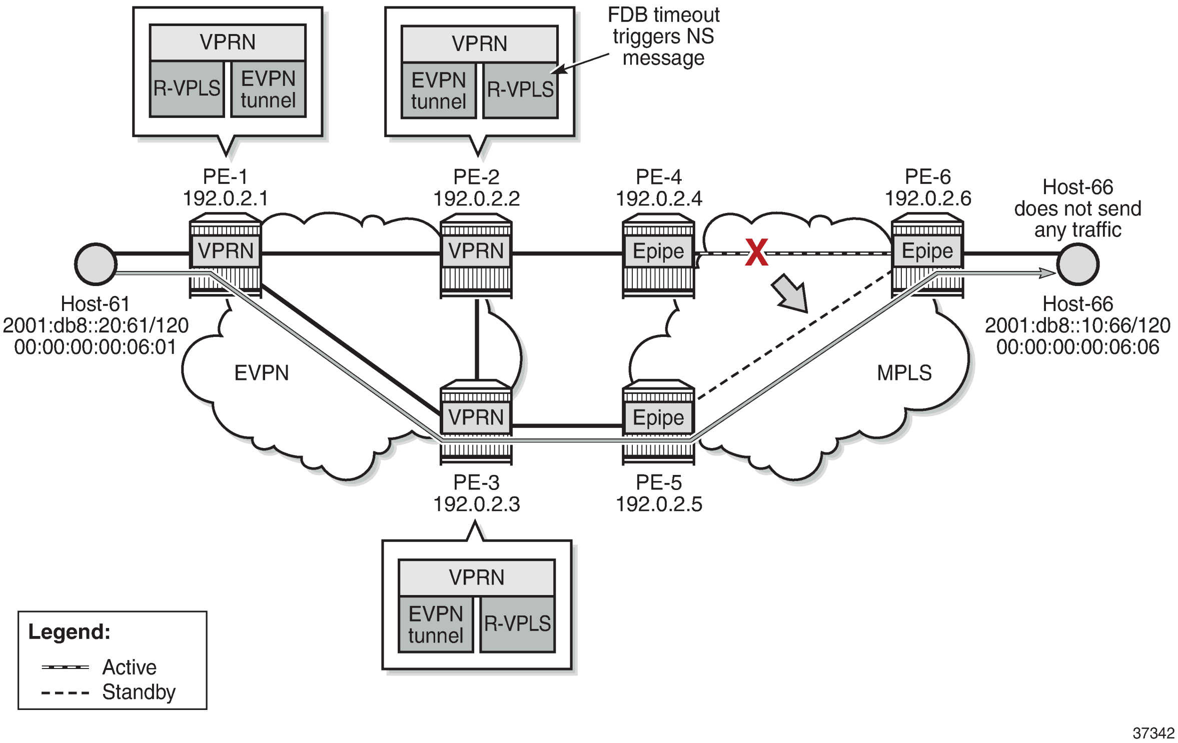
```
*A:PE-6# traceroute router 8 2001:db8::20:61 source 2001:db8::10:66
traceroute to 2001:db8::20:61 from 2001:db8::10:66, 30 hops max, 60 byte packets
  1  2001:db8::10:2 (2001:db8::10:2)    3.76 ms  4.23 ms  4.00 ms
  2  ::  * * *
  3  2001:db8::20:61 (2001:db8::20:61)    5.89 ms  5.40 ms  5.61 ms
```

The FDB on PE-2 shows that MAC address 00:00:00:00:06:06 is learned on the local SAP 1/1/1:7, as follows:

```
*A:PE-2# show service id 7 fdb detail

===============================================================================
Forwarding Database, Service 7
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
7          00:00:00:00:06:06 sap:1/1/1:7             L/30     01/28/22 13:36:03
7          00:00:00:00:2f:07 cpm                     Intf     01/28/22 13:17:40
7          00:00:00:00:3f:07 mpls-1:                 EvpnS:P  01/28/22 13:17:47
                             192.0.2.3:524281
           ldp:65538
-------------------------------------------------------------------------------
No. of MAC Entries: 3
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

A failure is simulated, causing a failover from the primary path via PE-4 to the secondary path via PE-5. Host-66 does not send any traffic after switchover. Figure 194: Host does not send any traffic after switchover shows that PE-2 sends an NS message when the FDB entry for host-66 ages out.

Figure 194: Host does not send any traffic after switchover



On PE-2, MAC address 00:00:00:00:06:06 expires in the FDB for R-VPLS "evi-7", which triggers PE-2 to send an NS message for 2001:db8::10:66. This CPM-generated NS message is flooded to the EVPN destinations PE-1 and PE-3.

```
# on PE-2:
202 2022/01/28 13:37:34.634 UTC MINOR: DEBUG #2001 vprn6 TIP
"TIP: NBR
Sending NS for nbr addr 2001:db8::10:66 nbr type dynamic"
```

The NS message reaches host-66, which replies with an NA message. PE-3 receives the NA message and updates its FDB and ND tables. PE-2 also receives the NA message, but it rejects NA messages received on interface "evi-7" when no learn-dynamic is configured:

```
225 2022/01/28 13:37:34.638 UTC MINOR: DEBUG #2001 vprn6 TIP
"TIP: NBR
Ignore NA for target address 2001:db8::10:66 on evpn endpoint evi-7 because learn-dynamic is
 disabled."
```

PE-2 receives the following EVPN-MAC routes from PE-3:

```
237 2022/01/28 13:43:03.001 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 191
    Flag: 0x90 Type: 14 Len: 146 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-MAC Len: 49 RD: 192.0.2.3:7 ESI: ESI-0, tag: 0, mac len: 48
            mac: 00:00:00:00:06:06, IP len: 16, IP: 2001:db8::10:66, label1: 8388512
        Type: EVPN-MAC Len: 49 RD: 192.0.2.3:7 ESI: ESI-0, tag: 0, mac len: 48
            mac: 00:00:00:00:06:06, IP len: 16, IP: fe80::10:6, label1: 8388512
        Type: EVPN-MAC Len: 33 RD: 192.0.2.3:7 ESI: ESI-0, tag: 0, mac len: 48
            mac: 00:00:00:00:06:06, IP len: 0, IP: NULL, label1: 8388512
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:7
        bgp-tunnel-encap:MPLS
        mac-mobility:Seq:3
"
```

PE-2 receives the following RT5 route from PE-3:

```
228 2022/01/28 13:37:35.638 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 106
    Flag: 0x90 Type: 14 Len: 69 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-IP-PREFIX Len: 58 RD: 192.0.2.3:5, tag: 0,
                  ip_prefix: 2001:db8::10:66/128 gw_ip fe80::7:b0d1:3fa3:2f60
                  Label: 8388512 (Raw Label: 0x7fffa0)
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:5
        bgp-tunnel-encap:MPLS
"
```

Upon receiving the routes, PE-2 updates its FDB and ARP tables. After the switchover, MAC address 00:00:00:00:06:06 is no longer learned on a local SAP on PE-2, but is learned via an EVPN-MAC route from PE-3, as follows:

```
*A:PE-2# show service id 7 fdb detail

===============================================================================
Forwarding Database, Service 7
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
7          00:00:00:00:06:06 mpls-1:                 Evpn     01/28/22 13:37:35
                             192.0.2.3:524281
           ldp:65538
7          00:00:00:00:2f:07 cpm                     Intf     01/28/22 13:17:40
7          00:00:00:00:3f:07 mpls-1:                 EvpnS:P  01/28/22 13:17:47
                             192.0.2.3:524281
           ldp:65538
-------------------------------------------------------------------------------
No. of MAC Entries: 3
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

Conclusion

EVPN host mobility is supported in SR OS as described in draft-ietf-bess-evpn-inter-subnet-forwarding. This chapter describes several cases when a host moves from a source PE to a target PE within the same broadcast domain.

Multi-Chassis Endpoint for VPLS Active/Standby Pseudowire

This chapter provides information about multi-chassis endpoint for VPLS active/standby pseudowire.

Topics in this chapter include:

Applicability

This chapter was initially written for SR OS Release 7.0.R6, but the CLI in this edition is based on SR OS Release 23.7.R2.

Overview

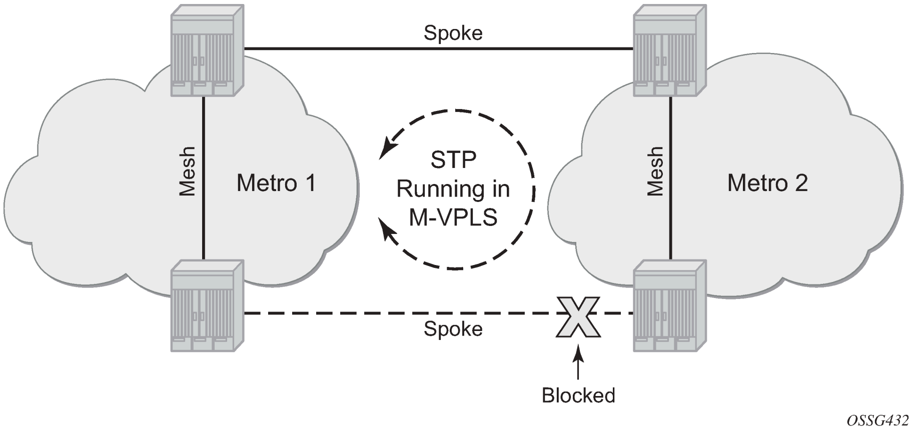
When implementing a large VPLS, one of the limiting factors is the number of T-LDP sessions required for the full mesh of SDPs. Mesh-SDPs are required between all PEs participating in the VPLS with a full mesh of T-LDP sessions.

This solution is not scalable, because the number of sessions grows more rapidly than the number of participating PEs. Several options exist to reduce the number of T-LDP sessions required in a large VPLS.

The first option is hierarchical VPLS (H-VPLS) with spoke SDPs. By using spoke SDPs between two clouds of fully meshed PEs, any-to-any T-LDP sessions for all participating PEs are not required.

However, if spoke SDP redundancy is required, STP must be used to avoid a loop in the VPLS. Management VPLS can be used to reduce the number of STP instances and separate customer and STP traffic, as illustrated in Figure 195: H-VPLS with STP.

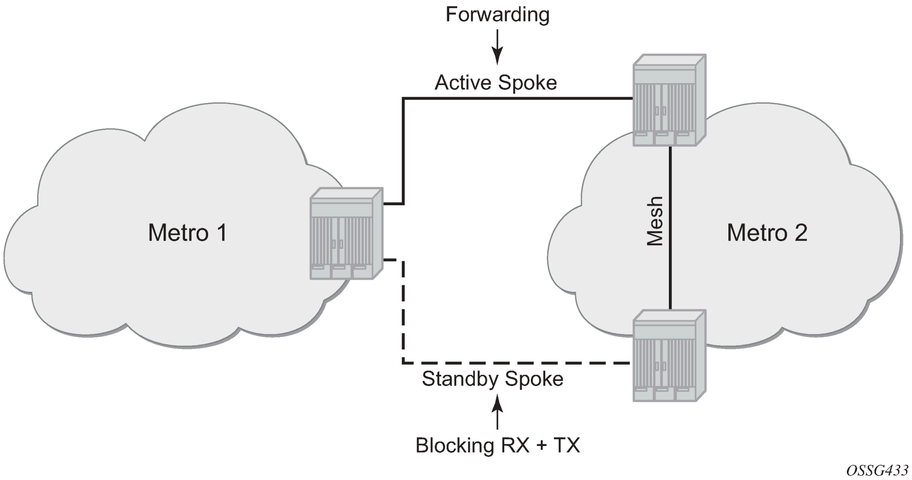
Figure 195: H-VPLS with STP



VPLS pseudowire redundancy provides H-VPLS redundant spoke connectivity. The active spoke SDP is in forwarding state, while the standby spoke SDP is in blocking state. Therefore, STP is not needed anymore to break the loop, as illustrated in Figure 196: VPLS pseudowire redundancy.

However, the PE implementing the active and standby spokes represents a single point of failure in the network.

Figure 196: VPLS pseudowire redundancy

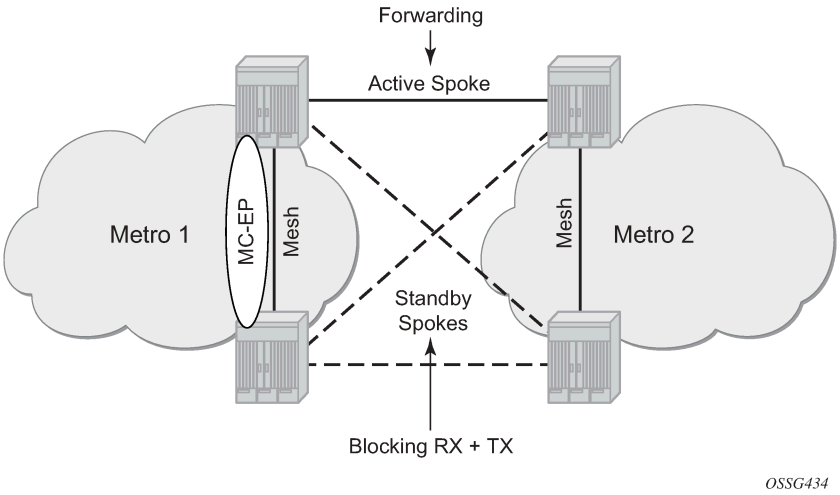


Multi-chassis endpoint (MC-EP) for VPLS active/standby pseudowire expands on the VPLS pseudowire redundancy and allows the removal of the single point of failure.

Only one spoke SDP is in forwarding state; all standby spoke SDPs are in blocking state. Mesh and square resiliency are supported.

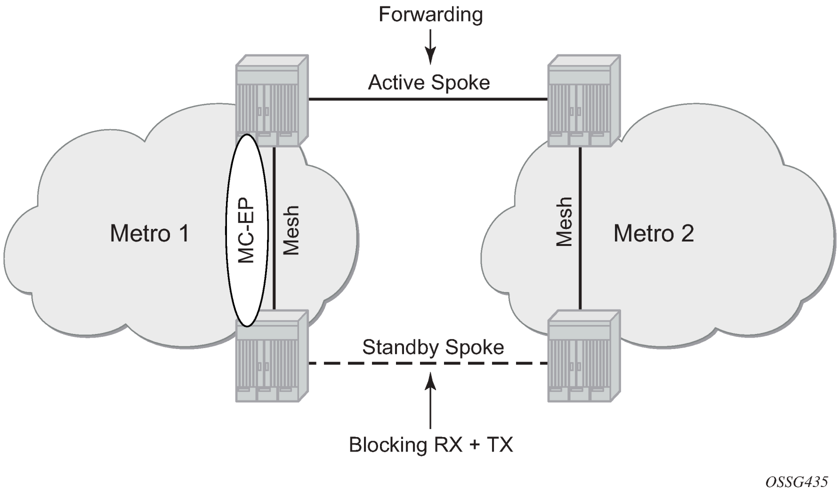
Mesh resiliency can protect against simultaneous node failure in the core and in the MC-EP (double failure), but requires more SDPs (and therefore more T-LDP sessions). Mesh resiliency is illustrated in Figure 197: Multi-chassis endpoint with mesh resiliency.

Figure 197: Multi-chassis endpoint with mesh resiliency



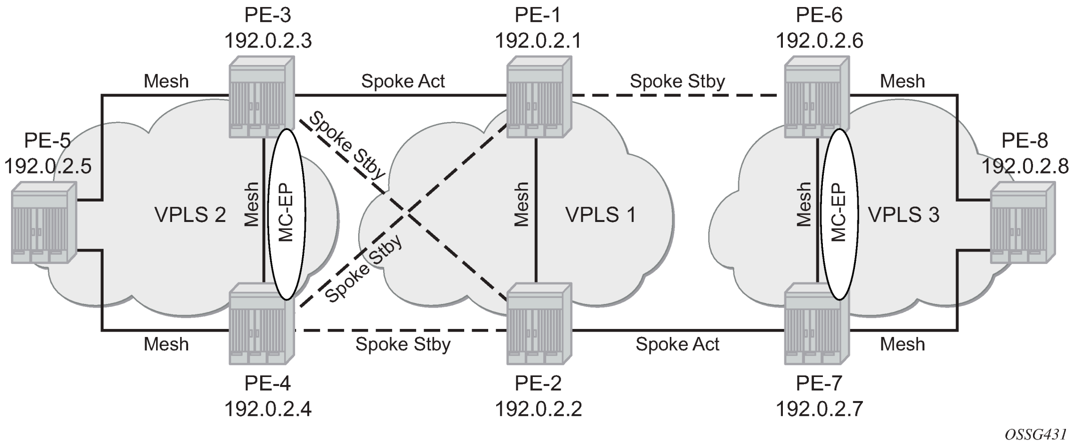
Square resiliency provides single failure node protection, and requires less SDPs (and thus less T-LDP sessions). Square resiliency is illustrated in Figure 198: Multi-chassis endpoint with square resiliency.

Figure 198: Multi-chassis endpoint with square resiliency



Example topology

Figure 199: Example topology



The network topology is displayed in Figure 199: Example topology.

The setup consists of:

- Two core nodes (PE-1 and PE-2), and three nodes for each metro area (PE-3, PE-4, PE-5 and PE-6, PE-7, PE-8, respectively).
- VPLS "Core VPLS-1" is the core VPLS, used to interconnect the two metro areas represented by VPLSs "Metro 1 VPLS-2" and "Metro 2 VPLS-3".
- VPLS "Metro 1 VPLS-2" is connected to the VPLS "Core VPLS-1" in mesh resiliency.
- VPLS "Metro 2 VPLS-3" is connected to the VPLS "Core VPLS-1" in square resiliency.

Three separate VPLS identifiers are used for clarity. However, the same identifier could be used for each. For interoperation, only the same VC-ID is required to be used on both ends of the spoke SDPs.

The initial configuration includes:

- Cards, MDAs, ports, router interfaces
- IS-IS on all router interfaces (alternatively, OSPF can be configured)
- LDP on all router interfaces (alternatively, RSVP-signaled LSPs can be configured over the paths used for mesh/spoke SDPs)

Configuration

SDP configuration

On each PE, SDPs are created to match the topology described in Figure 199: Example topology.

The convention for the SDP naming is: XY where X is the originating node and Y the target node.

The SDP configuration in PE-3 is as follows:

```
# on PE-3:
configure
    service
        sdp 31 mpls create
            far-end 192.0.2.1
            ldp
            no shutdown
        exit
        sdp 32 mpls create
            far-end 192.0.2.2
            ldp
            no shutdown
        exit
        sdp 34 mpls create
            far-end 192.0.2.4
            ldp
            no shutdown
        exit
        sdp 35 mpls create
            far-end 192.0.2.5
            ldp
            no shutdown
        exit
```

The following command shows that the SDPs on PE-3 are operationally up:

```
*A:PE-3# show service sdp

============================================================================
Services: Service Destination Points
============================================================================
SdpId  AdmMTU  OprMTU  Far End          Adm  Opr         Del     LSP   Sig
----------------------------------------------------------------------------
31     0       8914    192.0.2.1        Up   Up          MPLS    L     TLDP
32     0       8914    192.0.2.2        Up   Up          MPLS    L     TLDP
34     0       8914    192.0.2.4        Up   Up          MPLS    L     TLDP
35     0       8914    192.0.2.5        Up   Up          MPLS    L     TLDP
----------------------------------------------------------------------------
Number of SDPs : 4
----------------------------------------------------------------------------
Legend: R = RSVP, L = LDP, B = BGP, M = MPLS-TP, n/a = Not Applicable
        I = SR-ISIS, O = SR-OSPF, T = SR-TE, F = FPE
============================================================================
```

Full mesh VPLS configuration

Three fully meshed VPLS services are configured:

- VPLS "Core VPLS-1" on PE-1 and PE-2
- VPLS "Metro 1 VPLS-2" on PE-3, PE-4, and PE-5
- VPLS "Metro 2 VPLS-3" on PE-6, PE-7, and PE-8

VPLS "Core VPLS-1" is configured on PE-1 as follows. The configuration on PE-2 is similar.

```
# on PE-1:
configure
    service
```

```
        vpls 1 name "Core VPLS-1" customer 1 create
            description "core VPLS"
            mesh-sdp 12:1 create
            exit
            no shutdown
        exit
```

VPLS "Metro 1 VPLS-2" is configured on PE-3 as follows. The configuration on PE-4 and PE-5 is similar.

```
# on PE-3:
configure
    service
        vpls 2 name "Metro 1 VPLS-2" customer 1 create
            description "Metro 1 VPLS"
            mesh-sdp 34:2 create
            exit
            mesh-sdp 35:2 create
            exit
            no shutdown
        exit
```

VPLS "Metro 2 VPLS-3" is configured on PE-6 as follows. The configuration on PE-7 and PE-8 is similar.

```
# on PE-6:
configure
    service
        vpls 3 name "Metro 2 VPLS-3" customer 1 create
            description "Metro 2 VPLS"
            mesh-sdp 67:3 create
            exit
            mesh-sdp 68:3 create
            exit
            no shutdown
        exit
```

Verification of the VPLS:

- The service must be operationally up.
- All mesh SDPs must be up in the VPLS service.

On PE-6 (similar on other nodes):

```
*A:PE-6# show service id 3 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 3                   Vpn Id            : 0
Service Type      : VPLS
MACSec enabled    : no
Name              : Metro 2 VPLS-3
Description       : Metro 2 VPLS
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 10/03/2023 06:11:15
Last Mgmt Change  : 10/03/2023 06:11:09
Etree Mode        : Disabled
Admin State       : Up                  Oper State        : Up
MTU               : 1514
SAP Count         : 0                   SDP Bind Count    : 2
Snd Flush on Fail : Disabled            Host Conn Verify  : Disabled
SHCV pol IPv4     : None
```

```
Propagate MacFlush: Disabled            Per Svc Hashing   : Disabled
Allow IP Intf Bind: Disabled
Fwd-IPv4-Mcast-To*: Disabled            Fwd-IPv6-Mcast-To*: Disabled
Mcast IPv6 scope  : mac-based
Def. Gateway IP   : None
Def. Gateway MAC  : None
Temp Flood Time   : Disabled            Temp Flood        : Inactive
Temp Flood Chg Cnt: 0
SPI load-balance  : Disabled
TEID load-balance : Disabled
Lbl Eth/IP L4 TEID: Disabled
Src Tep IP        : N/A
Vxlan ECMP        : Disabled
MPLS ECMP         : Disabled
Ignore MTU Mismat*: Disabled
Tunnel ELMI       : Disabled

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sdp:67:3 M(192.0.2.7)                    Mesh         0       8914    Up   Up
sdp:68:3 M(192.0.2.8)                    Mesh         0       8914    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

Multi-chassis configuration

Multi-chassis is configured on the MC peers PE-3, PE-4 and PE-6, PE-7. The peer system address is configured, and **mc-endpoint** is enabled.

The multi-chassis configuration on PE-3 is as follows. The configuration on PE-4, PE-6, and PE-7 is similar.

```
# on PE-3:
configure
    redundancy
        multi-chassis
            peer 192.0.2.4 create
                mc-endpoint
                    no shutdown
                exit
                no shutdown
            exit
```

The multi-chassis synchronization (MCS) can be verified with the following command:

```
*A:PE-3# show redundancy multi-chassis mc-endpoint peer 192.0.2.4

===============================================================================
Multi-Chassis MC-Endpoint
===============================================================================
Peer Addr       : 192.0.2.4           Peer Name            :
Admin State     : up                  Oper State           : up
Last State chg  :                     Source Addr          :
System Id       : 02:11:ff:00:00:00   Sys Priority         : 0
Keep Alive Intvl: 10                  Hold on Nbr Fail     : 3
Passive Mode    : disabled            Psv Mode Oper        : No
Boot Timer      : 300                 BFD                  : disabled
```

```
Last update     : 10/03/2023 06:12:34 MC-EP Count          : 0
===============================================================================
```

If the MCS fails, both nodes will fall back to single-chassis mode. In that case, two spoke SDPs could become active at the same time. It is important to verify the MCS before enabling the redundant spoke SDPs.

Mesh resiliency configuration

PE-3 and PE-4 are connected to the core VPLS in mesh resiliency.

- First an endpoint is configured.
- The **no suppress-standby-signaling** is needed to block the standby spoke SDP.
- The multi-chassis endpoint peer is configured. The multi-chassis endpoint ID must match between the two peers.

The configuration on PE-3 and PE-4 is similar, but with a different multi-chassis endpoint peer.

```
# on PE-3:
configure
    service
        vpls "Metro 1 VPLS-2"
            endpoint "CORE" create
            no suppress-standby-signaling
            mc-endpoint 1
                mc-ep-peer 192.0.2.4
            exit
```

After this configuration, the MP-EP count in the preceding show command changes to 1, as follows:

```
*A:PE-3# show redundancy multi-chassis mc-endpoint peer 192.0.2.4

===============================================================================
Multi-Chassis MC-Endpoint
===============================================================================
Peer Addr       : 192.0.2.4          Peer Name            :
Admin State     : up                 Oper State           : up
Last State chg  :                    Source Addr          :
System Id       : 02:11:ff:00:00:00  Sys Priority         : 0
Keep Alive Intvl: 10                 Hold on Nbr Fail     : 3
Passive Mode    : disabled           Psv Mode Oper        : No
Boot Timer      : 300                BFD                  : disabled
Last update     : 10/03/2023 06:12:34 MC-EP Count          : 1
===============================================================================
```

Two spoke SDPs are configured on each peer of the multi-chassis to the two nodes of the core VPLS (mesh resiliency). Each spoke SDP refers to the endpoint CORE.

The precedence is defined on the spoke SDPs as follows:

- Spoke-SDP 31:1 on PE-3 is configured as primary (= precedence 0) and will be active.
- Spoke-SDP 32:1 on PE-3 is configured with precedence 1 and will be the first backup.
- Spoke-SDP 41:1 on PE-4 is configured with precedence 2 and will be the second backup.
- Spoke-SDP 42:1 on PE-4 is configured with precedence 3 and will be the third backup.

The following spoke SDPs are configured in VPLS "Metro 1 VPLS-2" on PE-3:

```
# on PE-3:
configure
    service
        vpls "Metro 1 VPLS-2"
            spoke-sdp 31:1 endpoint "CORE" create
                precedence primary
            exit
            spoke-sdp 32:1 endpoint "CORE" create
                precedence 1
            exit
```

The following spoke SDPs are configured in VPLS "Metro 1 VPLS-2" on PE-4:

```
# on PE-4:
configure
    service
        vpls "Metro 1 VPLS-2"
            spoke-sdp 41:1 endpoint "CORE" create
                precedence 2
            exit
            spoke-sdp 42:1 endpoint "CORE" create
                precedence 3
            exit
```

The following command is used to verify that the spoke and mesh SDPs in VPLS "Metro 1 VPLS-2" on PE-3 are operationally up:

```
*A:PE-3# show service id 2 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
31:1             Spok     192.0.2.1       Up      Up        524277    524278
32:1             Spok     192.0.2.2       Up      Up        524276    524278
34:2             Mesh     192.0.2.4       Up      Up        524279    524279
35:2             Mesh     192.0.2.5       Up      Up        524278    524279
-------------------------------------------------------------------------------
Number of SDPs : 4
-------------------------------------------------------------------------------
===============================================================================
```

The endpoints on PE-3 and PE-4 can be verified. One spoke SDP is in Tx-Active mode (31:1 on PE-1 because it is configured as primary).

```
*A:PE-3# show service id 2 endpoint "CORE" | match "Tx Active"
Tx Active (SDP)              : 31:1
Tx Active Up Time            : 0d 00:00:16
Tx Active Change Count       : 1
Last Tx Active Change        : 10/03/2023 06:16:03
```

There is no active spoke SDP on PE-4.

```
*A:PE-4# show service id 2 endpoint "CORE" | match "Tx Active"
Tx Active                    : none
Tx Active Up Time            : 0d 00:00:00
Tx Active Change Count       : 0
```

```
Last Tx Active Change        : 10/03/2023 05:58:09
```

On PE-1 and PE-2, the spoke SDPs are operationally up.

```
*A:PE-1# show service id 1 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
12:1             Mesh     192.0.2.2       Up      Up        524279    524279
13:1             Spok     192.0.2.3       Up      Up        524278    524277
14:1             Spok     192.0.2.4       Up      Up        524277    524277
-------------------------------------------------------------------------------
Number of SDPs : 3
-------------------------------------------------------------------------------
===============================================================================
```

However, because pseudowire signaling has been enabled, only one spoke SDP will be active, the others are set in standby.

On PE-1, spoke SDP 13:1 is active (no pseudowire bit signaled from peer PE-3) and the spoke SDP 14:1 is signaled in standby by peer PE-4.

```
*A:PE-1# show service id 1 sdp 13:1 detail | match "Peer Pw Bits"
Peer Pw Bits       : None
*A:PE-1# show service id 1 sdp 14:1 detail | match "Peer Pw Bits"
Peer Pw Bits       : pwFwdingStandby
```

On PE-2, both spoke SDPs are signaled in standby by peers PE-3 and PE-4.

```
*A:PE-2# show service id 1 sdp 23:1 detail | match "Peer Pw Bits"
Peer Pw Bits       : pwFwdingStandby
*A:PE-2# show service id 1 sdp 24:1 detail | match "Peer Pw Bits"
Peer Pw Bits       : pwFwdingStandby
```

There is one active and three standby spoke SDPs.

Square resiliency configuration

PE-6 and PE-7 will be connected to the core VPLS in square resiliency.

- First an endpoint is configured.
- The **no suppress-standby-signaling** is needed to block the standby spoke SDP.
- The multi-chassis endpoint peer is configured. The multi-chassis endpoint ID must match between the two peers.

On PE-7 and PE-6, one spoke SDP is configured on each peer of the multi-chassis to one node of the core VPLS (square resiliency). Each spoke SDP refers to the endpoint CORE.

```
# on PE-7:
configure
    service
        vpls "Metro 2 VPLS-3"
            endpoint "CORE" create
                no suppress-standby-signaling
                mc-endpoint 1
                    mc-ep-peer 192.0.2.6
                exit
```

The precedence will be defined on the spoke SDPs as follows:

- Spoke-SDP 72:1 on PE-7 is configured as primary (= precedence 0) and will be active.
- Spoke-SDP 61:1 on PE-6 is configured with precedence 1 and will be the first backup.

On PE-7, spoke SDP 72:1 is configured as primary, as follows:

```
# on PE-7:
configure
    service
        vpls "Metro 2 VPLS-3"
            spoke-sdp 72:1 endpoint "CORE" create
                precedence primary
            exit
```

On PE-6, spoke SDP 61:1 is configured with precedence 1, as follows:

```
# on PE-6:
configure
    service
        vpls "Metro 2 VPLS-3"
            spoke-sdp 61:1 endpoint "CORE" create
                precedence 1
            exit
```

The following command can be used to verify the spoke and mesh SDPs:

```
*A:PE-7# show service id 3 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
72:1             Spok     192.0.2.2       Up      Up        524277    524276
76:3             Mesh     192.0.2.6       Up      Up        524279    524279
78:3             Mesh     192.0.2.8       Up      Up        524278    524278
-------------------------------------------------------------------------------
Number of SDPs : 3
-------------------------------------------------------------------------------
===============================================================================
```

On PE-6 and PE-7, the spoke SDPs must be up.

The endpoints on PE-7 and PE-6 can be verified. Spoke-SDP 72:1 on PE-7 is configured as primary and is in Tx-Active mode.

```
*A:PE-7# show service id 3 endpoint | match "Tx Active"
Tx Active (SDP)              : 72:1
Tx Active Up Time            : 0d 00:00:16
Tx Active Change Count       : 1
Last Tx Active Change        : 10/03/2023 06:19:15
```

There is no active spoke SDP on PE-6.

```
*A:PE-6# show service id 3 endpoint | match "Tx Active"
```

```
Tx Active                       : none
Tx Active Up Time               : 0d 00:00:00
Tx Active Change Count          : 2
Last Tx Active Change           : 10/03/2023 06:19:15
```

The following output on PE-1 shows that spoke SDP 16:1 is signaled with peer in standby mode.

```
*A:PE-1# show service id 1 sdp 16:1 detail | match "Peer Pw Bits"
Peer Pw Bits       : pwFwdingStandby
```

The following output on PE-2 shows that the spoke SDP 27:1 is signaled with peer active (no pseudowire bits).

```
*A:PE-2# show service id 1 sdp 27:1 detail | match "Peer Pw Bits"
Peer Pw Bits       : None
```

There is one active and one standby spoke SDP.

Additional parameters

Multi-chassis

```
*A:PE-3# configure redundancy multi-chassis peer 192.0.2.4 mc-endpoint
  - mc-endpoint
  - no mc-endpoint

 [no] bfd-enable      - Configure BFD
 [no] boot-timer      - Configure boot timer interval
 [no] hold-on-neighb* - Configure hold time applied on neighbor failure
 [no] keep-alive-int* - Configure keep alive interval for this MC-Endpoint
 [no] passive-mode    - Configure passive-mode
 [no] shutdown        - Administratively enable/disable the multi-chassis
                        peer end-point
 [no] system-priority - Configure system priority
```

These parameters will be explained in the following sections.

Peer failure detection

The default mechanism is based on the keep-alive messages exchanged between the peers.

The keep-alive interval is the interval at which keep-alive messages are sent to the MC peer. It is set in tenths of a second from 5 to 500), with a default value of 10.

Hold-on-neighbor failure is the number of keep-alive intervals that the node will wait for a packet from the peer before assuming it has failed. After this interval, the node will revert to single chassis behavior. It can be set from 2 to 25 with a default value of 3.

BFD session

BFD is another peer failure detection mechanism. It can be used to speed up the convergence in case of peer loss.

```
# on PE-3:
configure
    redundancy
        multi-chassis
            peer 192.0.2.4
                mc-endpoint
                    bfd-enable
                exit
            exit
```

BFD must be enabled on the system interface.

```
# on PE-3:
configure
    router Base
        interface "system"
            address 192.0.2.3/32
            bfd 100 receive 100 multiplier 3
        exit
```

Verification of the BFD session:

```
*A:PE-3# show router bfd session

===============================================================================
Legend:
  Session Id = Interface Name | LSP Name | Prefix | RSVP Sess Name | Service Id
  wp = Working path   pp = Protecting path
===============================================================================
BFD Session
===============================================================================
Session Id                                        State      Tx Pkts    Rx Pkts
  Rem Addr/Info/SdpId:VcId                      Multipl     Tx Intvl   Rx Intvl
  Protocols                                        Type     LAG Port     LAG ID
  Loc Addr                                                             LAG name
-------------------------------------------------------------------------------
system                                               Up          N/A        N/A
  192.0.2.4                                           3         1000       1000
  mcep                                           cpm-np          N/A        N/A
  192.0.2.3
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
```

Note:
Simulators are used in the test environment. A limitation of working with simulators is that the minimum BFD transmit or receive interval on simulators equals 1000 ms. Therefore, the timer values in the show command may not reflect the configured timer intervals.

Boot timer

The **boot-timer** command specifies the time after a reboot that the node will try to establish a connection with the MC peer before assuming a peer failure. In case of failure, the node will revert to single chassis behavior.

System priority

The system priority influences the selection of the MC master. The lowest priority node will become the primary.

In case of equal priorities, the lowest system ID (that is, the lowest chassis MAC address) will become the primary.

VPLS endpoint and spoke SDP

Ignore standby pseudowire bits

```
*A:PE-1>config>service>vpls# spoke-sdp 14:1 ?

---snip---
 [no] ignore-standby* - Ignore 'standby-bit' received from LDP peer
---snip---
```

The peer pseudowire status bits are ignored and traffic is forwarded over the spoke SDP, which can speed up convergence for multicast traffic in case of spoke SDP failure. Traffic sent over the standby spoke SDP will be discarded by the peer.

In this topology, if the **ignore-standby-signaling** command is enabled on PE-1, it sends MC traffic to PE-3 and PE-4 (and to PE-6). If PE-3 fails, PE-4 can start forwarding traffic in the VPLS as soon as it detects PE-3 being down. There is no signaling needed between PE-1 and PE-4.

Block-on-mesh failure

```
*A:PE-3>config>service>vpls>endpoint# ?

---snip---
 [no] block-on-mesh-* - Block traffic on mesh-SDP failure
---snip---
```

In case a PE loses all the mesh SDPs of a VPLS, it should block the spoke SDPs to the core VPLS, and inform the MC-EP peer that can activate one of its spoke SDPs.

If **block-on-mesh-failure** is enabled, the PE will signal all the pseudowires of the endpoint in standby.

In this topology, if PE-3 does not have any valid mesh SDP to the VPLS "Metro 1 VPLS-2" mesh, it will set the spoke SDPs under endpoint CORE in standby.

When **block-on-mesh-failure** is activated under an endpoint, it is automatically set under the spoke SDPs belonging to this endpoint.

```
*A:PE-3>config>service>vpls# info
----------------------------------------------
            description "Metro 1 VPLS"
            endpoint "CORE" create
                no suppress-standby-signaling
                mc-endpoint 1
                    mc-ep-peer 192.0.2.4
                exit
            exit
            stp
                shutdown
            exit
            spoke-sdp 31:1 endpoint "CORE" create
                stp
                    shutdown
                exit
                precedence primary
                no shutdown
            exit
            spoke-sdp 32:1 endpoint "CORE" create
                stp
                    shutdown
                exit
                precedence 1
                no shutdown
            exit
            mesh-sdp 34:2 create
                no shutdown
            exit
            mesh-sdp 35:2 create
                no shutdown
            exit
            no shutdown
----------------------------------------------
*A:PE-3>config>service>vpls>endpoint# block-on-mesh-failure
*A:PE-3>config>service>vpls>endpoint# exit
*A:PE-3>config>service>vpls# info
----------------------------------------------
            description "Metro 1 VPLS"
            endpoint "CORE" create
                no suppress-standby-signaling
                block-on-mesh-failure
                mc-endpoint 1
                    mc-ep-peer 192.0.2.4
                exit
            exit
            stp
                shutdown
            exit
            spoke-sdp 31:1 endpoint "CORE" create
                stp
                    shutdown
                exit
                block-on-mesh-failure
                precedence primary
                no shutdown
            exit
            spoke-sdp 32:1 endpoint "CORE" create
                stp
                    shutdown
                exit
                block-on-mesh-failure
                precedence 1
                no shutdown
            exit
            mesh-sdp 34:2 create
                no shutdown
            exit
            mesh-sdp 35:2 create
                no shutdown
            exit
            no shutdown
----------------------------------------------
```

Precedence

```
*A:PE-3>config>service>vpls# spoke-sdp 31:1 ?

---snip---
 [no] precedence      - Configure the spoke-sdp precedence
---snip---
```

The precedence is used to indicate in which order the spoke SDPs should be used. The value is from 0 to 4 (0 being primary), the lowest having higher priority. The default value is 4.

Revert time

```
*A:PE-3>config>service>vpls# endpoint "CORE" ?

---snip---
 [no] revert-time     - Configure the time to wait before reverting to primary spoke-sdp
---snip---
```

If the precedence is equal between the spoke SDPs, there is no revertive behavior. Changing the precedence of a spoke SDP will not trigger a revert. The default is **no revert**.

MAC flush parameters

When a spoke SDP goes from standby to active (due to the active spoke SDP failure), the node will send a flush-all-but-mine message.

After a restoration of the spoke SDP, a new **flush-all-but-mine** message will be sent.

```
# on PE-1:
configure
    service
        vpls "Core VPLS-1"
            propagate-mac-flush
```

A node configured with **propagate-mac-flush** forwards the flush messages received on the spoke SDP to its other mesh or spoke SDPs.

A node configured with **send-flush-on-failure** sends a flush-all-from-me message when one of its mesh or spoke SDPs goes down.

```
# on PE-1:
configure
    service
        vpls "Core VPLS-1"
            send-flush-on-failure
```

Failure scenarios

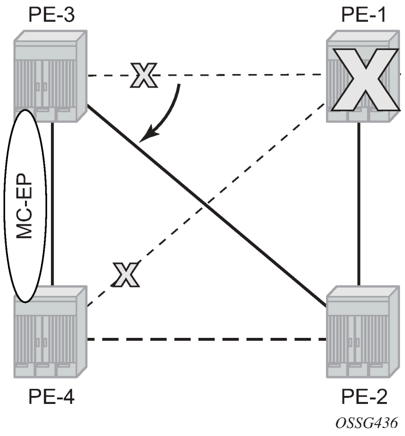
For the subsequent failure scenarios, the configuration of the nodes is as described in the Configuration section.

Core node failure

When the core node PE-1 fails, the spoke SDPs 31:1 from PE-3 and 41:1 from PE-4 go down.

Because the spoke SDP 31:1 was active, the MC master (PE-3 in this case) will select the next best spoke SDP, which will be 32:1 (precedence 1). See Figure 200: Core node failure.

Figure 200: Core node failure



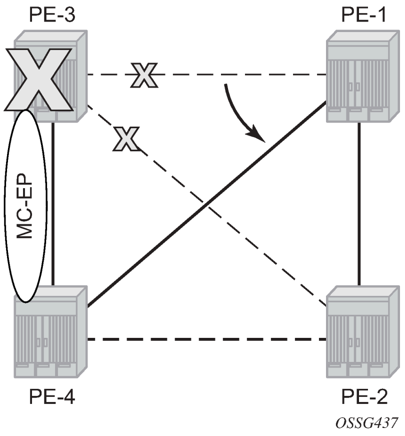
```
*A:PE-3# show service id 2 endpoint

===============================================================================
Service 2 endpoints
===============================================================================
Endpoint name                : CORE
Description                  : (Not Specified)
Creation Origin              : manual
Revert time                  : 0
Act Hold Delay               : 0
Ignore Standby Signaling     : false
Suppress Standby Signaling   : false
Block On Mesh Fail           : true
Multi-Chassis Endpoint       : 1
MC Endpoint Peer Addr        : 192.0.2.4
Psv Mode Active              : No
```

```
Tx Active (SDP)              : 32:1
Tx Active Up Time            : 0d 00:00:11
Revert Time Count Down       : never
Tx Active Change Count       : 2
Last Tx Active Change        : 10/03/2023 06:31:56
-------------------------------------------------------------------------------
Members
-------------------------------------------------------------------------------
Spoke-sdp: 31:1 Prec:0                              Oper Status: Down
Spoke-sdp: 32:1 Prec:1                              Oper Status: Up
===============================================================================
===============================================================================
```

```
*A:PE-4# show service id 2 endpoint

===============================================================================
Service 2 endpoints
===============================================================================
Endpoint name                : CORE
Description                  : (Not Specified)
Creation Origin              : manual
Revert time                  : 0
Act Hold Delay               : 0
Ignore Standby Signaling     : false
Suppress Standby Signaling   : false
Block On Mesh Fail           : false
Multi-Chassis Endpoint       : 1
MC Endpoint Peer Addr        : 192.0.2.3
Psv Mode Active              : No
Tx Active                    : none
Tx Active Up Time            : 0d 00:00:00
Revert Time Count Down       : never
Tx Active Change Count       : 0
Last Tx Active Change        : 10/03/2023 05:58:09
-------------------------------------------------------------------------------
Members
-------------------------------------------------------------------------------
Spoke-sdp: 41:1 Prec:2                              Oper Status: Down
Spoke-sdp: 42:1 Prec:3                              Oper Status: Up
===============================================================================
===============================================================================
```

Multi-chassis node failure

Figure 201: Multi-chassis node failure



When the multi-chassis node PE-3 fails, both spoke SDPs 31:1 and 32:1 from PE-3 go down.

PE-4 reverts to single chassis mode and selects the best spoke SDP, which will be 41:1 between PE-4 and PE-1 (precedence 2). See Figure 201: Multi-chassis node failure.

```
*A:PE-4# show redundancy multi-chassis mc-endpoint peer 192.0.2.3

===============================================================================
Multi-Chassis MC-Endpoint
===============================================================================
Peer Addr       : 192.0.2.3           Peer Name            :
Admin State     : up                  Oper State           : down
Last State chg  :                     Source Addr          :
System Id       : 02:17:ff:00:00:00   Sys Priority         : 0
Keep Alive Intvl: 10                  Hold on Nbr Fail     : 3
Passive Mode    : disabled            Psv Mode Oper        : No
Boot Timer      : 300                 BFD                  : enabled
Last update     : 10/03/2023 06:22:54 MC-EP Count          : 1
===============================================================================
```

```
*A:PE-4# show service id 2 endpoint

===============================================================================
Service 2 endpoints
===============================================================================
Endpoint name                : CORE
Description                  : (Not Specified)
Creation Origin              : manual
Revert time                  : 0
Act Hold Delay               : 0
Ignore Standby Signaling     : false
Suppress Standby Signaling   : false
Block On Mesh Fail           : false
Multi-Chassis Endpoint       : 1
MC Endpoint Peer Addr        : 192.0.2.3
Psv Mode Active              : No
Tx Active (SDP)              : 41:1
Tx Active Up Time            : 0d 00:00:03
Revert Time Count Down       : never
```

```
Tx Active Change Count       : 1
Last Tx Active Change        : 10/03/2023 06:37:38
-------------------------------------------------------------------------------
Members
-------------------------------------------------------------------------------
Spoke-sdp: 41:1 Prec:2                            Oper Status: Up
Spoke-sdp: 42:1 Prec:3                            Oper Status: Up
===============================================================================
```

Multi-chassis communication failure

If the multi-chassis communication is interrupted, both nodes will revert to single chassis mode.

To simulate a communication failure between the two nodes, define a static route on PE-3 that will blackhole the system address of PE-4.

```
# on PE-3:
configure
    router Base
        static-route-entry 192.0.2.4/32
            black-hole
                no shutdown
            exit
        exit
```

Verify that the MC synchronization is operationally down.

```
*A:PE-4# show redundancy multi-chassis mc-endpoint peer 192.0.2.3

===============================================================================
Multi-Chassis MC-Endpoint
===============================================================================
Peer Addr       : 192.0.2.3           Peer Name            :
Admin State     : up                  Oper State           : down
Last State chg  :                     Source Addr          :
System Id       : 02:17:ff:00:00:00   Sys Priority         : 0
Keep Alive Intvl: 10                  Hold on Nbr Fail     : 3
Passive Mode    : disabled            Psv Mode Oper        : No
Boot Timer      : 300                 BFD                  : enabled
Last update     : 10/03/2023 06:22:54 MC-EP Count          : 1
===============================================================================
```

The spoke SDPs are active on PE-3 and on PE-4.

```
*A:PE-3# show service id 2 endpoint | match "Tx Active"
Tx Active (SDP)              : 31:1
Tx Active Up Time            : 0d 00:01:27
Tx Active Change Count       : 10
Last Tx Active Change        : 10/03/2023 06:43:02
```

```
*A:PE-4# show service id 2 endpoint | match "Tx Active"
Tx Active (SDP)              : 41:1
Tx Active Up Time            : 0d 00:00:34
Tx Active Change Count       : 5
Last Tx Active Change        : 10/03/2023 06:43:53
```

This can potentially cause a loop in the system. The Passive mode subsection describes how to avoid this loop.

Passive mode

As in the preceding Multi-chassis communication failure subsection, if there is a failure in the multi-chassis communication, both nodes will assume that the peer is down and will revert to single-chassis mode. This can create loops because two spoke SDPs can become active.

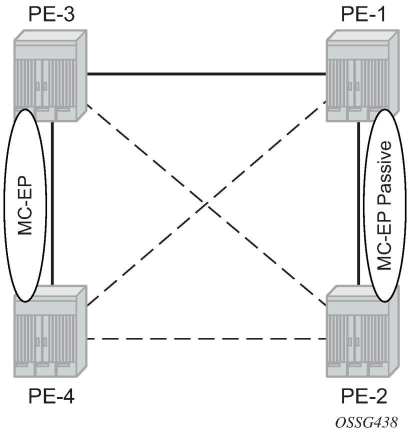
One solution is to synchronize the two core nodes, and configure them in passive mode, as illustrated in Figure 202: Multi-chassis passive mode.

In passive mode, both peers will stay dormant as long as one active spoke SDP is signaled from the remote end. If more than one spoke SDP becomes active, the MC-EP algorithm will select the best SDP. All other spoke SDPs are blocked locally (in Rx and Tx directions). There is no signaling sent to the remote PEs.

If one peer is configured in passive mode, the other peer will be forced to passive mode as well.

The **no suppress-standby-signaling** and **no ignore-standby-signaling** commands are required.

Figure 202: Multi-chassis passive mode



The following output shows the multi-chassis configuration on PE-1 (similar on PE-2).

```
# on PE-1:
configure
    redundancy
        multi-chassis
            peer 192.0.2.2 create
                mc-endpoint
                    no shutdown
                    passive-mode
                exit
                no shutdown
            exit
```

The following output shows the VPLS spoke SDPs configuration on PE-1 (similar on PE-2)

```
# on PE-1:
configure
```

```
    service
        vpls "Core VPLS-1"
            endpoint "METRO1" create
                no suppress-standby-signaling
                mc-endpoint 1
                    mc-ep-peer 192.0.2.2
                exit
            exit
            spoke-sdp 13:1 endpoint "METRO1" create
            exit
            spoke-sdp 14:1 endpoint "METRO1" create
            exit
            no shutdown
```

To simulate a communication failure between the two nodes, a static route is defined on PE-3 that will blackhole the system address of PE-4.

```
# on PE-3:
configure
    router Base
        static-route-entry 192.0.2.4/32
            black-hole
                no shutdown
            exit
        exit
```

The spoke SDPs are active on PE-3 and on PE-4.

```
*A:PE-3# show service id 2 endpoint | match "Tx Active"
Tx Active (SDP)             : 31:1
Tx Active Up Time           : 0d 00:00:37
Tx Active Change Count      : 12
Last Tx Active Change       : 10/03/2023 07:12:39
```

```
*A:PE-4# show service id 2 endpoint | match "Tx Active"
Tx Active (SDP)             : 41:1
Tx Active Up Time           : 0d 00:00:24
Tx Active Change Count      : 7
Last Tx Active Change       : 10/03/2023 07:12:54
```

PE-1 and PE-2 have blocked one spoke SDP which avoids a loop in the VPLS.

```
*A:PE-1# show service id 1 endpoint | match "Tx Active"
Tx Active (SDP)             : 13:1
Tx Active Up Time           : 0d 00:01:16
Tx Active Change Count      : 5
Last Tx Active Change       : 10/03/2023 07:12:39
```

```
*A:PE-2# show service id 1 endpoint | match "Tx Active"
Tx Active                   : none
Tx Active Up Time           : 0d 00:00:00
Tx Active Change Count      : 2
Last Tx Active Change       : 10/03/2023 07:12:54
```

The passive nodes do not set the pseudowire status bits; therefore, the nodes PE-3 and PE-4 are not aware that one spoke SDP is blocked.

Conclusion

Multi-chassis endpoint for VPLS active/standby pseudowire allows the building of hierarchical VPLS without single point of failure, and without requiring STP to avoid loops.

Care must be taken to avoid loops. The multi-chassis peer communication is important and should be possible on different interfaces.

Passive mode can be a solution to avoid loops in case of multi-chassis communication failure.

Multi-Instance VPRN with EVPN-IFL Using SRv6 Transport

This chapter provides information about multi-instance VPRN services with EVPN-IFL using SRv6 transport.

Topics in this chapter include:

Applicability

The information and configuration in this chapter are based on SR OS Release 23.10.R2.

Overview

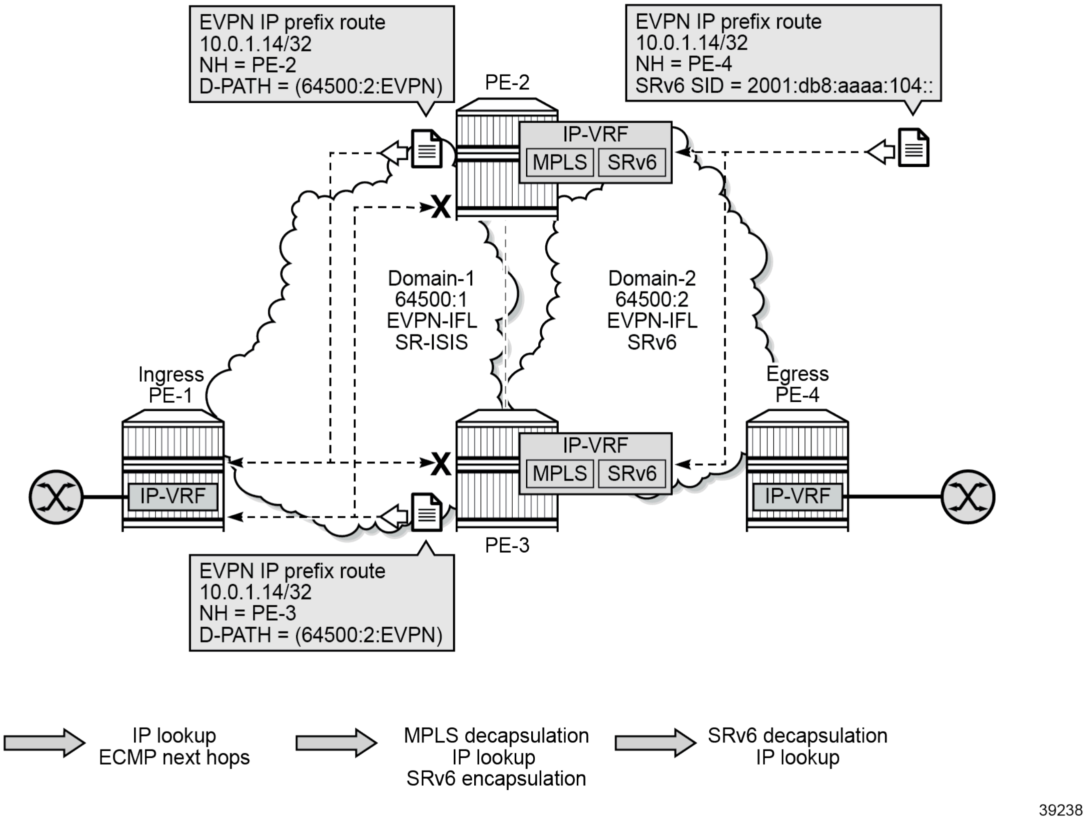
SRv6 transport in VPRN services with EVPN-IFL is supported in SR OS Release 22.5.R1 and later. Maximum two BGP instances per VPRN are supported and these BGP instances can be associated with the same BGP address family or different BGP address families. When configuring a VPRN with EVPN in interface-less mode (EVPN-IFL) over SRv6 transport, the associated SRv6 locator must have the End.DT4, End.DT6, or End.DT46 functions which can be statically configured or dynamically allocated by the router.

BGP path attribute propagation for SRv6 routes does not require a dedicated CLI command. When multiple BGP owners coexist in the same VPRN route table, BGP path propagation is supported in the following cases, regardless of the encapsulation (MPLS or SRv6) of the route:

- between VPN-IPv4/v6 and EVPN-IFL
- between VPN-IPv4/v6 and VPN-IPv4/v6 – when **allow-export-bgp-vpn** is enabled
- between EVPN-IFL and EVPN-IFL – when **allow-export-bgp-vpn** is enabled
- between VPN-IPv4/v6 and IPv4/v6
- between EVPN-IFL and IPv4/v6
- between VPN-IPv4/v6 and EVPN-IFF – when **iff-attribute-uniform-propagation** is enabled
- between EVPN-IFL and EVPN-IFF – when **iff-attribute-uniform-propagation** is enabled

Figure 203: EVPN IP prefix routes readvertised between domains shows how an EVPN IP prefix route originating from PE-4 is advertised for a VPRN with EVPN-IFL configured on all nodes. The VPRN with EVPN-IFL uses SRv6 transport in domain 2 and SR-ISIS tunnels in domain 1.

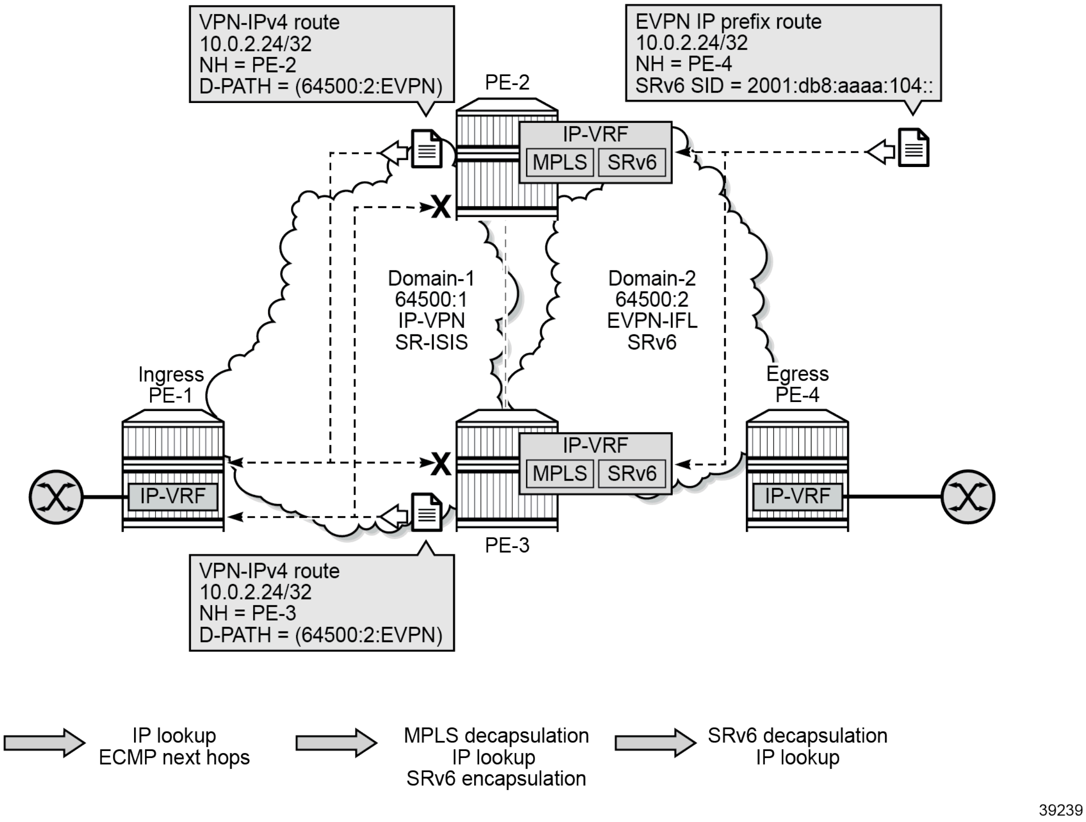
Figure 203: EVPN IP prefix routes readvertised between domains



PE-2 and PE-3 act as service gateways (GWs) that import routes and readvertise them between domains. On the service GWs, the VPRN has two BGP instances that are associated with the EVPN address family. The domain path attribute is used as automated loop prevention, as described in the Domain Path Attribute for VPRN BGP Routes chapter. Each service GW imports the IP prefix route and prepends the domain ID of origin when readvertising these IP prefix routes. When GW PE-2 receives the IP prefix route from PE-4, it prepends domain ID 64500:2 and advertises the IP prefix to PE-1 and PE-3. PE-1 accepts and uses this IP prefix route, but PE-3 does not install this IP prefix route in the VRF because the domain ID 64500:2 is local to PE-3.

Interworking between EVPN-IFL and IP-VPN is supported, as shown in Figure 204: Interworking between EVPN-IFL and IP-VPN.

Figure 204: Interworking between EVPN-IFL and IP-VPN



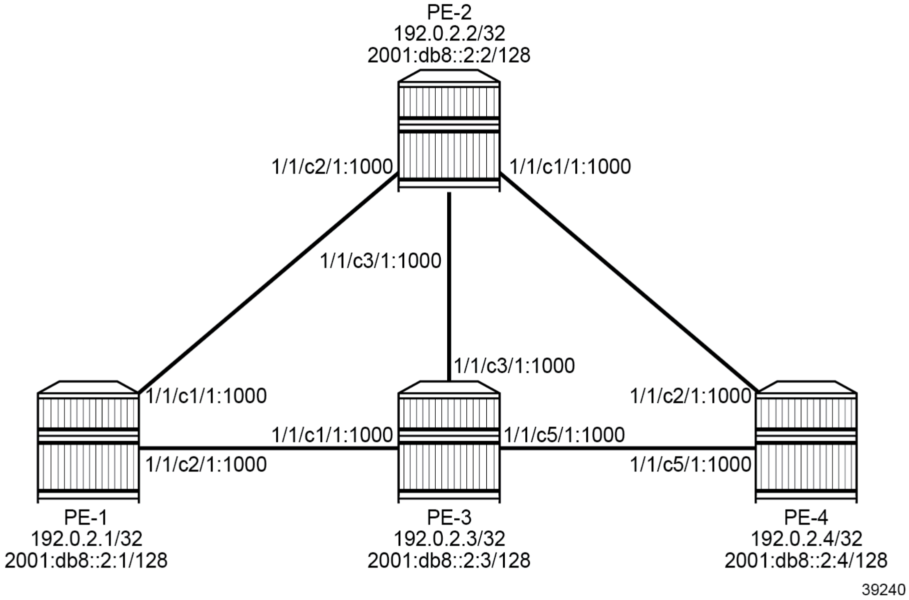
On the service GWs PE-2 and PE-3, one BGP instance is associated with the EVPN address family while the other BGP instance is associated with the VPN-IPv4 address family.

When GW PE-2 receives the IP prefix route from PE-4, it prepends domain ID 64500:2 and advertises the IP prefix 10.0.2.24/32 in a VPN-IPv4 route to PE-1 and PE-3. PE-1 accepts and uses this VPN-IPv4 route, but PE-3 does not install this VPN-IPv4 route in the VRF because the domain ID 64500:2 is local to PE-3.

Configuration

Figure 205: Example topology shows the example topology with four SR OS nodes:

Figure 205: Example topology



The initial configuration on the nodes includes:

- cards, MDAs, ports
- router interfaces
- IS-IS on all router interfaces: IS-IS level 1 between PE-1, PE-2, and PE-3; IS-IS level 2 between PE-2, PE-3, and PE-4
- SR-ISIS between PE-1, PE-2, and PE-3
- SRv6 between PE-2, PE-3, and PE-4

As an example, the initial configuration on PE-2 is as follows:

```
# on PE-2:
configure
    card 1
        mda 1
            xconnect
                mac 1 create
                    loopback 1 create
                    exit
                    loopback 2 create
                    exit
                exit
            exit
            no shutdown
        exit
    exit
    port-xc
        pxc 1 create
            port 1/1/m1/1
            no shutdown
        exit
        pxc 2 create
            port 1/1/m1/2
            no shutdown
        exit
    exit
    port pxc-1.a
        ethernet
        exit
        no shutdown
    exit
    port pxc-1.b
        ethernet
        exit
        no shutdown
    exit
    port pxc-2.a
        ethernet
        exit
        no shutdown
    exit
    port pxc-2.b
        ethernet
        exit
        no shutdown
    exit
    port 1/1/m1/1
        no shutdown
    exit
    port 1/1/m1/2
        no shutdown
    exit
    fwd-path-ext
        fpe 1 create
            path pxc 1
            srv6 origination
                interface-a
                exit
                interface-b
                exit
            exit
        exit
        fpe 2 create
            path pxc 2
            srv6 termination
                interface-a
                exit
                interface-b
                exit
            exit
        exit
    exit
---snip---
    router Base
        interface "int-PE-2-PE-1"
            address 192.168.12.2/30
            port 1/1/c2/1:1000
        exit
        interface "int-PE-2-PE-3"
            address 192.168.23.1/30
            port 1/1/c3/1:1000
            ipv6
                address 2001:db8::168:23:1/126
            exit
        exit
       interface "int-PE-2-PE-4"
            port 1/1/c1/1:1000
            ipv6
                address 2001:db8::168:24:1/126
            exit
        exit
        interface "system"
            address 192.0.2.2/32
            ipv6
                address 2001:db8::2:2/128
            exit
        exit
        segment-routing
            segment-routing-v6
                origination-fpe 1
                source-address 2001:db8::2:2
                locator "PE2-loc"
                    block-length 48
                    termination-fpe 2
                    prefix
                        ip-prefix 2001:db8:aaaa:102::/64
                    exit
                    static-function
                    exit
                    no shutdown
                exit
                base-routing-instance
                    locator "PE2-loc"
                        function
                            end-x-auto-allocate srh-mode psp protection unprotected
                            end 1
                                srh-mode usp
                            exit
                        exit
                    exit
                exit
            exit
        exit
        mpls-labels
            sr-labels start 20000 end 20099
        exit
        isis 0
            area-id 49.0001
            advertise-router-capability as
            traffic-engineering
            traffic-engineering-options
                ipv6
                application-link-attributes
                exit
            exit
            advertise-passive-only
            ipv6-routing native
            level 1
                wide-metrics-only
            exit
            level 2
                wide-metrics-only
            exit
            segment-routing
                prefix-sid-range global
                no shutdown
            exit
            interface "system"
                ipv4-node-sid index 2
                passive
                no shutdown
            exit
            segment-routing-v6
                locator "PE2-loc"
                    level-capability level-2
                    level 1
                    exit
                    level 2
                    exit
                exit
                no shutdown
            exit
            interface "int-PE-2-PE-1"
                level-capability level-1
                interface-type point-to-point
            exit
            interface "int-PE-2-PE-3"
                level-capability level-1/2      # default
                interface-type point-to-point
            exit
            interface "int-PE-2-PE-4"
                level-capability level-2
                interface-type point-to-point
            exit
            no shutdown
        exit
```

The following scenarios are described in this section:

- Multi-instance VPRN with one EVPN-IFL domain using SRv6 transport
  - Multi-instance VPRN with EVPN-IFL over SRv6 and EVPN-IFL over SR-ISIS
  - Multi-instance VPRN with EVPN-IFL over SRv6 and VPN-IPv4/v6 over SR-ISIS
- Multi-instance VPRN with two EVPN-IFL domains using SRv6 transport
  - VPRN with two BGP-EVPN instances pointing at the same SRv6 locator
  - VPRN with two BGP-EVPN instances pointing at different SRv6 locators

Multi-instance VPRN with one EVPN-IFL domain using SRv6 transport

The following two scenarios are described in this section:

- Multi-instance VPRN with EVPN-IFL over SRv6 and EVPN-IFL over SR-ISIS where EVPN-IFL is used in both domains and only the transport is different
- Multi-instance VPRN with EVPN-IFL over SRv6 and VPN-IPv4/v6 over SR-ISIS with interworking between EVPN-IFL and VPN-IPv4/v6 and different transport tunnels in both domains

Multi-instance VPRN with EVPN-IFL over SRv6 and EVPN-IFL over SR-ISIS

BGP configuration

BGP is configured on all nodes for the EVPN address family. The configuration on PE-1 is as follows:

```
# on PE-1:
configure
```

```
    router Base
        autonomous-system 64500
        bgp
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "access-mpls"
                family evpn
                peer-as 64500
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
            no shutdown
```

The BGP configuration on the service GW PE-2 has two different groups. The BGP configuration for the "access-mpls" group is similar to the BGP configuration on PE-1, whereas the BGP configuration for the "core-srv6" has IPv6 peers and advertises IPv6 next hops for EVPN routes:

```
# on PE-2:
configure
    router Base
        autonomous-system 64500
        bgp
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "core-srv6"
                family evpn
                peer-as 64500
                advertise-ipv6-next-hops evpn
                neighbor 2001:db8::2:3
                exit
                neighbor 2001:db8::2:4
                exit
            exit
            group "access-mpls"
                family evpn
                peer-as 64500
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.3
                exit
            exit
            no shutdown
```

The BGP configuration on PE-3 is identical, but with different peer addresses.

On PE-4, the BGP configuration is as follows:

```
# on PE-4:
configure
    router Base
        autonomous-system 64500
        bgp
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "core-srv6"
                family evpn
                peer-as 64500
                advertise-ipv6-next-hops evpn
                neighbor 2001:db8::2:2
                exit
                neighbor 2001:db8::2:3
                exit
            exit
            no shutdown
```

Service configuration

VPRN-1 is configured with EVPN-IFL. On PE-1, VPRN-1 has only one BGP instance and MPLS (SR-ISIS) tunnels are used:

```
# on PE-1:
configure
    service
        vprn 1 name "VPRN-1" customer 1 create
            interface "loopback" create
                address 10.0.1.11/32
                ipv6
                    address 2001:db8::1:11/128
                exit
                loopback
            exit
            bgp-evpn
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    route-distinguisher 192.0.2.1:11
                    vrf-target target:64500:11
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

On GW PE-2, the VPRN-1 service is configured as follows. The SRv6 locator from the **router Base segment-routing segment-routing-v6** context is used and the End.DT4, End.DT6, and End.DT46 functions are configured for it. EVPN-IFL is used in domain 1 and in domain 2. The **allow-export-bgp-vpn** command is required between two EVPN-IFL instances. The route distinguishers and the route targets have different values in the different domains. The domain IDs are configured on the service GWs to avoid loops. For SRv6, the IPv6 system address is used as source address.

```
# on PE-2:
configure
    service
        vprn 1 name "VPRN-1" customer 1 create
            segment-routing-v6 1 create
                locator "PE2-loc"
                    function
                        end-dt4
                        end-dt6
                        end-dt46
                    exit
                exit
```

```
           exit
           bgp-evpn
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    domain-id 64500:1
                    route-distinguisher 192.0.2.2:11
                    vrf-target target:64500:11
                    no shutdown
                exit
                segment-routing-v6 bgp 1
                    domain-id 64500:2
                    route-distinguisher 192.0.2.2:12
                    srv6-instance 1 default-locator "PE2-loc"
                    source-address 2001:db8::2:2
                    vrf-target target:64500:12
                    no shutdown
                exit
            exit
            allow-export-bgp-vpn    # required between two EVPN-IFL instances
            no shutdown
        exit
```

The service configuration on PE-3 is similar.

On PE-4, the VPRN-1 service is configured as follows:

```
# on PE-4:
configure
    service
        vprn 1 name "VPRN-1" customer 1 create
            interface "loopback" create
                address 10.0.1.14/32
                ipv6
                    address 2001:db8::1:14/128
                exit
                loopback
            exit
            segment-routing-v6 1 create
                locator "PE4-loc"
                    function
                        end-dt4
                        end-dt6
                        end-dt46
                    exit
                exit
           exit
           bgp-evpn
                segment-routing-v6 bgp 1
                    route-distinguisher 192.0.2.4:12
                    srv6-instance 1 default-locator "PE4-loc"
                    source-address 2001:db8::2:4
                    vrf-target target:64500:12
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

Verification

GW PE-2 accepts and uses the IP prefix route received from PE-4:

```
*A:PE-2# show router bgp routes evpn ip-prefix rd 192.0.2.4:12
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
Flag  Route Dist.         Prefix
      Tag                 Gw Address
                          NextHop
                          Label
                          ESI
-------------------------------------------------------------------------------
u*>i  192.0.2.4:12        10.0.1.14/32
      0                   00:00:00:00:00:00
                          2001:db8::2:4
                          524288
                          ESI-0

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

The details for this IP prefix route include SRv6 information such as the SID, the End.DT4 function and so on:

```
*A:PE-2# show router bgp routes evpn ip-prefix rd 192.0.2.4:12 detail
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
Original Attributes

Network        : n/a
Nexthop        : 2001:db8::2:4
Path Id        : None
From           : 2001:db8::2:4
Res. Nexthop   : fe80::1a:1ff:fe01:b
Local Pref.    : 100                    Interface Name : int-PE-2-PE-4
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64500:12
Cluster        : No Cluster Members
```

```
Originator Id  : None                   Peer Router Id : 192.0.2.4
Origin         : IGP
Flags          : Used Valid Best
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : IP-PREFIX
ESI            : ESI-0
Tag            : 0
Gateway Address: 00:00:00:00:00:00
Prefix         : 10.0.1.14/32
Route Dist.    : 192.0.2.4:12
MPLS Label     : 524288
Route Tag      : 0
Neighbor-AS    : n/a
DB Orig Val    : N/A                    Final Orig Val : N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h03m46s
SRv6 TLV Type  : SRv6 L3 Service TLV (5)
SRv6 SubTLV    : SRv6 SID Information (1)
Sid            : 2001:db8:aaaa:104::
Full Sid       : 2001:db8:aaaa:104:8000::
Behavior       : End.DT4 (19)
SRv6 SubSubTLV : SRv6 SID Structure (1)
Loc-Block-Len  : 48                     Loc-Node-Len   : 16
Func-Len       : 20                     Arg-Len        : 0
Tpose-Len      : 20                     Tpose-offset   : 64
---snip---
```

PE-2 readvertises this IP prefix route to PE-1 and PE-3 after prepending the domain ID 64500:2. PE-1 accepts the route, but PE-3 has domain ID 64500:2 locally, so it does not install the IP prefix in its VRF. The following shows that PE-3 does not use the IP prefix route for prefix 10.0.1.14/32 with RD 192.0.2.2:11 and D-path [64500:2:(evpn)] . PE-3 detects a domain path loop in VRF 1.

```
*A:PE-3# show router bgp routes evpn ip-prefix rd 192.0.2.2:11 detail
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
---snip---

-------------------------------------------------------------------------------
Original Attributes

Network        : n/a
Nexthop        : 192.0.2.2
Path Id        : None
From           : 192.0.2.2
Res. Nexthop   : 192.168.23.1
Local Pref.    : 100                    Interface Name : int-PE-3-PE-2
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64500:11 bgp-tunnel-encap:MPLS
Cluster        : No Cluster Members
```

```
Originator Id  : None                    Peer Router Id : 192.0.2.2
Origin         : IGP
Flags          : Valid Best
Route Source   : Internal
AS-Path        : No As-Path
D-Path         : [64500:2:(evpn)]
EVPN type      : IP-PREFIX
ESI            : ESI-0
Tag            : 0
Gateway Address: 00:00:00:00:00:00
Prefix         : 10.0.1.14/32
Route Dist.    : 192.0.2.2:11
MPLS Label     : LABEL 524280
Route Tag      : 0
Neighbor-AS    : n/a
DB Orig Val    : N/A                     Final Orig Val : N/A
Source Class   : 0                       Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h03m39s
DPath Loop VRFs: 1
---snip---
```

Likewise, when PE-3 receives an IP prefix route for prefix 10.0.1.14/32 from PE-4, it imports the route and it readvertises this IP prefix route to PE-1 and PE-2 after prepending the domain ID 64500:2. PE-1 accepts and uses the route, but PE-2 has domain ID 64500:2 locally, so it does not install the IP prefix route in its VRF. The following shows that PE-2 does not use the IP prefix route for prefix 10.0.1.14/32 with RD 192.0.2.3:11 and D-path [64500:2:(evpn)] . PE-2 detects a domain path loop in VRF 1.

```
*A:PE-2# show router bgp routes evpn ip-prefix rd 192.0.2.3:11 detail
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
---snip---
-------------------------------------------------------------------------------
Original Attributes

Network        : n/a
Nexthop        : 192.0.2.3
Path Id        : None
From           : 192.0.2.3
Res. Nexthop   : 192.168.23.2
Local Pref.    : 100                     Interface Name : int-PE-2-PE-3
Aggregator AS  : None                    Aggregator     : None
Atomic Aggr.   : Not Atomic              MED            : None
AIGP Metric    : None                    IGP Cost       : 10
Connector      : None
Community      : target:64500:11 bgp-tunnel-encap:MPLS
Cluster        : No Cluster Members
Originator Id  : None                    Peer Router Id : 192.0.2.3
Origin         : IGP
Flags          : Valid Best
Route Source   : Internal
AS-Path        : No As-Path
D-Path         : [64500:2:(evpn)]
```

```
EVPN type      : IP-PREFIX
ESI            : ESI-0
Tag            : 0
Gateway Address: 00:00:00:00:00:00
Prefix         : 10.0.1.14/32
Route Dist.    : 192.0.2.3:11
MPLS Label     : LABEL 524280
Route Tag      : 0
Neighbor-AS    : n/a
DB Orig Val    : N/A                    Final Orig Val : N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h01m44s
DPath Loop VRFs: 1      # Domain ID is local --> Domain path loop detected in VRF 1
---snip---
```

Besides IP prefix routes, the GWs also receive IPv6 prefix routes. PE-2 receives the following IPv6 route from PE-4:

```
*A:PE-2# show router bgp routes evpn ipv6-prefix rd 192.0.2.4:12
================================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
================================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

================================================================================
BGP EVPN IPv6-Prefix Routes
================================================================================
Flag  Route Dist.         Prefix
      Tag                 Gw Address
                          NextHop
                          Label
                          ESI
--------------------------------------------------------------------------------
u*>i  192.0.2.4:12        2001:db8::1:14/128
      0                   00:00:00:00:00:00
                          2001:db8::2:4
                          524287
                          ESI-0

--------------------------------------------------------------------------------
Routes : 1
================================================================================
```

The detailed information for this IPv6 prefix route shows that an SRv6 tunnel with End.DT6 function is used:

```
*A:PE-2# show router bgp routes evpn ipv6-prefix rd 192.0.2.4:12 detail
================================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
================================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

================================================================================
BGP EVPN IPv6-Prefix Routes
================================================================================
Original Attributes

Network        : n/a
Nexthop        : 2001:db8::2:4
Path Id        : None
From           : 2001:db8::2:4
Res. Nexthop   : fe80::1a:1ff:fe01:b
Local Pref.    : 100                    Interface Name : int-PE-2-PE-4
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64500:12
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.4
Origin         : IGP
Flags          : Used Valid Best
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : IP-PREFIX
ESI            : ESI-0
Tag            : 0
Gateway Address: 00:00:00:00:00:00
Prefix         : 2001:db8::1:14/128
Route Dist.    : 192.0.2.4:12
MPLS Label     : 524287
Route Tag      : 0
Neighbor-AS    : n/a
DB Orig Val    : N/A                    Final Orig Val : N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h04m25s
SRv6 TLV Type  : SRv6 L3 Service TLV (5)
SRv6 SubTLV    : SRv6 SID Information (1)
Sid            : 2001:db8:aaaa:104::
Full Sid       : 2001:db8:aaaa:104:7fff:f000::
Behavior       : End.DT6 (18)
SRv6 SubSubTLV : SRv6 SID Structure (1)
Loc-Block-Len  : 48                     Loc-Node-Len   : 16
Func-Len       : 20                     Arg-Len        : 0
Tpose-Len      : 20                     Tpose-offset   : 64
---snip---
```

The IPv4 route table for VPRN-1 on PE-1 shows an EVPN-IFL route to 10.0.1.14/32 that uses an SR-ISIS tunnel to PE-2:

```
*A:PE-1# show router service-name "VPRN-1" route-table

===============================================================================
Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.1.11/32                                  Local   Local     00h05m35s  0
       loopback                                                     0
10.0.1.14/32                                  Remote  EVPN-IFL  00h05m02s  170
       192.0.2.2 (tunneled:SR-ISIS:524290)                          10
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The IPv4 route table for VPRN-1 on PE-2 shows an EVPN-IFL route to 10.0.1.11/32 that uses an SR-ISIS tunnel to PE-1 and an EVPN-IFL route to 10.0.1.14/32 that uses an SRv6 tunnel to PE-4:

```
*A:PE-2# show router service-name "VPRN-1" route-table

===============================================================================
Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.1.11/32                                  Remote  EVPN-IFL  00h05m23s  170
       192.0.2.1 (tunneled:SR-ISIS:524290)                          10
10.0.1.14/32                                  Remote  EVPN-IFL  00h05m03s  170
       2001:db8:aaaa:104:8000:: (tunneled:SRV6)                     10
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The IPv4 route table for VPRN-1 on PE-3 is similar:

```
*A:PE-3# show router service-name "VPRN-1" route-table

===============================================================================
Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.1.11/32                                  Remote  EVPN-IFL  00h05m09s  170
       192.0.2.1 (tunneled:SR-ISIS:524292)                          10
10.0.1.14/32                                  Remote  EVPN-IFL  00h05m04s  170
       2001:db8:aaaa:104:8000:: (tunneled:SRV6)                     10
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

On PE-4, the route table for VPRN-1 is as follows:

```
*A:PE-4# show router service-name "VPRN-1" route-table

===============================================================================
Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.1.11/32                                  Remote  EVPN-IFL  00h05m01s  170
       2001:db8:aaaa:102:7fff:c000:: (tunneled:SRV6)                10
10.0.1.14/32                                  Local   Local     00h05m05s  0
       loopback                                                     0
```

```
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The IPv6 route tables for VPRN-1 on the different PEs are similar; for example, on PE-2:

```
*A:PE-2# show router service-name "VPRN-1" route-table ipv6

===============================================================================
IPv6 Route Table (Service: 1)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
2001:db8::1:11/128                            Remote  EVPN-IFL  00h06m24s  170
       192.0.2.1 (tunneled:SR-ISIS:524290)                          10
2001:db8::1:14/128                            Remote  EVPN-IFL  00h06m03s  170
       2001:db8:aaaa:104:7fff:f000:: (tunneled:SRV6)                10
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

EVPN-IFL IPv4 routes are advertised with End.DT4 or End.DT46 in that preference order and EVPN-IFL IPv6 routes are advertised with End.DT6 or End.DT46 in that preference order. The following command shows the SID values for the End.DT4, End.DT6, and End.DT46 functions on PE-4:

```
*A:PE-4# show service id "VPRN-1" segment-routing-v6 instance 1

===============================================================================
Segment Routing v6 Instance 1 Service 1
===============================================================================
Locator
 Type          Function  SID                                   Status
-------------------------------------------------------------------------------
PE4-loc
  End.DT4      *524288   2001:db8:aaaa:104:8000::              ok
  End.DT6      *524287   2001:db8:aaaa:104:7fff:f000::         ok
  End.DT46     *524286   2001:db8:aaaa:104:7fff:e000::         ok
===============================================================================
Legend: * - System allocated
```

The following command displays the configured BGP-EVPN parameters for MPLS and for SRv6:

```
*A:PE-2# show service id "VPRN-1" bgp-evpn

===============================================================================
BGP EVPN MPLS Table
===============================================================================
Admin State        : Up                    Oper State     : Up
VRF Import         : None
VRF Export         : None
Route Dist.        : 192.0.2.2:11
Oper Route Dist.   : 192.0.2.2:11
Oper RD Type       : configured
Route Target       : target:64500:11
Route Target Import: None
Route Target Export: None
Default Route Tag  : None
Domain-Id          : 64500:1
Dyn Egr Lbl Limit  : Disabled
EVI                : 0

Advertise          : Disabled
Weighted ECMP      : Disabled

Auto-Bind Tunnel
Resolution         : any                    Strict Tnl Tag : False
ECMP               : 1                      Flex Algo FB   : False
Bgp Instance       : 1
Filter Tunnel Types: (Not Specified)

Tunnel Encap
MPLS               : True                   MPLSoUDP       : False
===============================================================================

===============================================================================
Service 1 BGP-EVPN Segment-Routing-V6 Information
===============================================================================

Admin State       : Up                  Oper State        : Up
EVI               : <default>
VRF Import        : None
VRF Export        : None
Route Dist.       : 192.0.2.2:12
Oper Route Dist   : 192.0.2.2:12
Oper RD Type      : configured
Route Target      : target:64500:12
Route Target Expor: None
Route Target Impor: None
Def Route Tag     : 0x0
Route Resolution  : route-table

Srv6 Instance     : 1
Default Locator   : PE2-loc
Source Address    : 2001:db8::2:2
Domain-Id         : 64500:2

Advertise         : Disabled
Weighted ECMP     : Disabled

===============================================================================
```

Multi-instance VPRN with EVPN-IFL over SRv6 and VPN-IPv4/v6 over SR-ISIS

This section describes a use case with interworking between EVPN-IFL and VPN-IPv4.

BGP configuration

Between PE-1, PE-2, and PE-3, BGP is supported for the VPN-IPv4 and VPN-IPv6 address families. The BGP configuration on PE-1 is as follows:

```
# on PE-1:
```

```
configure
    router Base
        autonomous-system 64500
        bgp
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "access-mpls"
                family vpn-ipv4 vpn-ipv6
                peer-as 64500
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
```

The BGP configuration on PE-2 is as follows:

```
# on PE-2:
configure
    router Base
        autonomous-system 64500
        bgp
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "access-mpls"
                family vpn-ipv4 vpn-ipv6
                peer-as 64500
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.3
                exit
            exit
            group "core-srv6"
                family evpn
                peer-as 64500
                advertise-ipv6-next-hops evpn
                neighbor 2001:db8::2:3
                exit
                neighbor 2001:db8::2:4
                exit
            exit
        exit
```

The BGP configuration on PE-3 is similar.

The BGP configuration on PE-4 remains unchanged.

Service configuration

On PE-1, VPRN-2 is configured as follows:

```
# on PE-1:
configure
    service
        vprn 2 name "VPRN-2" customer 1 create
            interface "loopback" create
                address 10.0.2.21/32
                ipv6
                    address 2001:db8::2:21/128
                exit
                loopback
            exit
            bgp-ipvpn
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    route-distinguisher 192.0.2.1:21
                    vrf-target target:64500:21
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

On PE-2, VPRN-2 is configured as follows:

```
# on PE-2:
configure
    service
        vprn 2 name "VPRN-2" customer 1 create
            segment-routing-v6 1 create
                locator "PE2-loc"
                    function
                        end-dt4
                        end-dt6
                        end-dt46
                    exit
                exit
           exit
           bgp-ipvpn
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    domain-id 64500:1
                    route-distinguisher 192.0.2.2:21
                    vrf-target target:64500:21
                    no shutdown
                exit
           exit
           bgp-evpn
                segment-routing-v6 bgp 1
                    domain-id 64500:2
                    route-distinguisher 192.0.2.2:22
                    srv6-instance 1 default-locator "PE2-loc"
                    source-address 2001:db8::2:2
                    vrf-target target:64500:22
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

The configuration on PE-3 is similar.

On PE-4, VPRN-2 is configured as follows:

```
# on PE-4:
configure
```

```
    service
        vprn 2 name "VPRN-2" customer 1 create
            interface "loopback" create
                address 10.0.2.24/32
                ipv6
                    address 2001:db8::2:24/128
                exit
                loopback
            exit
            segment-routing-v6 1 create
                locator "PE4-loc"
                    function
                        end-dt4
                        end-dt6
                        end-dt46
                    exit
                exit
           exit
           bgp-evpn
                segment-routing-v6 bgp 1
                    route-distinguisher 192.0.2.4:22
                    srv6-instance 1 default-locator "PE4-loc"
                    source-address 2001:db8::2:4
                    vrf-target target:64500:22
                    no shutdown
                exit
            exit
            no shutdown
        exit
        no shutdown
```

Verification

GW PE-2 receives and uses the following IP prefix route from PE-4:

```
*A:PE-2# show router bgp routes evpn ip-prefix rd 192.0.2.4:22
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
Flag  Route Dist.         Prefix
      Tag                 Gw Address
                          NextHop
                          Label
                          ESI
-------------------------------------------------------------------------------
u*>i  192.0.2.4:22        10.0.2.24/32
      0                   00:00:00:00:00:00
                          2001:db8::2:4
                          524285
                          ESI-0

-------------------------------------------------------------------------------
Routes : 1
```

```
===============================================================================
```

The detailed information for this IP prefix route shows that the End.DT4 function is used:

```
*A:PE-2# show router bgp routes evpn ip-prefix rd 192.0.2.4:22 detail
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
Original Attributes

Network        : n/a
Nexthop        : 2001:db8::2:4
Path Id        : None
From           : 2001:db8::2:4
Res. Nexthop   : fe80::1a:1ff:fe01:b
Local Pref.    : 100                    Interface Name : int-PE-2-PE-4
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64500:22
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.4
Origin         : IGP
Flags          : Used Valid Best
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : IP-PREFIX
ESI            : ESI-0
Tag            : 0
Gateway Address: 00:00:00:00:00:00
Prefix         : 10.0.2.24/32
Route Dist.    : 192.0.2.4:22
MPLS Label     : 524285
Route Tag      : 0
Neighbor-AS    : n/a
DB Orig Val    : N/A                    Final Orig Val : N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h03m27s
SRv6 TLV Type  : SRv6 L3 Service TLV (5)
SRv6 SubTLV    : SRv6 SID Information (1)
Sid            : 2001:db8:aaaa:104::
Full Sid       : 2001:db8:aaaa:104:7fff:d000::
Behavior       : End.DT4 (19)
SRv6 SubSubTLV : SRv6 SID Structure (1)
Loc-Block-Len  : 48                     Loc-Node-Len   : 16
Func-Len       : 20                     Arg-Len        : 0
Tpose-Len      : 20                     Tpose-offset   : 64
---snip---
```

PE-2 readvertises this prefix in a VPN-IPv4 route to PE-1 and PE-3 after prepending the domain ID 64500:2. PE-1 accepts this route, but PE-3 has domain ID 64500:2 locally, so it does not add this route to its VRF. The following shows that PE-3 does not use the VPN-IPv4 route received from PE-2 and that PE-3 detects a domain path loop in VRF 2:

```
*A:PE-3# show router bgp routes vpn-ipv4 rd 192.0.2.2:21 detail
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP VPN-IPv4 Routes
===============================================================================
Original Attributes

Network        : 10.0.2.24/32
Nexthop        : 192.0.2.2
Route Dist.    : 192.0.2.2:21           VPN Label      : 524276
Path Id        : None
From           : 192.0.2.2
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : int-PE-3-PE-2
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64500:21
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.2
Fwd Class      : None                   Priority       : None
Origin         : IGP
Flags          : Valid Best
Route Source   : Internal
AS-Path        : No As-Path
D-Path         : [64500:2:(evpn)]
Route Tag      : 0
Neighbor-AS    : n/a
DB Orig Val    : N/A                    Final Orig Val : N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h02m32s
VPRN Imported  : None
DPath Loop VRFs: 2
---snip---
```

The IPv4 route table on PE-1 shows a BGP-VPN route to 10.0.2.24/32 that uses an SR-ISIS tunnel to PE-2:

```
*A:PE-1# show router service-name "VPRN-2" route-table

===============================================================================
Route Table (Service: 2)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.2.21/32                                  Local   Local     00h06m02s  0
       loopback                                                     0
10.0.2.24/32                                  Remote  BGP VPN   00h05m25s  170
       192.0.2.2 (tunneled:SR-ISIS:524290)                          10
```

```
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The IPv4 route table on PE-2 shows a BGP-VPN route to 10.0.2.21/32 that uses an SR-ISIS tunnel to PE-1 and an EVPN-IFL route to 10.0.2.24/32 that uses an SRv6 tunnel to PE-4:

```
*A:PE-2# show router service-name "VPRN-2" route-table

===============================================================================
Route Table (Service: 2)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.2.21/32                                  Remote  BGP VPN   00h05m51s  170
       192.0.2.1 (tunneled:SR-ISIS:524290)                           10
10.0.2.24/32                                  Remote  EVPN-IFL  00h05m42s  170
       2001:db8:aaaa:104:7fff:d000:: (tunneled:SRV6)                 10
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The route table on PE-3 is similar.

The route table on PE-4 is as follows:

```
*A:PE-4# show router service-name "VPRN-2" route-table

===============================================================================
Route Table (Service: 2)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.2.21/32                                  Remote  EVPN-IFL  00h06m18s  170
       2001:db8:aaaa:102:7fff:8000:: (tunneled:SRV6)                 10
10.0.2.24/32                                  Local   Local     00h06m22s  0
       loopback                                                     0
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The IPv6 route tables for VPRN-2 are similar; for example, on PE-2:

```
*A:PE-2# show router service-name "VPRN-2" route-table ipv6

===============================================================================
IPv6 Route Table (Service: 2)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
2001:db8::2:21/128                            Remote  BGP VPN   00h06m48s  170
       192.0.2.1 (tunneled:SR-ISIS:524290)                          10
2001:db8::2:24/128                            Remote  EVPN-IFL  00h06m38s  170
       2001:db8:aaaa:104:7fff:c000:: (tunneled:SRV6)                10
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The next hop value 2001:db8:aaaa:104:7fff:c000:: in the preceding output corresponds to the SID value for End.DT6 in the following command on PE-4:

```
*A:PE-4# show service id "VPRN-2" segment-routing-v6 instance 1

===============================================================================
Segment Routing v6 Instance 1 Service 2
===============================================================================
Locator
 Type          Function  SID                                     Status
-------------------------------------------------------------------------------
PE4-loc
  End.DT4      *524285   2001:db8:aaaa:104:7fff:d000::           ok
  End.DT6      *524284   2001:db8:aaaa:104:7fff:c000::           ok
  End.DT46     *524283   2001:db8:aaaa:104:7fff:b000::           ok
===============================================================================
Legend: * - System allocated
```

The following command shows the BGP-IPVPN information for VPRN-2 on PE-2:

```
*A:PE-2# show service id "VPRN-2" bgp-ipvpn

===============================================================================
Service 2 BGP-IPVPN MPLS Information
===============================================================================
Admin State       : Up                  Oper State        : Up
VRF Import        : None
VRF Export        : None
Route Dist.       : None
Oper Route Dist   : 192.0.2.2:21
Oper RD Type      : configured
Route Target      : target:64500:21
Route Target Impor: None
Route Target Expor: None
Domain-Id         : 64500:1
Dyn Egr Lbl Limit : Disabled

Auto-Bind Tunnel
Resolution        : any                 Strict Tnl Tag    : False
ECMP              : 1                   Flex Algo FB      : False
Weighted ECMP     : False
BGP Instance      : 1
Filter Tunnel Type: bgp
===============================================================================
```

The following command shows the BGP-EVPN information for VPRN-2 on PE-2:

```
*A:PE-2# show service id "VPRN-2" bgp-evpn

===============================================================================
Service 2 BGP-EVPN Segment-Routing-V6 Information
===============================================================================

Admin State       : Up                  Oper State        : Up
EVI               : <default>
VRF Import        : None
VRF Export        : None
Route Dist.       : 192.0.2.2:22
Oper Route Dist   : 192.0.2.2:22
Oper RD Type      : configured
Route Target      : target:64500:22
Route Target Expor: None
Route Target Impor: None
Def Route Tag     : 0x0
Route Resolution  : route-table

Srv6 Instance     : 1
Default Locator   : PE2-loc
Source Address    : 2001:db8::2:2
Domain-Id         : 64500:2

Advertise         : Disabled
Weighted ECMP     : Disabled

===============================================================================
```

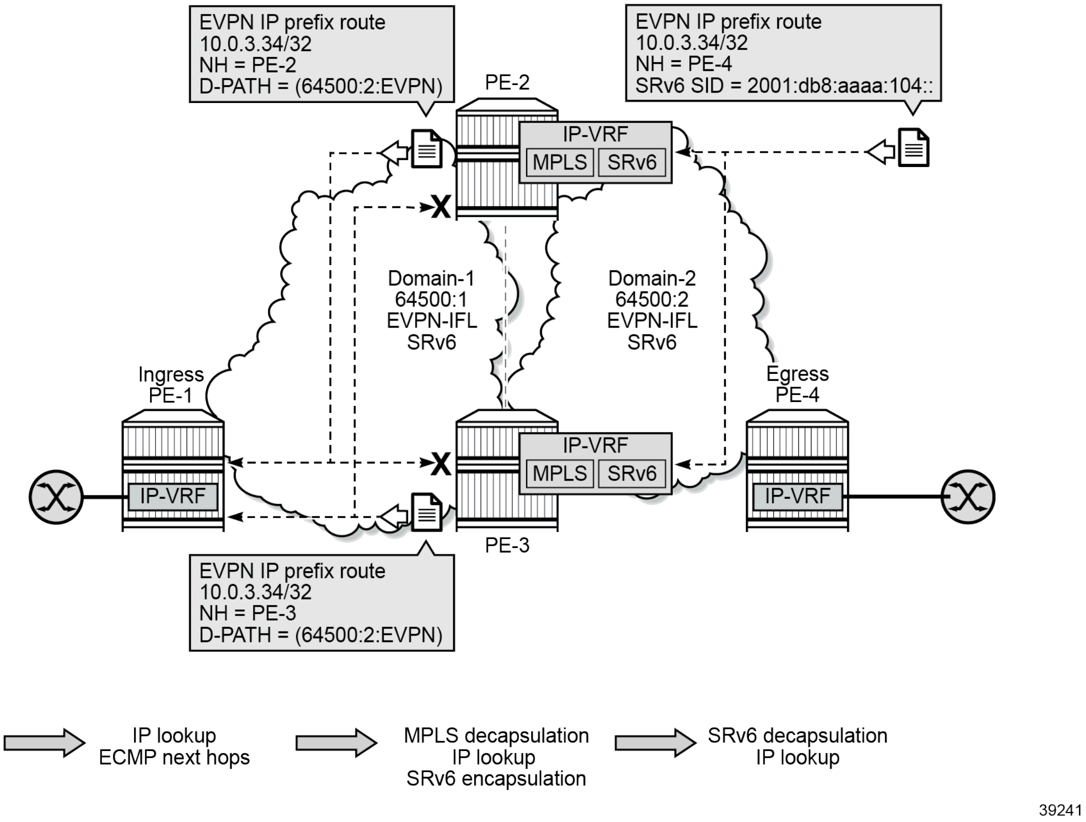
Multi-instance VPRN with two EVPN-IFL domains using SRv6 transport

Multi-instance VPRNs with two EVPN-IFL domains using SRv6 transport are supported in SR OS Release 23.10.R1 and later. In this section, the following two scenarios are described:

- VPRN with two BGP-EVPN instances pointing at the same SRv6 locator
- VPRN with two BGP-EVPN instances pointing at different SRv6 locators

Figure 206: EVPN IP prefix routes readvertised between SRv6 domains shows how IP prefix 10.0.3.34/32 is advertised in VPRN-3 with two BGP-EVPN instances pointing at the same SRv6 locator.

Figure 206: EVPN IP prefix routes readvertised between SRv6 domains



For a VPRN with two BGP-EVPN instances pointing at different SRv6 locators, the behavior is identical but the SRv6 SID on PE-4 is different for a different SRv6 locator.

VPRN with two BGP-EVPN instances pointing at the same SRv6 locator

This section describes VPRN-3 which has two BGP-EVPN instances that both use the same locator.

SRv6 configuration

SRv6 was already configured among PE-2, PE-3, and PE-4. In this scenario and the following, SRv6 is also configured among PE-1, PE-2, and PE-3. The IS-IS configuration on PE-1 is as follows:

```
# on PE-1:
configure
    router Base
        isis 0
            area-id 49.0001
            traffic-engineering
            traffic-engineering-options
                ipv6
                application-link-attributes
                exit
            exit
            advertise-passive-only
            advertise-router-capability as
            ipv6-routing native
            level 1
                wide-metrics-only
            exit
            level 2
                wide-metrics-only
            exit
            segment-routing-v6
                locator "PE1-loc"
                    level-capability level-1    # on PE-2, PE-3: level-1/2 (default)
                    level 1
                    exit
                    level 2
                    exit
                exit
                no shutdown
            exit
            interface "system"
                passive
                no shutdown
            exit
            interface "int-PE-1-PE-2"
                level-capability level-1
                interface-type point-to-point
                no shutdown
            exit
            interface "int-PE-1-PE-3"
                level-capability level-1
                interface-type point-to-point
                no shutdown
            exit
            no shutdown
```

On the GWs PE-2 and PE-3, the existing locators "PE2-loc" and "PE3-loc" are used on both SRv6 domains and these SRv6 locators are configured with level-capability level-1/2, which is the default value.

BGP configuration

In this example, BGP uses IPv6 peer addresses. The BGP configuration on PE-1 is as follows:

```
# on PE-1:
configure
    router Base
        autonomous-system 64500
        bgp
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "access-srv6"
                family evpn
                peer-as 64500
                advertise-ipv6-next-hops evpn
                neighbor 2001:db8::2:2
                exit
                neighbor 2001:db8::2:3
                exit
            exit
            no shutdown
```

The BGP configuration on the service GWs PE-2 and PE-3 is as follows:

```
# on PE-2:
configure
    router Base
        autonomous-system 64500
        bgp
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "core-srv6"
                family evpn
                peer-as 64500
                advertise-ipv6-next-hops evpn
                neighbor 2001:db8::2:3            # on PE-3: 2001:db8::2:2
                exit
                neighbor 2001:db8::2:4
                exit
            exit
            group "access-srv6"
                family evpn
                peer-as 64500
                advertise-ipv6-next-hops evpn
                neighbor 2001:db8::2:1
                exit
            exit
            no shutdown
```

The BGP configuration on PE-4 remains the same as in the preceding use cases.

Service configuration

On PE-1 and PE-4, the configuration of VPRN-3 is similar. VPRN-3 is configured on PE-1 as follows:

```
# on PE-1:
configure
    service
        vprn 3 name "VPRN-3" customer 1 create
            interface "loopback" create
                address 10.0.3.31/32              # on PE-4: 10.0.3.34/32
                ipv6
                    address 2001:db8::3:31/128    # on PE-4: 2001:db8::3:34/128
                exit
                loopback
            exit
            segment-routing-v6 1 create
                locator "PE1-loc"                  # on PE-4: "PE4-loc"; same functions
                    function
                        end-dt4
                        end-dt6
                        end-dt46
                    exit
                exit
           exit
           bgp-evpn
                segment-routing-v6 bgp 1
                    route-distinguisher 192.0.2.1:31            # on PE-4: 192.0.2.4:32
                    srv6-instance 1 default-locator "PE1-loc"   # on PE-4: "PE4-loc"
                    source-address 2001:db8::2:1                # on PE-4: 2001:db8::2:4
                    vrf-target target:64500:31                  # on PE-4: target:64500:32
                    no shutdown
                exit
            exit
            no shutdown
```

On GWs PE-2 and PE-3, VPRN-3 has two BGP-EVPN instances that both point to the same locator, as follows:

```
# on PE-2:
configure
    service
        vprn 3 name "VPRN-3" customer 1 create
            segment-routing-v6 1 create
                locator "PE2-loc"                # on PE-3: "PE3-loc"; same functions
                    function
                        end-dt4
                        end-dt6
                        end-dt46
                    exit
                exit
            exit
            bgp-evpn
                segment-routing-v6 bgp 1
                    domain-id 64500:1
                    route-distinguisher 192.0.2.2:31            # on PE-3: 192.0.2.3:31
                    srv6-instance 1 default-locator "PE2-loc"   # on PE-3: "PE3-loc"
                    source-address 2001:db8::2:2                # on PE-3: 2001:db8::2:3
                    vrf-target target:64500:31
                    no shutdown
                exit
                segment-routing-v6 bgp 2
                    domain-id 64500:2
                    route-distinguisher 192.0.2.2:32           # on PE-3: 192.0.2.3:32
                    srv6-instance 1 default-locator "PE2-loc"  # on PE-3: "PE3-loc"
                    source-address 2001:db8::2:2               # on PE-3: 2001:db8::2:3
                    vrf-target target:64500:32
                    no shutdown
                exit
            exit
            allow-export-bgp-vpn
            no shutdown
```

Verification

The domain path attribute is used for loop prevention. GW PE-2 does not use the IP prefix routes readvertised by GW PE-3:

```
*A:PE-2# show router bgp routes evpn ip-prefix rd 192.0.2.3:31
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
Flag  Route Dist.         Prefix
      Tag                 Gw Address
                          NextHop
                          Label
                          ESI
-------------------------------------------------------------------------------
*>i   192.0.2.3:31        10.0.3.31/32
      0                   00:00:00:00:00:00
                          2001:db8::2:3
                          524281
                          ESI-0

*>i   192.0.2.3:31        10.0.3.34/32
      0                   00:00:00:00:00:00
                          2001:db8::2:3
                          524281
                          ESI-0

-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

The detailed output of the preceding command on PE-2 shows that the End.DT4 function is used and that PE-2 detects a domain path loop in VRF 3 for EVPN IP prefix routes with RD 192.0.2.3:31:

```
*A:PE-2# show router bgp routes evpn ip-prefix rd 192.0.2.3:31 detail
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
Original Attributes

Network        : n/a
Nexthop        : 2001:db8::2:3
Path Id        : None
From           : 2001:db8::2:3
Res. Nexthop   : fe80::203:feff:fe00:0
Local Pref.    : 100                    Interface Name : int-PE-2-PE-3
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64500:31
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.3
Origin         : IGP
Flags          : Valid Best
Route Source   : Internal
AS-Path        : No As-Path
D-Path         : [64500:1:(evpn)]
EVPN type      : IP-PREFIX
ESI            : ESI-0
Tag            : 0
Gateway Address: 00:00:00:00:00:00
```

```
Prefix         : 10.0.3.31/32
Route Dist.    : 192.0.2.3:31
MPLS Label     : 524281
Route Tag      : 0
Neighbor-AS    : n/a
DB Orig Val    : N/A                    Final Orig Val : N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 18h46m24s
SRv6 TLV Type  : SRv6 L3 Service TLV (5)
SRv6 SubTLV    : SRv6 SID Information (1)
Sid            : 2001:db8:aaaa:103::
Full Sid       : 2001:db8:aaaa:103:7fff:9000::
Behavior       : End.DT4 (19)
SRv6 SubSubTLV : SRv6 SID Structure (1)
Loc-Block-Len  : 48                     Loc-Node-Len   : 16
Func-Len       : 20                     Arg-Len        : 0
Tpose-Len      : 20                     Tpose-offset   : 64
DPath Loop VRFs: 3
---snip---
```

The IPv4 route table for VPRN-3 on PE-2 is as follows:

```
*A:PE-2# show router service-name "VPRN-3" route-table

===============================================================================
Route Table (Service: 3)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.3.31/32                                  Remote  EVPN-IFL  19h55m28s  170
       2001:db8:aaaa:101:8000:: (tunneled:SRV6)                     10
10.0.3.34/32                                  Remote  EVPN-IFL  19h55m28s  170
       2001:db8:aaaa:104:7fff:a000:: (tunneled:SRV6)                10
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The behavior is the same as in the preceding use cases. The same **show** commands can be used to verify that.

VPRN with two BGP-EVPN instances pointing at different SRv6 locators

This section describes VPRN-4, which has two BGP-EVPN instances pointing at different locators.

SRv6 locator configuration

On PE-2, PE-3, and PE-4, an additional SRv6 locator is configured. In this example, the only difference is the IP prefix. On PE-2, the "PE2-loc" locator was already configured and the "PE2-loc-2" locator is added:

```
# on PE-2:
configure
    router Base
```

```
        segment-routing
            segment-routing-v6
                origination-fpe 1
                source-address 2001:db8::2:2
                locator "PE2-loc"
                    block-length 48
                    termination-fpe 2
                    prefix
                        ip-prefix 2001:db8:aaaa:102::/64
                    exit
                    static-function
                    exit
                    no shutdown
                exit
                locator "PE2-loc-2"
                    block-length 48
                    termination-fpe 2
                    prefix
                        ip-prefix 2001:db8:aaaa:122::/64
                    exit
                    static-function
                    exit
                    no shutdown
                exit
                base-routing-instance
                    locator "PE2-loc"
                        function
                            end-x-auto-allocate srh-mode psp protection unprotected
                            end 1
                                srh-mode usp
                            exit
                        exit
                    exit
                    locator "PE2-loc-2"
                        function
                            end-x-auto-allocate srh-mode psp protection unprotected
                            end 1
                                srh-mode usp
                            exit
                        exit
                    exit
                exit
            exit
        exit
        isis 0
            segment-routing-v6
                locator "PE2-loc"
                    level 1
                    exit
                    level 2
                    exit
                exit
                locator "PE2-loc-2"
                    level 1
                    exit
                    level 2
                    exit
                exit
                no shutdown
            exit
```

Likewise, PE-3 gets additional locator "PE3-loc-2" and PE-4 gets additional locator "PE4-loc-2".

BGP configuration

The BGP configuration is the same as for VPRN-3: BGP is enabled for the EVPN address family and the peer addresses are the IPv6 system addresses.

Service configuration

On PE-1, VPRN-4 is configured as follows:

```
#  on PE-1:
configure
    service
        vprn 4 name "VPRN-4" customer 1 create
            interface "loopback" create
                address 10.0.4.41/32
                ipv6
                    address 2001:db8::4:41/128
                exit
                loopback
            exit
            segment-routing-v6 1 create
                locator "PE1-loc"
                    function
                        end-dt4
                        end-dt6
                        end-dt46
                    exit
                exit
           exit
           bgp-evpn
                segment-routing-v6 bgp 1
                    route-distinguisher 192.0.2.1:41
                    srv6-instance 1 default-locator "PE1-loc"
                    source-address 2001:db8::2:1
                    vrf-target target:64500:41
                    no shutdown
                exit
            exit
            no shutdown
```

On PE-2, VPRN-4 is configured with two SRv6 instances that use different locators, as follows. The configuration on PE-3 is similar.

```
#  on PE-2:
configure
    service
        vprn 4 name "VPRN-4" customer 1 create
            segment-routing-v6 1 create
                locator "PE2-loc"               # on PE-3: "PE3-loc"; same functions
                    function
                        end-dt4
                        end-dt6
                        end-dt46
                    exit
                exit
            exit
            segment-routing-v6 2 create
                locator "PE2-loc-2"              # on PE-3: "PE3-loc-2"; same function
                    function
```

```
                        end-dt46
                    exit
                exit
            exit
            bgp-evpn
                segment-routing-v6 bgp 1
                    domain-id 64500:1
                    route-distinguisher 192.0.2.2:41           # on PE-3: 192.0.2.3:41
                    srv6-instance 1 default-locator "PE2-loc"   # on PE-3: "PE3-loc"
                    source-address 2001:db8::2:2                # on PE-3: 2001:db8::2:3
                    vrf-target target:64500:41
                    no shutdown
                exit
                segment-routing-v6 bgp 2
                    domain-id 64500:2
                    route-distinguisher 192.0.2.2:42            # on PE-3: 192.0.2.3:42
                    srv6-instance 2 default-locator "PE2-loc-2"  # on PE-3: "PE3-loc-2"
                    source-address 2001:db8::2:2                 # on PE-3: 2001:db8::2:3
                    vrf-target target:64500:42
                    no shutdown
                exit
            exit
            allow-export-bgp-vpn
            no shutdown
```

On PE-4, VPRN-4 is configured as follows:

```
#  on PE-4:
configure
    service
        vprn 4 name "VPRN-4" customer 1 create
            interface "loopback" create
                address 10.0.4.44/32
                ipv6
                    address 2001:db8::4:44/128
                exit
                loopback
            exit
            segment-routing-v6 1 create
                locator "PE4-loc-2"
                    function
                        end-dt46
                    exit
                exit
            exit
            bgp-evpn
                segment-routing-v6 bgp 1
                    route-distinguisher 192.0.2.4:42
                    srv6-instance 1 default-locator "PE4-loc-2"
                    source-address 2001:db8::2:4
                    vrf-target target:64500:42
                    no shutdown
                exit
            exit
            no shutdown
```

Verification

The behavior is similar as in the preceding use cases. Loops are prevented using the domain path attribute. The following shows that PE-3 detects a domain path loop in VRF 4 for a route originating from PE-2:

```
*A:PE-3# show router bgp routes evpn ip-prefix rd 192.0.2.2:41 detail
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN IP-Prefix Routes
===============================================================================
---snip---

-------------------------------------------------------------------------------
Original Attributes

Network        : n/a
Nexthop        : 2001:db8::2:2
Path Id        : None
From           : 2001:db8::2:2
Res. Nexthop   : fe80::202:feff:fe00:0
Local Pref.    : 100                    Interface Name : int-PE-3-PE-2
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64500:41
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.2
Origin         : IGP
Flags          : Valid Best
Route Source   : Internal
AS-Path        : No As-Path
D-Path         : [64500:2:(evpn)]
EVPN type      : IP-PREFIX
ESI            : ESI-0
Tag            : 0
Gateway Address: 00:00:00:00:00:00
Prefix         : 10.0.4.44/32
Route Dist.    : 192.0.2.2:41
MPLS Label     : 524278
Route Tag      : 0
Neighbor-AS    : n/a
DB Orig Val    : N/A                    Final Orig Val : N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 02h23m15s
SRv6 TLV Type  : SRv6 L3 Service TLV (5)
SRv6 SubTLV    : SRv6 SID Information (1)
Sid            : 2001:db8:aaaa:102::
Full Sid       : 2001:db8:aaaa:102:7fff:6000::
Behavior       : End.DT4 (19)
SRv6 SubSubTLV : SRv6 SID Structure (1)
Loc-Block-Len  : 48                     Loc-Node-Len   : 16
Func-Len       : 20                     Arg-Len        : 0
```

```
Tpose-Len      : 20                     Tpose-offset   : 64
DPath Loop VRFs: 4
---snip---
```

The route tables for IPv4 and IPv6 are similar to the ones in the preceding use cases. The IPv4 route tables for VPRN-4 are the following:

```
*A:PE-1# show router service-name "VPRN-4" route-table

===============================================================================
Route Table (Service: 4)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.4.41/32                                  Local   Local     22h51m57s  0
       loopback                                                     0
10.0.4.44/32                                  Remote  EVPN-IFL  02h25m28s  170
       2001:db8:aaaa:102:7fff:6000:: (tunneled:SRV6)                10
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================


*A:PE-2# show router service-name "VPRN-4" route-table

===============================================================================
Route Table (Service: 4)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.4.41/32                                  Remote  EVPN-IFL  22h50m43s  170
       2001:db8:aaaa:101:7fff:f000:: (tunneled:SRV6)                10
10.0.4.44/32                                  Remote  EVPN-IFL  02h25m29s  170
       2001:db8:aaaa:124:7fff:7000:: (tunneled:SRV6)                10
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================



*A:PE-3# show router service-name "VPRN-4" route-table

===============================================================================
Route Table (Service: 4)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.0.4.41/32                                  Remote  EVPN-IFL  22h43m18s  170
       2001:db8:aaaa:101:7fff:f000:: (tunneled:SRV6)                10
10.0.4.44/32                                  Remote  EVPN-IFL  02h25m30s  170
       2001:db8:aaaa:124:7fff:7000:: (tunneled:SRV6)                10
-------------------------------------------------------------------------------
```

```
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================




*A:PE-4# show router service-name "VPRN-4" route-table

===============================================================================
Route Table (Service: 4)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                   Metric
-------------------------------------------------------------------------------
10.0.4.41/32                                  Remote  EVPN-IFL  02h16m56s  170
       2001:db8:aaaa:122:7fff:3000:: (tunneled:SRV6)               10
10.0.4.44/32                                  Local   Local     22h48m18s  0
       loopback                                                    0
-------------------------------------------------------------------------------
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

Conclusion

Multi-instance VPRN services with EVPN-IFL can use SRv6 transport as well as MPLS transport. Interworking between EVPN-IFL and IP-VPN is supported. Multi-instance VPRN services can be used as Service Gateways to connect two SRv6 domains together.

Multi-Segment Pseudowire Routing

This chapter describes advanced multi-segment pseudowire routing configurations.

Topics in this chapter include:

Applicability

Multi-Segment Pseudowire (MS-PW) routing is supported in SR OS Release 9.0.R3, and later. This chapter was initially written for SR OS Release 10.0.R4. The CLI in this edition is based on SR OS Release 21.2.R1. There are no specific prerequisites for this configuration.

Overview

SR OS supports the use of Multi-Segment Pseudowire (MS-PW) routing for Epipe services. MS-PW routing is described in RFC 7267, *Dynamic Placement of Multi-Segment Pseudowires*, and it is an extension of the procedures proposed in RFC 6073 (static MS-PW) to enable multi-segment pseudowires to be dynamically placed. Ultimately, MS-PW Routing provides the capability of setting up MS-PWs without provisioning the Switching PEs (S-PEs).

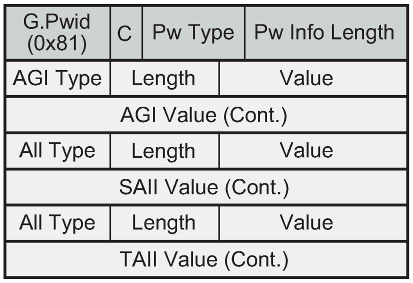
This chapter will go through the configuration process required to set up MS-PW routing and will provide two configuration examples typically deployed by service providers: MS-PW within the same Autonomous System (AS) and MS-PW across two different ASs. Different configuration options are shown and described for each example.

MS-PW routing

From a data plane perspective, MS-PW routing does not introduce any changes with respect to the existing MS-PW architecture. However, from the control plane perspective, MS-PW routing brings a new information model and set of procedures to set up a MS-PW. These are the building blocks defined by the MS-PW routing feature:

- A new information model is introduced for dynamic MS-PW based on the FEC129, Attachment Individual Identifier (AII) Type 2. Static MS-PW uses FEC128 whereas VPLS with BGP-AD uses FEC129, but with AII Type 1 instead.
  - FEC129 is suitable for applications where the local PE with a Source Attachment Individual Identifier (SAII) must automatically learn the remote Target Attachment Individual Identifier (TAII), normally through BGP, before launching the LDP mapping message for the pseudowire setup. Figure 207: FEC129 structure shows the FEC129 structure:

Figure 207: FEC129 structure



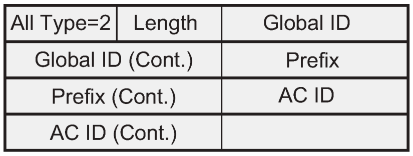
G.Pwid (0x81)	C	Pw Type	Pw Info Length
AGI Type	Length	Value	
AGI Value (Cont.)			
AII Type	Length	Value	
SAII Value (Cont.)			
AII Type	Length	Value	
TAII Value (Cont.)			

ACG0004A

- – The Attachment Group Identifier (AGI) is not used in dynamic MS-PW signaling. In VPLS, it typically carries the instance identifier. It is zero in dynamic MS-PWs.
- – The SAII and TAII (or pseudowire end-point identifiers) are encoded in FEC129 and can have two different formats: AII Type 1 or AII Type 2.

- AII Type 1 is composed of a fixed 32-bit value unique on the local PE. This AII type is used by VPLS when BGP-AD is needed.
- Figure 208: AII type 2 format shows the AII type 2 format. AII type 2 is composed of global-ID:prefix: attachment-circuit-ID (GID:prefix:AC-ID) and allows for summarization, thereby enhancing scalability in large networks. The GID is normally derived from the AS number, the prefix from the node system address, and the AC-ID is the local pseudowire end-point identifier. The combination of the three identifiers gives us a globally unique 96-bit AII value. In general, the same global ID and prefix are assigned for all ACs belonging to the same Terminating PE (T-PE). This is not a strict requirement though.

Figure 208: AII type 2 format

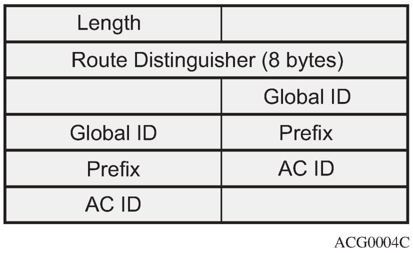


AII Type=2	Length	Global ID
Global ID (Cont.)		Prefix
Prefix (Cont.)		AC ID
AC ID (Cont.)		

ACG0004B

- A MS-PW routing table must be built in all the T-PEs and S-PEs through one of the following two mechanisms:
  - – Multi-protocol BGP (MP-BGP), using a dedicated NLRI and SAFI (pseudowire routing SAFI=6, with AFI=25 for L2-VPN). The FEC129 AII Type 2 global values are mapped in the pseudowire routing NLRI and advertised by BGP. SR OS supports an NLRI comprising a Length, RD, Global ID, and 32-bit prefix, that is, the AC ID is not included in the advertised NLRI. The AC ID is not included as indicated in RFC 7267 because the source T-PE knows by provisioning the AC ID on the terminating T-PE to use in signaling. Therefore, there is no need to advertise a "fully qualified" 96-bit address on a per pseudowire attachment circuit basis. Only the T-PE Global ID, Prefix, and prefix length need to be advertised as part of well-known BGP procedures. This also minimizes the amount of routing information that is advertised in BGP to only what is necessary to reach the far-end T-PE. Figure 209: Pseudowire routing NLRI (the AC ID is always zero) shows the MS-PW routing NLRI:

Figure 209: Pseudowire routing NLRI (the AC ID is always zero)



Length	
Route Distinguisher (8 bytes)	
	Global ID
Global ID	Prefix
Prefix	AC ID
AC ID	

   - Static routes, configurable via CLI
- Once the MS-PW routing table is populated, Targeted LDP (T-LDP) will make use of it to signal the MS-PW all the way from the originating T-PE to the terminating T-PE as well as in the reverse direction. The following methods will be used:
   - At the originating T-PE, a longest-match lookup will be performed in the pseudowire routing table for the configured TAII. Based on the lookup outcome, a label mapping message will be sent to the Next Signaling Hop (NSH).

     Note:
     The "originating T-PE" will be the T-PE initiating the MS-PW signaling. See the Active/passive signaling and auto-configuration section for further information.

   - At the intermediate S-PEs and destination T-PE, a longest-match lookup between the TAII Type 2 included in the T-LDP signaling message and entries installed in the pseudowire routing table will be performed.
   - Alternatively to the pseudowire routing table lookup, T-LDP can also use explicit routing, as per section 7.4.2 of RFC 7267. If that is the case, a "path" must be configured on the T-PEs. The originating T-PE will include an Explicit Route Object (ERO) in the T-LDP label mapping, containing all the S-PE hops specified in the configured path. Each S-PE along the path will remove its own entry from the ERO and will forward the label mapping message to the next hop.

SR OS supports the information model and all the previously described methods:

- Dynamic placement through MP-BGP, with the pseudowire routing NLRI
- Static routes
- Explicit paths

In addition to the above, the following features are supported on dynamic MS-PW:

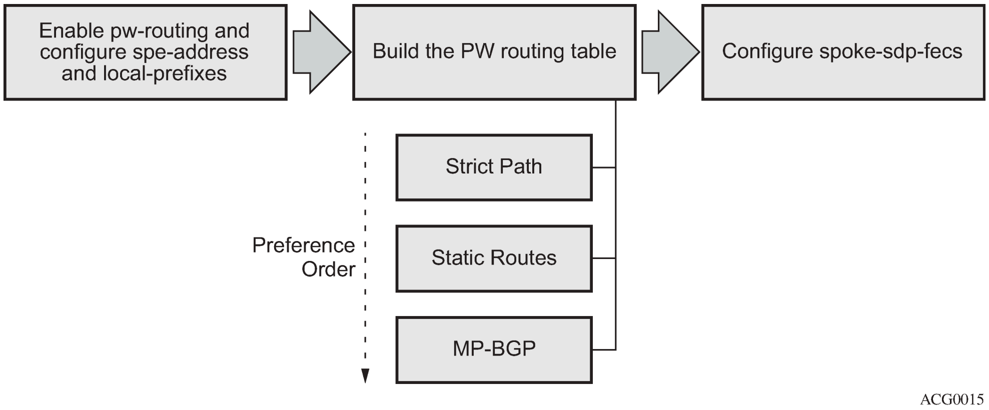
- Auto-configuration of spoke SDPs at T-PE (if enabled on a T-PE, there is no need for configuring the TAII of the remote T-PE, see Active/passive signaling and auto-configuration. The auto-configuration is typically used in hub-and-spoke scenarios. The TAII would only be configured on the spoke T-PE whereas the TAII would be automatically provisioned on the hub T-PE if the auto-config parameter is added.
- OAM using virtual circuit connectivity verification vccv-ping and vccv-trace
- Pseudowire redundancy
- Control word
- Hash label

- Standby-signaling-master and standby-signaling-slave commands
- Filters

Configuration

The following flowchart in Figure 210: Configuration flowchart shows the configuration process to be followed when setting up MS-PW routing. Base IGP and MPLS configuration is assumed to be in place before these configuration tasks can be carried out.

Figure 210: Configuration flowchart


Enable pw-routing and configure spe-address and local-prefixes
Build the PW routing table
Configure spoke-sdp-fecs
Strict Path
Static Routes
MP-BGP
Preference Order
ACG0015

The following subsections review these three steps, including all the options in detail.

- Pseudowire routing enablement
- Building the pseudowire routing table
- Spoke-SDP FEC timers

Pseudowire routing enablement

The first step in the configuration is to enable **pw-routing** and configure the required pseudowire routing basic parameters: the **spe-address** (in S-PEs and T-PEs) and the **local-prefix**/prefixes (only required in T-PEs). The following CLI examples show the configuration of the spe-address and local-prefixes.

```
# on PE-1:
configure
    service
        pw-routing
            spe-address 65536:192.0.2.1
            local-prefix 65536:192.0.2.11 create
                advertise-bgp route-distinguisher 65536:11 community 65535:11
            exit
            local-prefix 65536:192.0.2.12 create
                advertise-bgp route-distinguisher 65536:12 community 65535:12
            exit
        exit
```

In order to enable support for MS-PW routing on an SR OS router, a single, globally unique, S-PE ID (known as the spe-address) is first configured in the **config>service>pw-routing** context on each SR OS router to be used as a T-PE or S-PE. The S-PE address has the format global-id:prefix. It is not possible to configure any local prefixes used for pseudowire routing or to configure spoke SDPs using dynamic MS-PWs at a T-PE, unless an S-PE address has already been configured. The S-PE address is used as the address of a node when populating the switching point TLV in the LDP label mapping message and the pseudowire status notification sent for faults at an S-PE. The following output shows the spe-address configuration format:

```
*A:PE-1# configure service pw-routing spe-address ?
  - no spe-address
  - spe-address <global-id:prefix>

 <global-id:prefix>   : <global-id>:{<prefix>|<ipaddress>}
                        global-id - [1..4294967295]
                        prefix    - [1..4294967295]
                        ipaddress - a.b.c.d
```

Where:

- <global-id> is normally the 2 or 4-byte ASN identifying the network (although nothing prevents the operator from configuring any value here)
- <prefix> is normally the system address of the node (although any value in IP address or decimal format can be used)

If an S-PE is capable of dynamic MS-PW signaling, but is not assigned with an S-PE address, then on receiving a dynamic MS-PW label mapping message, the S-PE will return a label release with the "LDP_RESOURCES_UNAVAILABLE" (0x38)" status code. The S-PE address cannot be changed unless the dynamic MS-PW configuration is completely removed; therefore Nokia recommends to configure the spe-address carefully and keep it for the life of the services.

The second basic pw-routing context parameter is the local-prefix:

```
*A:PE-1# configure service pw-routing local-prefix ?
  - local-prefix <local-prefix> [create]
  - no local-prefix <local-prefix>

 <local-prefix>       : <global-id>:<ip-addr>|<raw-prefix>
                        ip-addr      - a.b.c.d
                        raw-prefix   - [1..4294967295]
                        global-id    - [1..4294967295]

 [no] advertise-bgp   - Configure BGP advertisement
```

One or more local (Layer 2) prefixes (up to a maximum of 16), which are formatted in the style of <global-id>:<ipv4-address>, are supported. A local prefix identifies a T-PE in the pseudowire routing domain. When using explicit paths or static routes, the definition of the local-prefixes without any further attribute is enough. However, when BGP is used, the **advertise-bgp** parameter along with a Route Distinguisher (RD) value and an optional BGP community is required.

```
A:PE-1# configure service pw-routing local-prefix 65536:192.0.2.11 advertise-bgp ?
  - advertise-bgp route-distinguisher <rd> [community <community>]
  - no advertise-bgp route-distinguisher <rd>

 <rd>                 : <ip-addr:comm-val>|<2byte-asnumber:ext-comm-val>|
                                   <4byte-asnumber:comm-val>
                        ip-addr        - a.b.c.d
                        comm-val       - [0..65535]
                        2byte-asnumber - [1..65535]
                        ext-comm-val   - [0..4294967295]
                        4byte-asnumber - [1..4294967295]
 <community>          : <asnumber:comm-val>
                        asnumber - [1..65535]
                        comm-val - [0..65535]
```

Up to four unique RDs (and communities) can be configured per each local prefix. Different RDs for the same prefix allow the operator to advertise the same prefix coming from up to four different Next Signaling Hops (NSHs). Route Reflectors (RRs) would reflect the four routes in that case, whereas only one would be reflected should the same RD be used.

```
*A:PE-1>config>service>pw-routing>local-prefix# info
----------------------------------------------
                advertise-bgp route-distinguisher 400:20
                advertise-bgp route-distinguisher 500:3
                advertise-bgp route-distinguisher 600:300
                advertise-bgp route-distinguisher 65536:11 community 65535:11
```

```
*A:PE-1>config>service>pw-routing>local-prefix# advertise-bgp route-distinguisher 700:100
MINOR: SVCMGR #6072 Maximum number of RD's has been reached
```

For each local prefix, BGP then advertises each global ID/prefix tuple and unique RD and community (if configured) using the MS-PW NLRI, based on the aggregated FEC129 AII Type 2 and the Layer 2 VPN/PW routing AFI/SAFI 25/6, to each BGP neighbor, subject to local BGP policies.

Building the pseudowire routing table

Once the S-PE address and the local prefixes have been configured and before configuring the Epipe service itself on the T-PE nodes, we need to populate the pseudowire routing table in all the participating T-PE and S-PE nodes, so that T-LDP knows what the Next Signaling Hop (NSH) is and sends LDP label mapping messages.

The pseudowire routing table will be populated with local prefixes, static routes, and BGP routes, where the static routes have preference over the BGP-learned routes. The pseudowire routing table can be overridden by the explicit paths, should the operator want to configure them. Therefore, when T-LDP signals an LDP Label Mapping for a TAII, it will:

- First check if there is an explicit path configured for that spoke-SDP FEC.
- Otherwise, it will look up the TAII prefix into the pseudowire routing table, where static routes take precedence over BGP routes.

An aggregation scheme, similar to that used for classless IPv4 addresses, can be employed in the pseudowire routing table, where a longest match is used to find a route. Except for the default pseudowire route, which is encoded with a zero mask, masks included in the pw-routing table are:

- /64 for regular prefixes, including a global ID and prefix (as previously mentioned; the AC-ID is not included in the BGP NLRI).
- /96 for local prefixes, including the AC-ID, as well as global-id and prefix.

Each S-PE and T-PE must have a pseudowire routing table that contains a reference to the T-LDP session to use to signal to a set of next hop S-PEs to reach a T-PE (or the T-PE if that is the next hop). For Epipes, this table contains aggregated AII Type 2 FECs and may be populated with routes that are learned through MP-BGP or that are statically configured.

Explicit paths

A set of default explicit routes to a remote T-PE prefix may be configured on a T-PE under **config>services>pw-routing** using the path name command. Explicit paths are used to populate the explicit route TLV used by MS-PW T-LDP signaling. Only strict (fully qualified) explicit paths are supported. It is possible to configure explicit paths independently of the configuration of BGP or static routing.

The following CLI excerpt shows an explicit path example for a MS-PW following the PE-1–PE-3–PE-5–PE-2 path (see the example topology in Figure 211: Intra-AS MS-PW example topology). The IP addresses are the system addresses of all the S-PE and T-PE along the path (except for PE-1).

```
# on PE-1:
configure
    service
        pw-routing
            path "path-1" create
                hop 1 192.0.2.3
                hop 2 192.0.2.5
                hop 3 192.0.2.2
                no shutdown
            exit
```

Static routes

In addition to support for BGP routing, static MS-PW routes may also be configured using the **config services pw-routing static-route** command. Each static route comprises of the target T-PE global ID and prefix, and the IP address of the T-LDP session to the next hop S-PE or T-PE that should be used:

```
*A:PE-1# configure service pw-routing static-route ?
  - no static-route <route-name>
  - static-route <route-name>

 <route-name>         : <global-id>:<prefix>:<next-hop-ip_addr>
                        global-id      - 0..4294967295
                        prefix         - a.b.c.d|0..4294967295
                        ip_addr        - a.b.c.d
```

If a static route <global-id>:<prefix> is set to 0, then this represents the default route.

```
# on PE-1:
configure
    service
        pw-routing
            static-route 0:0.0.0.0:192.0.2.3
            static-route 0:0.0.0.0:192.0.2.4
```

Even though several default routes can be configured, only one default route is added to the PW routing table. The following command shows the PW routing table content where only one default route (out of the two previously configured ones) is added. The default route added to the PW routing table is the first valid route added to the configuration.

```
*A:PE-1# show service pw-routing route-table all-routes

===============================================================================
Service PW L2 Routing Information
```

```
===============================================================================
AII-Type2/Prefix-Len                           Next-Hop       Owner  Age
 Route-Distinguisher                            Community      Best
-------------------------------------------------------------------------------
0:0.0.0.0:0/0                                  192.0.2.3      static 00h00m11s
 0:0                                            0:0            yes
---snip---
```

If a static route exists to a T-PE, then this is used in preference to any BGP route that may exist.

BGP routes

As already mentioned, the dynamic advertisement of the PW routes is enabled for each prefix and RD using the **advertise-bgp** command in the **config>services>pw-routing>local-prefix** context. A BGP export policy is required in order to export MS-PW routes in MP-BGP. This can be done using a default policy matching all the MS-PW routes, such as the following:

```
# on PE-1:
configure
    router
        autonomous-system 65536
        policy-options
            begin
            policy-statement "export_ms-pw"
                entry 10
                    from
                        family ms-pw
                    exit
                    action accept
                        origin igp
                    exit
                exit
            exit
            commit
        exit
        bgp
            enable-peer-tracking
            rapid-withdrawal
            group "region"
                family ms-pw
                export "export_ms-pw"
                peer-as 65536
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
            exit
        exit
```

MS-PW routes advertised/received can be debugged and shown on the log sessions (**debug router bgp update**). A dedicated MS-PW address family and NLRI are used to distribute the MS-PW prefixes. The following BGP update is sent by PE-1 to PE-3:

```
# on PE-1:
2 2021/03/03 08:55:53.651 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.3
"Peer 1: 192.0.2.3: UPDATE
Peer 1: 192.0.2.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 51
    Flag: 0x90 Type: 14 Len: 26 Multiprotocol Reachable NLRI:
        Address Family MSPW
        NextHop len 4 NextHop 192.0.2.1
        [MSPW] rd: 65536:12, global-id 65536, prefix 192.0.2.12,  ac-id 0, preflen 128
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        65535:12
"
```

MS-PW BGP routes can also be displayed in the pseudowire routing table along with the static routes and the local prefixes.

```
*A:PE-1# show service pw-routing route-table

===============================================================================
Service PW L2 Routing Information
===============================================================================
AII-Type2/Prefix-Len                           Next-Hop       Owner  Age
 Route-Distinguisher                            Community      Best
-------------------------------------------------------------------------------
0:0.0.0.0:0/0                                  192.0.2.3      static 00h05m58s
 0:0                                            0:0            yes
65536:192.0.2.11:0/64                          192.0.2.1      local  00h09m53s
 0:0                                            0:0            yes
65536:192.0.2.11:0/64                          192.0.2.1      local  00h09m53s
 65536:11                                       65535:11       yes
65536:192.0.2.12:0/64                          192.0.2.1      local  00h09m53s
 0:0                                            0:0            yes
65536:192.0.2.12:0/64                          192.0.2.1      local  00h09m53s
 65536:12                                       65535:12       yes
65536:192.0.2.21:0/64                          192.0.2.3      bgp    00h02m29s
 65536:21                                       65535:11       yes
65536:192.0.2.22:0/64                          192.0.2.4      bgp    00h02m44s
 65536:22                                       65535:12       yes
-------------------------------------------------------------------------------
Entries found: 7
===============================================================================
```

If there are two (or more) equal cost BGP MS-PW routes with identical <global-ID:prefix> and different RDs in the RIB, they are both tagged as best/used and both will be added to the pseudowire routing table; however, only the one with a higher RD will be shown as “Best” and as a result of that, only that one will be used by T-LDP for the NSH.

The **pw-routing** context on PE-2 contains the following **advertise-bgp** entries with different RDs for local-prefix 65536:192.0.2.2:

```
# on PE-2:
configure
    service
        pw-routing
            local-prefix 65536:192.0.2.2 create
                advertise-bgp route-distinguisher 65536:21 community 65535:11
                advertise-bgp route-distinguisher 65536:22 community 65535:12
            exit
```

The following CLI output shows an example of two equal cost MS-PW routes. The route 65536:192.0.2.2 with RD 65536:21 and next-hop 192.0.2.3 is tagged as best and used; the route with RD 65536:22 and next-hop 192.0.2.4 is also best and used (u*>).

```
*A:PE-1# show router bgp routes ms-pw aii-type2 65536:192.0.2.2:0
===============================================================================
 BGP Router ID:192.0.2.1        AS:65536       Local AS:65536
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP MSPW Routes
===============================================================================
Flag  Network                 RD
      Nexthop                 AII-Type2/Preflen
      As-Path
-------------------------------------------------------------------------------
u*>i  65536:192.0.2.2         65536:21
      192.0.2.3               65536:192.0.2.2:0/64
      No As-Path
*i    65536:192.0.2.2         65536:21
      192.0.2.4               65536:192.0.2.2:0/64
      No As-Path
u*>i  65536:192.0.2.2         65536:22
      192.0.2.4               65536:192.0.2.2:0/64
      No As-Path
*i    65536:192.0.2.2         65536:22
      192.0.2.3               65536:192.0.2.2:0/64
      No As-Path
-------------------------------------------------------------------------------
Routes : 4
===============================================================================
```

However, only the one with RD 65536:22 (higher RD) is added as “Best” to the pseudowire routing table and T-LDP will use 192.0.2.4 as the NSH:

```
*A:PE-1# show service pw-routing route-table all-routes

===============================================================================
Service PW L2 Routing Information
===============================================================================
AII-Type2/Prefix-Len                           Next-Hop       Owner  Age
 Route-Distinguisher                            Community      Best
-------------------------------------------------------------------------------
---snip---
65536:192.0.2.2:0/64                           192.0.2.3      bgp    00h01m13s
 65536:21                                       65535:11       no
65536:192.0.2.2:0/64                           192.0.2.4      bgp    00h01m31s
 65536:22                                       65535:12       yes
---snip---
```

In the preceding example, the routes have different RDs, but that need not be the case. When the originating T-PE or any intermediate S-PE receives two (or more) equal cost MS-PW routes with the *same RD* but from different Next-Hops (NHs), all the MS-PW routes will be added to the MS-PW routing table as “Best”.

In case multiple equal-cost MS-PW routes are available, T-LDP will pick up the NSH out of an ECMP hashing algorithm applied to the <global-ID:prefix:AC-ID> for the SAII and the TAII of the pseudowires pointing at the same prefix. The output of that hashing algorithm will determine what the NSH will be for a spoke-SDP FEC.

When path diversity for an active and a standby pseudowire (hot standby pseudowire redundancy) is desired and the two pseudowires of the same Epipe endpoint are pointing at the same remote <global-ID:prefix> coming from two different NHs, the operator has to make sure T-LDP chooses a different NSH for the standby pseudowire. Only in that case, hot standby pseudowire redundancy can be achieved. As a rule of thumb, if the SAII/TAII of the active and standby pseudowires are separated by 16 or more AC-ID values, T-LDP will select a different NSH for both pseudowires.

For example:

- Given the following SAII/TAII AC-ID values for the active/standby pseudowires on the originating T-PE, T-LDP will select the same NSH:
  - Active pseudowire: saii-type2 — 65536:192.0.2.1:1, taii-type2 — 65536:192.0.2.2:1
  - Standby pseudowire: saii-type2 — 65536:192.0.2.1:2, taii-type2 — 65536:192.0.2.2:2
- However, the following SAII/TAII AC-ID values for the active/standby pseudowires on the originating T-PE will allow the ECMP hashing algorithm to make T-LDP select different NSHs for the active and the standby pseudowires:
  - Active pseudowire: saii-type2 — 65536:192.0.2.1:1, taii-type2 — 65536:192.0.2.2:1
  - Standby pseudowire: saii-type2 — 65536:192.0.2.1:16, taii-type2 — 65536:192.0.2.2:16

Other AC-ID values greater than 16 (for the standby pseudowire) would also have achieved next hop diversity.

Configuring dynamic pseudowires on the T-PEs

Before any LDP signaling can take place, T-LDP sessions must be explicitly configured on T-PEs and S-PEs.

One or more spoke-SDPs may be configured for distributed Epipe VLL services. Dynamic MS-PWs use FEC129 (also known as the Generalized ID FEC) with AII Type 2 to identify the pseudowire, as opposed to FEC128 (also known as the PW ID FEC) used for traditional single segment pseudowires and for pseudowire switching. FEC129 spoke-SDPs are configured under the **spoke-sdp-fec** command in the CLI. Spoke-SDP FECs (or FEC129 spoke-SDPs) are by default FEC type 129 and AII type 2. Spoke-SDP FECs can be part of an endpoint and even an Inter-Chassis Backup (ICB) pseudowire.

```
*A:PE-1# configure service epipe "Epipe2" spoke-sdp-fec ?
  - no spoke-sdp-fec <spoke-sdp-fec-id>
  - spoke-sdp-fec <spoke-sdp-fec-id> [fec <fec-type>] [aii-type <aii-type>] [create]
  - spoke-sdp-fec <spoke-sdp-fec-id> no-endpoint
  - spoke-sdp-fec <spoke-sdp-fec-id> [fec <fec-type>] [aii-type <aii-type>] [create]
                  endpoint <name> [icb]

 <spoke-sdp-fec-id>   : [1..4294967295]
 <fec-type>           : [129..130]
 <aii-type>           : [1..2]
 <name>               : [32 chars max]
 <icb>                : keyword - configure spoke-sdp as inter-chassis backup
```

FEC129 AII Type 2 uses a SAII and a TAII to identify the ends of a PW at the T-PE. The SAII identifies the local end, while the TAII identifies the remote end. The SAII and TAII are each structured as follows:

- Global ID: this is a 4-byte identifier that uniquely identifies an operator or the local network. Normally, this matches the ASN
- Prefix: a 4-byte prefix, which should correspond to one of the local prefixes assigned under pw-routing
- AC ID: a 4-byte identifier for this end of the PW. This should be locally unique within the scope of the global-id:prefix

In terms of the SDP tunnel being used by each spoke-SDP FEC, PW routing chooses the MS-PW path in terms of the sequence of S-PEs to use to reach a T-PE. It does not select the SDP to use on each hop, which is instead determined at signaling time. When a label mapping is sent for a PW segment, an LDP SDP will be used to reach the next-hop S-PE/T-PE, if such an SDP exists. If not, and an RFC 3107 labeled BGP SDP is available, then that will be used. Otherwise, the label mapping will fail and a label release will be sent.

Note:
The RSVP SDPs might be picked at the T-PE through the use of pw-template <policy-id> [use-provisioned-sdp], however there is no way to select an RSVP SDP on an S-PE.

The following CLI output shows one example of two spoke-SDP FECs belonging to an endpoint:

```
# on PE-1:
configure
    service
        pw-template 1 name "PW1" create
            controlword
        exit
        epipe 2 name "Epipe2" customer 1 create
            description "ms-pw epipe with bgp - using 2 prefixes"
            endpoint "CORE" create
                description "endpoint for epipe A/S PW redundancy"
                revert-time 10
                standby-signaling-master
            exit
            sap 1/1/4:2 create
            exit
            spoke-sdp-fec 21 fec 129 aii-type 2 create endpoint CORE
                precedence primary
                pw-template-bind 1
                saii-type2 65536:192.0.2.11:1
                taii-type2 65536:192.0.2.21:1
                no shutdown
            exit
            spoke-sdp-fec 22 fec 129 aii-type 2 create endpoint CORE
                pw-template-bind 1
                saii-type2 65536:192.0.2.12:1
                taii-type2 65536:192.0.2.22:1
                no shutdown
            exit
            no shutdown
        exit
```

The following options are available in the **spoke-sdp-fec** context:

```
*A:PE-1# configure service epipe "Epipe2" spoke-sdp-fec ?
  - no spoke-sdp-fec <spoke-sdp-fec-id>
  - spoke-sdp-fec <spoke-sdp-fec-id> [fec <fec-type>] [aii-type <aii-type>] [create]
  - spoke-sdp-fec <spoke-sdp-fec-id> no-endpoint
  - spoke-sdp-fec <spoke-sdp-fec-id> [fec <fec-type>] [aii-type <aii-type>] [create]
                                     endpoint <name> [icb]
```

```
 <spoke-sdp-fec-id>   : [1..4294967295]
 <fec-type>           : [129..130]
 <aii-type>           : [1..2]
 <name>               : [32 chars max]
 <icb>                : keyword - configure spoke-sdp as inter-chassis backup

 [no] auto-config     - Configure auto-configuration
 [no] path            - Configure path-name
 [no] precedence      - Configure precedence
 [no] pw-template-bi* - Configure Pseudo-Wire template-binding policy
 [no] retry-count     - Configure retry count
 [no] retry-timer     - Configure retry timer
 [no] saii-type2      - Configure Source Attachment Individual Identifier (SAII)
 [no] shutdown        - Administratively enable/disable the spoke SDP FEC binding
      signaling       - Configure Spoke-SDP FEC signaling
 [no] standby-signal* - Enable PW standby-signaling slave
 [no] taii-type2      - Configure Target Attachment Individual Identifier (TAII)
```

Active/passive signaling and auto-configuration

When an MS-PW is signaled, each T-PE might independently initiate signaling of the MS-PW. This could result in a different path being used in each direction of the PW. To avoid this situation, one of the T-PEs will start the PW signaling (active role), while the other T-PE waits to receive the LDP label mapping message before sending the LDP label mapping message for the reverse direction of the PW (passive role).

Debugging for LDP messages is enabled on PE-2, as follows:

```
# on PE-2:
debug
    router "Base"
        ldp
            peer 192.0.2.5
                packet
                    init detail
                    label detail
                exit
            exit
            peer 192.0.2.6
                packet
                    init detail
                    label detail
                exit
            exit
        exit
```

By default, the T-PE with SAII>TAII will have the active role and will send the label mapping first. When spoke-SDP FEC 21 is first disabled, and then enabled, PE-2 sends a label mapping to PE-5 first (message 77 in following output). Afterward, it receives a label mapping packet from PE-5 (message 78).

```
# on PE-2:
configure
    service
        epipe "Epipe2"
            spoke-sdp-fec 21
                shutdown
                no shutdown
```

```
33 2021/03/03 09:06:57.536 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Label Mapping packet (msgId 146) to 192.0.2.5:0
Protocol version = 1
Label 524277 advertised for the following FECs
Service FEC GENPWE3: ENET(5)
AGI = type: 1, len: 8, val: 00:00:00:00:00:00:00:00
SAII = T: 2, L: 12, Global-id: 65536, Prefix: 192.0.2.21, AcId: 1
TAII = T: 2, L: 12, Global-id: 65536, Prefix: 192.0.2.11, AcId: 1
Group ID = 0 cBit = 1
Interface parameter Mtu = 1500
Interface parameter VCCV = 0x306
PW status bits = 0x0
"
```

```
34 2021/03/03 09:06:57.550 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Recv Label Mapping packet (msgId 154) from 192.0.2.5:0
Protocol version = 1
Label 524277 advertised for the following FECs
Service FEC GENPWE3: ENET(5)
AGI = type: 1, len: 8, val: 00:00:00:00:00:00:00:00
SAII = T: 2, L: 12, Global-id: 65536, Prefix: 192.0.2.11, AcId: 1
TAII = T: 2, L: 12, Global-id: 65536, Prefix: 192.0.2.21, AcId: 1
Group ID = 0 cBit = 1
Interface parameter Mtu = 1500
Interface parameter VCCV = 0x106
PW status bits = 0x18
Switching hop: System = 192.0.2.3, Remote System = 192.0.2.1
previous segment fec AGI = type: 1, len: 8, val: 00:00:00:00:00:00:00:00
SAII = T: 2, L: 12, Global-id: 65536, Prefix: 192.0.2.11, AcId: 1
TAII = T: 2, L: 12, Global-id: 65536, Prefix: 192.0.2.21, AcId: 1
S-PE = T: 2, L: 12, Global-id: 65536, Prefix: 192.0.2.3, AcId: 0
Switching hop: System = 192.0.2.5, Remote System = 192.0.2.3
previous segment fec AGI = type: 1, len: 8, val: 00:00:00:00:00:00:00:00
SAII = T: 2, L: 12, Global-id: 65536, Prefix: 192.0.2.11, AcId: 1
TAII = T: 2, L: 12, Global-id: 65536, Prefix: 192.0.2.21, AcId: 1
S-PE = T: 2, L: 12, Global-id: 65536, Prefix: 192.0.2.5, AcId: 0
"
```

For the other T-PE, it is the other way round. PE-1 receives a label mapping packet first before it sends a label mapping packet back.

This default behavior can be modified by the **signaling** command. When set to master, the T-PE will send a label mapping message regardless of the SAII and TAII. By default the parameter is set to auto (which means the T-PE will trigger label mapping if SAII>TAII).

```
*A:PE-1# configure service epipe "Epipe2" spoke-sdp-fec 21 signaling ?
  - signaling <signaling>

 <signaling>          : auto|master
```

```
# on PE-1:
configure
    service
        epipe "Epipe2"
            spoke-sdp-fec 21
                shutdown
                signaling master
                no shutdown
```

The MS-PW routing implementation on SR OS supports single-sided auto-provisioning. This allows it to have “hub” T-PEs where the TAII is not required to be configured and as such simplifies the provisioning. In this case, the spoke T-PE PWs would be configured with specific SAII and TAII as well as signaling master, whereas the hub T-PE PWs would be configured with only the SAII and the auto-config parameter. When the auto-config attribute is set for a spoke-SDP FEC, the T-PE always passively waits for the label mapping to be received before issuing a label mapping message (because it does not know the TAII beforehand). This is a CLI example for a hub T-PE spoke-SDP FEC:

```
# on PE-2:
configure
    service
        epipe "Epipe2"
            spoke-sdp-fec 21 fec 129 aii-type 2 create
                auto-config
                precedence primary
                pw-template-bind 1
                saii-type2 65536:192.0.2.21:1
                no shutdown
            exit
```

Spoke-SDP FEC timers

MS-PW routing provides a few timers that can be configured at the global pw-routing level or at each specific spoke-SDP FEC level:

```
# on PE-1:
configure
    service
        pw-routing
            boot-timer 20
            retry-timer 40
            retry-count 50
```

```
# on PE-1:
configure
    service
        epipe "Epipe2"
            spoke-sdp-fec 21
                retry-timer 10
                retry-count 10
```

Where:

- **Boot-timer** (the default is 10 seconds with values 0 — 600 seconds allowed): Configures a hold-off timer for MS-PW routing advertisements and signaling that is used at boot time. This timer helps to make sure all the network infrastructure is up and running before setting up the PWs.
- **Retry-timer** (the default is 30 seconds with values 10 — 480 seconds allowed): The exponential back-off timer that determines the interval between consecutive retries to re-establish a spoke-SDP. The configured value gives the initial retry time. The attempt fails if a label withdrawal is received. If configured at global and spoke-SDP FEC level, the latter overrides the value set by the global settings.

- **Retry-count** (the default 30 with values 10 — 10000): Specifies the number of attempts the system should make to re-establish the spoke-SDP after it has failed. After each successful attempt, the counter is reset to zero. When the specified number is reached, no more attempts are made and the spoke-SDP is put into the disabled state. Use the **no shutdown** command to bring up the path after the retry limit is exceeded. It is present at the PW routing level as well as the spoke-SDP FEC level. If configured at global and spoke-SDP FEC level, the latter overrides the value set by the global settings.
- The usual endpoint level timers are also available for MS-PW routing:
  - **Revert-time** <*time-value*> **| infinite** (default is 0, values 0 — 600 sec): configures the time to wait before reverting to the primary spoke-SDP FEC.
  - **Active-hold-delay** (the default is 0, values 0 — 60 deci-seconds): It specifies that the node will delay sending the T-LDP status bits for VLL endpoint when the MC-LAG transitions the LAG subgroup which hosts the SAP from active to standby (MC-Ring or MC-APS are supported too) or when any object in the endpoint—SAP, ICB, or regular spoke SDP—transitions from up to down operational state. The active-hold-delay range starts from 1 (in units of deci-seconds) via CLI, and the only way to get the default value of zero is to use the **no active-hold-delay** command

Standby signaling

Just as with a regular endpoint with regular spoke-SDPs, there can also be standby-signaling-master and standby-signaling-slave parameters for spoke-SDP FECs.

The **standby-signaling-master** command is configured in the **endpoint** context and makes sure that standby signaling (T-LDP PW status bits 0x20) is sent for the selected standby PW.

```
# on PE-1:
configure
    service
        epipe "Epipe2"
            endpoint "CORE"
                standby-signaling-master
```

It is not allowed to add a SAP associated to an endpoint configured as standby-signaling-master to an Epipe.

```
*A:PE-1>config>service>epipe# sap 1/1/4:2 endpoint "CORE" create
MINOR: SVCMGR #6025 The endpoint has standby-signaling-master configured
```

The **standby-signaling-slave** can be configured at endpoint or spoke-SDP FEC level.

```
# at endpoint on PE-1:
configure
    service
        epipe "Epipe2"
            endpoint "CORE"
                no standby-signaling-master
                standby-signaling-slave
```

```
# at spoke-SDP FEC on PE-1:
configure
    service
        epipe "Epipe2"
            spoke-sdp-fec 21
                standby-signaling-slave
```

When **standby-signaling-slave** is configured, the node will block the transmit forwarding direction of a spoke-SDP based on the PW standby bit received from a T-LDP peer.

Spoke-SDP FEC templates and filters

PW templates are the way to configure the control word for this type of PW as well as ingress/egress filters (IPv4/MAC/IPv6). Filters are only supported on the T-PEs, because there is no provisioning of a PW template (or Epipe at all) on the S-PEs.

```
# on PE-1:
configure
    service
        pw-template 1 name "PW1" create
            controlword
            egress
                filter ip 1
            exit
        exit
        epipe 2 name "Epipe2" customer 1 create
---snip---
            spoke-sdp-fec 22 fec 129 aii-type 2 create endpoint CORE
                pw-template-bind 1
                saii-type2 65536:192.0.2.12:1
                taii-type2 65536:192.0.2.22:1
                no shutdown
            exit
```

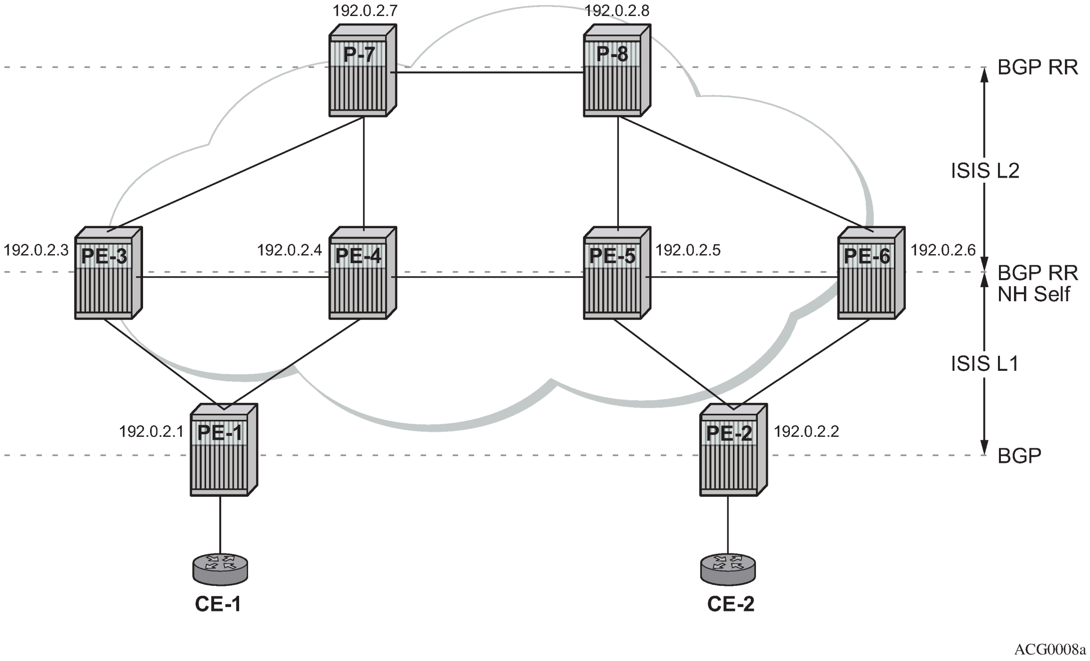
PW template changes (just like for VPLS with BGP-AD or BGP-VPLS) are not automatically propagated. A **tools perform** command is provided to evaluate and distribute the changes at the service level to one or all the services that use that template (if the service ID is omitted, then all the services will be updated).

```
# on PE-1:
tools perform service id 2 eval-pw-template 1 allow-service-impact
```

Intra-AS MS-PW routing

This section provides a configuration example for an intra-AS scenario. Figure 211: Intra-AS MS-PW example topology shows the example topology that will be used for this section.

Figure 211: Intra-AS MS-PW example topology



Multiple MS-PW routing Epipes are to be configured between PE-1 and PE-2, with PE-3, PE-4, PE-5, and PE-6 being S-PE routers. P-7 and P-8 are pure P routers from a data plane perspective.

All the PEs are pre-configured with IS-IS as the IGP, as shown in Figure 211: Intra-AS MS-PW example topology: PE-1 and PE-2 are level-1 routers, P-7 and P-8 are level-2 only routers and the rest of the routers are level-1/level-2. Link level LDP is also pre-configured on all the network interfaces and targeted LDP is configured between PE-1 and PE-3/PE-4, between PE-2 and PE-5/PE-6 and among PE-3, PE-4, PE-5, and PE-6. There are no targeted LDP sessions configured on P-7 and P-8.

As outlined in Figure 210: Configuration flowchart, the configuration is a three-step process where the pw-routing context is configured first, then the required configuration so that routing tables get populated accordingly and finally the services themselves.

MS-PW using BGP routing

In this subsection, Epipe 2 will be configured between PE-1 and PE-2, where T-LDP will use the BGP routes populated in the MS-PW routing table to signal the MS-PW.

The first step is the provisioning of the pw-routing context on all the T-PEs and S-PEs. The **spe-address** will be configured on all the T-PEs and S-PEs—all the routers except for P-7 and P-8—using the ASN as the global ID and the system address as the prefix. On PE-1 and PE-2, (only) the prefixes used for setting up Epipe 2 are configured. Two prefixes are configured per T-PE so that PW redundancy with path diversity for the standby PW can be carried out. The **spe-address** and local prefixes for the T-PEs are shown in the following CLI output. The **advertise-bgp** parameter is required because BGP is used here.

```
# on PE-1:
configure
    service
```

```
        pw-routing
            spe-address 65536:192.0.2.1
            local-prefix 65536:192.0.2.11 create
                advertise-bgp route-distinguisher 65536:11 community 65535:11
            exit
            local-prefix 65536:192.0.2.12 create
                advertise-bgp route-distinguisher 65536:12 community 65535:12
            exit
```

```
# on PE-2:
configure
    service
        pw-routing
            spe-address 65536:192.0.2.2
            local-prefix 65536:192.0.2.21 create
                advertise-bgp route-distinguisher 65536:21 community 65535:11
            exit
            local-prefix 65536:192.0.2.22 create
                advertise-bgp route-distinguisher 65536:22 community 65535:12
            exit
```

The second step is the configuration of BGP.

As shown in Figure 211: Intra-AS MS-PW example topology, BGP is enabled in all the routers. The middle routers (PE-3, PE-4 and PE-5, PE-6) are BGP RRs for PE-1 and PE-2 and they reflect MS-PW routes while changing the next-hop to their own system address. This is required so that T-LDP knows where to send the label mapping message for a particular prefix. P-7 and P-8 are regular RRs reflecting routes among all the S-PEs. The BGP configuration of PE-1, PE-3, PE-4, and P-7 is as follows. Similar commands are configured on the other PEs depending on their T-PE, S-PE, or RR function.

The T-PEs have dual-homed BGP sessions to the S-PEs. Example for PE-1:

```
# on PE-1:
configure
    router
        autonomous-system 65536
        policy-options
            begin
            policy-statement "export_ms-pw"
                entry 10
                    from
                        family ms-pw
                    exit
                    action accept
                        origin igp
                    exit
                exit
            exit
            commit
        exit
        bgp
            enable-peer-tracking
            rapid-withdrawal
            group "region"
                family ms-pw
                export "export_ms-pw"
                peer-as 65536
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
            exit
        exit
```

The S-PEs are reflecting routes and also changing the NH and local preference based on the communities accordingly, so that PW diversity can be ensured.

```
# on PE-3:
configure
    router
        autonomous-system 65536
        policy-options
            begin
            community "65535:11"
                members "65535:11"
            exit
            community "65535:12"
                members "65535:12"
            exit
            policy-statement "export_ms-pw_ABR-to-core"
                entry 10
                    from
                        protocol bgp
                        community "65535:11"
                        family ms-pw
                    exit
                    action accept
                        origin igp
                        local-preference 150
                        next-hop-self
                    exit
                exit
                entry 20
                    from
                        protocol bgp
                        community "65535:12"
                        family ms-pw
                    exit
                    action accept
                        origin igp
                        local-preference 100
                        next-hop-self
                    exit
                exit
            exit
            policy-statement "export_ms-pw_ABR-to-region"
                entry 10
                    from
                        protocol bgp
                        community "65535:11"
                        family ms-pw
                    exit
                    action accept
                        origin igp
                        local-preference 150
                        next-hop-self
                    exit
                exit
                entry 20
                    from
                        protocol bgp
                        community "65535:12"
                        family ms-pw
                    exit
                    action accept
```

```
                        origin igp
                        local-preference 100
                        next-hop-self
                    exit
                exit
            exit
            commit
        exit
        bgp
            rapid-withdrawal
            group "core"
                family ms-pw
                export "export_ms-pw_ABR-to-core"
                peer-as 65536
                neighbor 192.0.2.7
                exit
                neighbor 192.0.2.8
                exit
            exit
            group "region"
                family ms-pw
                cluster 3.3.3.3
                export "export_ms-pw_ABR-to-region"
                peer-as 65536
                enable-peer-tracking
                neighbor 192.0.2.1
                exit
            exit
```

The second S-PE to which PE-1 is connected has the following BGP configuration:

```
# on PE-4:
configure
    router
        autonomous-system 65536
        policy-options
            begin
            community "65535:11"
                members "65535:11"
            exit
            community "65535:12"
                members "65535:12"
            exit
            policy-statement "export_ms-pw_ABR-to-core"
                entry 10
                    from
                        protocol bgp
                        community "65535:12"
                        family ms-pw
                    exit
                    action accept
                        origin igp
                        local-preference 150
                        next-hop-self
                    exit
                exit
                entry 20
                    from
                        protocol bgp
                        community "65535:11"
                        family ms-pw
                    exit
                    action accept
```

```
                        origin igp
                        local-preference 100
                        next-hop-self
                    exit
                exit
            exit
            policy-statement "export_ms-pw_ABR-to-region"
                entry 10
                    from
                        protocol bgp
                        community "65535:12"
                        family ms-pw
                    exit
                    action accept
                        origin igp
                        local-preference 150
                        next-hop-self
                    exit
                exit
                entry 20
                    from
                        protocol bgp
                        community "65535:11"
                        family ms-pw
                    exit
                    action accept
                        origin igp
                        local-preference 100
                        next-hop-self
                    exit
                exit
            exit
            commit
        exit
        bgp
            rapid-withdrawal
            group "core"
                family ms-pw
                export "export_ms-pw_ABR-to-core"
                peer-as 65536
                neighbor 192.0.2.7
                exit
                neighbor 192.0.2.8
                exit
            exit
            group "region"
                family ms-pw
                cluster 4.4.4.4
                export "export_ms-pw_ABR-to-region"
                peer-as 65536
                enable-peer-tracking
                neighbor 192.0.2.1
                exit
            exit
        exit
```

The following is the BGP configuration for the RRs P-7 and P-8:

```
# on P-7 and P-8:
configure
    router
        autonomous-system 65536
        bgp
```

```
            enable-peer-tracking
            rapid-withdrawal
            group "core"
                family ms-pw
                cluster 1.1.1.1
                peer-as 65536
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
                neighbor 192.0.2.6
                exit
            exit
        exit
```

After BGP is properly configured and the BGP update exchange takes place, the RIBs are properly populated and the required prefixes uploaded into the MS-PW routing table.

The following command shows the BGP MS-PW routes on PE-1:

```
*A:PE-1# show router bgp routes ms-pw
===============================================================================
 BGP Router ID:192.0.2.1        AS:65536       Local AS:65536
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP MSPW Routes
===============================================================================
Flag  Network                 RD
      Nexthop                 AII-Type2/Preflen
      As-Path
-------------------------------------------------------------------------------
---snip---
u*>i  65536:192.0.2.21        65536:21
      192.0.2.3               65536:192.0.2.21:0/64
      No As-Path
*i    65536:192.0.2.21        65536:21
      192.0.2.4               65536:192.0.2.21:0/64
      No As-Path
u*>i  65536:192.0.2.22        65536:22
      192.0.2.4               65536:192.0.2.22:0/64
      No As-Path
*i    65536:192.0.2.22        65536:22
      192.0.2.3               65536:192.0.2.22:0/64
      No As-Path
---snip---
-------------------------------------------------------------------------------
```

On PE-1, the MS-PW routing table is as follows:

```
*A:PE-1# show service pw-routing route-table

===============================================================================
Service PW L2 Routing Information
===============================================================================
AII-Type2/Prefix-Len                           Next-Hop       Owner  Age
 Route-Distinguisher                            Community      Best
```

```
-------------------------------------------------------------------------------
---snip---
65536:192.0.2.11:0/64                          192.0.2.1      local  00h31m21s
 65536:11                                       65535:11       yes
---snip---
65536:192.0.2.12:0/64                          192.0.2.1      local  00h31m21s
 65536:12                                       65535:12       yes
65536:192.0.2.21:0/64                          192.0.2.3      bgp    00h23m56s
 65536:21                                       65535:11       yes
65536:192.0.2.22:0/64                          192.0.2.4      bgp    00h24m11s
 65536:22                                       65535:12       yes
-------------------------------------------------------------------------------
```

The two prefixes advertised by PE-2 are properly learned by PE-1 through two different next hops. Now, use each one with a different PW and make sure that the active and standby PW follow different paths in the network.

Once the routes are installed in the MS-PW routing table, the services are configured on PE-1 and PE-2, as follows:

```
# on PE-1:
configure
    service
        pw-template 1 name "PW1" create
            controlword
        exit
        epipe 2 name "Epipe2" customer 1 create
            description "ms-pw epipe with bgp - using 2 prefixes"
            endpoint "CORE" create
                description "endpoint for epipe A/S PW redundancy"
                revert-time 10
                standby-signaling-master
            exit
            sap 1/1/4:2 create
            exit
            spoke-sdp-fec 21 fec 129 aii-type 2 create endpoint CORE
                precedence primary
                pw-template-bind 1
                saii-type2 65536:192.0.2.11:1
                taii-type2 65536:192.0.2.21:1
                no shutdown
            exit
            spoke-sdp-fec 22 fec 129 aii-type 2 create endpoint CORE
                pw-template-bind 1
                saii-type2 65536:192.0.2.12:1
                taii-type2 65536:192.0.2.22:1
                no shutdown
            exit
            no shutdown
        exit
```

```
# on PE-2:
configure
    service
        pw-template 1 name "PW1" create
            controlword
        exit
        epipe 2 name "Epipe2" customer 1 create
            description "ms-pw epipe with bgp - using 2 prefixes"
            endpoint "CORE" create
                description "endpoint for epipe A/S PW redundancy"
                revert-time 10
            exit
            sap 1/1/4:2 create
            exit
            spoke-sdp-fec 21 fec 129 aii-type 2 create endpoint CORE
                precedence primary
                pw-template-bind 1
                saii-type2 65536:192.0.2.21:1
                taii-type2 65536:192.0.2.11:1
                no shutdown
            exit
            spoke-sdp-fec 22 fec 129 aii-type 2 create endpoint CORE
                pw-template-bind 1
                saii-type2 65536:192.0.2.22:1
                taii-type2 65536:192.0.2.12:1
                no shutdown
            exit
            no shutdown
        exit
```

The following command can be executed to verify that the service and spoke-SDP FECs are up:

```
*A:PE-1# show service id 2 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 2                   Vpn Id            : 0
Service Type      : Epipe
---snip---

Admin State       : Up                  Oper State        : Up
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:2                              q-tag        1518    1518    Up   Up
sdp:32766:4294967294 SB(192.0.2.4)       MS-PW        0       1556    Up   Up
sdp:32767:4294967295 SB(192.0.2.3)       MS-PW        0       1556    Up   Up
===============================================================================
```

The SDP-binding identifiers and SDP identifiers are automatically generated by the system.

Use **vccv-trace** to check that the spoke-SDP FECs for the active and standby pseudowires follow different and disjoint paths:

```
*A:PE-1# oam vccv-trace spoke-sdp-fec 21
VCCV-TRACE  with 120 bytes of MPLS payload
1  192.0.2.3  rtt=2.31ms rc=8(DSRtrMatchLabel)
2  192.0.2.5  rtt=4.21ms rc=8(DSRtrMatchLabel)
3  192.0.2.2  rtt=4.67ms rc=3(EgressRtr)
```

```
*A:PE-1# oam vccv-trace spoke-sdp-fec 22
VCCV-TRACE  with 120 bytes of MPLS payload
1  192.0.2.4  rtt=1.72ms rc=8(DSRtrMatchLabel)
2  192.0.2.6  rtt=3.96ms rc=8(DSRtrMatchLabel)
3  192.0.2.2  rtt=4.70ms rc=3(EgressRtr)
```

MS-PW using static routing

In this subsection, Epipe 3 will be configured between PE-1 and PE-2, where T-LDP will use static routes in the MS-PW routing table to signal the MS-PW.

On PE-1 and PE-2 (only), the prefixes used for setting up Epipe 3 are configured. These prefixes could be the same as used for Epipe 2, however, different prefixes are used in this example. The **no advertise-bgp** parameter is required now. The static routes for each remote prefix are also configured. Because we will also have PW redundancy for Epipe 3, two prefixes with static routes pointing at different next-hops will be used:

```
# on PE-1:
configure
    service
        pw-routing
            spe-address 65536:192.0.2.1
            local-prefix 65536:192.0.2.13 create
            exit
            local-prefix 65536:192.0.2.14 create
            exit
            static-route 65536:192.0.2.23:192.0.2.3
            static-route 65536:192.0.2.24:192.0.2.4
```

```
# on PE-2:
configure
    service
        pw-routing
            spe-address 65536:192.0.2.2
            local-prefix 65536:192.0.2.23 create
            exit
            local-prefix 65536:192.0.2.24 create
            exit
            static-route 65536:192.0.2.13:192.0.2.5
            static-route 65536:192.0.2.14:192.0.2.6
```

Static routes are also required at all S-PEs along the path (keeping the path diversity for the prefixes as well) and for both directions:

```
# on PE-3:
configure
    service
        pw-routing
            spe-address 65536:192.0.2.3
            static-route 65536:192.0.2.13:192.0.2.1
            static-route 65536:192.0.2.23:192.0.2.5
```

```
# on PE-4:
configure
    service
        pw-routing
            spe-address 65536:192.0.2.4
            static-route 65536:192.0.2.14:192.0.2.1
            static-route 65536:192.0.2.24:192.0.2.6
```

Finally, once the MS-PW routing tables are properly populated, the services can be configured and brought up:

```
# on PE-1:
configure
    service
        pw-template 1 name "PW1" create
            controlword
        exit
        epipe 3 name "Epipe3" customer 1 create
            description "ms-pw epipe with static routes"
            endpoint "CORE" create
                description "endpoint for epipe A/S PW redundancy"
                revert-time 10
                standby-signaling-master
            exit
            sap 1/1/4:3 create
            exit
            spoke-sdp-fec 31 fec 129 aii-type 2 create endpoint CORE
                precedence primary
                pw-template-bind 1
                saii-type2 65536:192.0.2.13:31
                taii-type2 65536:192.0.2.23:31
                no shutdown
            exit
            spoke-sdp-fec 32 fec 129 aii-type 2 create endpoint CORE
                pw-template-bind 1
                saii-type2 65536:192.0.2.14:32
                taii-type2 65536:192.0.2.24:32
                no shutdown
            exit
            no shutdown
        exit
```

```
# on PE-2:
configure
    service
        pw-template 1 name "PW1" create
            controlword
        exit
        epipe 3 name "Epipe3" customer 1 create
            description "ms-pw epipe with static routes"
            endpoint "CORE" create
                description "endpoint for epipe A/S PW redundancy"
                revert-time 10
                standby-signaling-master
            exit
            sap 1/1/4:3 create
            exit
            spoke-sdp-fec 31 fec 129 aii-type 2 create endpoint CORE
                precedence primary
                pw-template-bind 1
                saii-type2 65536:192.0.2.23:31
                taii-type2 65536:192.0.2.13:31
                no shutdown
            exit
            spoke-sdp-fec 32 fec 129 aii-type 2 create endpoint CORE
                pw-template-bind 1
                saii-type2 65536:192.0.2.24:32
                taii-type2 65536:192.0.2.14:32
                no shutdown
            exit
            no shutdown
        exit
```

Check the status and path of the spoke-SDP FECs with the proper **show** commands and **oam vccv-trace/ping** commands (see previous subsection MS-PW using BGP routing).

MS-PW using explicit paths

In this subsection, Epipe 4 will be configured between PE-1 and PE-2, where T-LDP will use explicit paths to signal the MS-PW, overriding the information given by the MS-PW routing table. Although this mode requires the specific configuration of the hops, one by one, the configuration is only done on the T-PEs, as opposed to the static routes where all the S-PEs must be configured with static routes (a mix of static routes and BGP routes can coexist). The local prefixes shown for Epipe 3 will be re-used here for Epipe 4.

Path-1 and path-2 will be configured hop by hop, using diverse paths. All the S-PE nodes as well as the terminating T-PE must be included in the path.

```
# on PE-1:
configure
    service
        pw-routing
            spe-address 65536:192.0.2.1
            local-prefix 65536:192.0.2.13 create
            exit
            local-prefix 65536:192.0.2.14 create
            exit
            path "path-1" create
                hop 1 192.0.2.3
                hop 2 192.0.2.5
                hop 3 192.0.2.2
                no shutdown
            exit
            path "path-2" create
                hop 1 192.0.2.4
                hop 2 192.0.2.6
                hop 3 192.0.2.2
                no shutdown
            exit
        exit
```

```
# on PE-2:
configure
    service
        pw-routing
            spe-address 65536:192.0.2.2
            local-prefix 65536:192.0.2.23 create
            exit
            local-prefix 65536:192.0.2.24 create
            exit
            path "path-1" create
                hop 1 192.0.2.5
                hop 2 192.0.2.3
                hop 3 192.0.2.1
                no shutdown
            exit
            path "path-2" create
                hop 1 192.0.2.6
                hop 2 192.0.2.4
                hop 3 192.0.2.1
                no shutdown
            exit
        exit
```

Those paths must be specified when configuring the Epipe:

```
# on PE-1:
configure
    service
        epipe 4 name "Epipe4" customer 1 create
            description "ms-pw epipe with explicit paths"
            endpoint "CORE" create
                description "endpoint for epipe A/S PW redundancy"
                revert-time 10
                standby-signaling-master
            exit
            sap 1/1/4:4 create
            exit
            spoke-sdp-fec 41 fec 129 aii-type 2 create endpoint CORE
                precedence primary
                saii-type2 65536:192.0.2.13:41
                taii-type2 65536:192.0.2.23:41
                path "path-1"
                no shutdown
            exit
            spoke-sdp-fec 42 fec 129 aii-type 2 create endpoint CORE
                saii-type2 65536:192.0.2.14:42
                taii-type2 65536:192.0.2.24:42
                path "path-2"
                no shutdown
            exit
            no shutdown
        exit
```

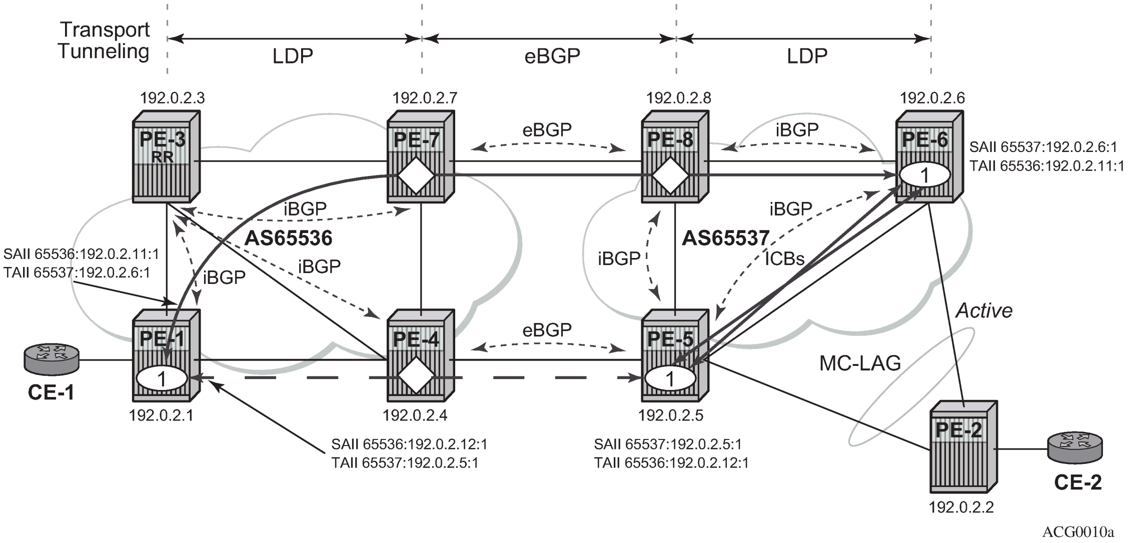
```
# on PE-2:
configure
    service
        epipe 4 name "Epipe4" customer 1 create
            description "ms-pw epipe with explicit paths"
            endpoint "CORE" create
                description "endpoint for epipe A/S PW redundancy"
                revert-time 10
            exit
            sap 1/1/4:4 create
            exit
            spoke-sdp-fec 41 fec 129 aii-type 2 create endpoint CORE
                precedence primary
                saii-type2 65536:192.0.2.23:41
                taii-type2 65536:192.0.2.13:41
                path "path-1"
                no shutdown
            exit
            spoke-sdp-fec 42 fec 129 aii-type 2 create endpoint CORE
                saii-type2 65536:192.0.2.24:42
                taii-type2 65536:192.0.2.14:42
                path "path-2"
                no shutdown
            exit
            no shutdown
        exit
```

Verify the status and path of the spoke-SDP FECs with the proper **show** commands and **oam vccv-trace/ping** commands (see subsection MS-PW using BGP routing).

Inter-AS MS-PW routing

This configuration example for an inter-AS scenario uses BGP tunnels between ASBRs and BGP as the MS-PW routing mechanism. Figure 212: Inter-AS MS-PW example topology shows the example topology used in this section.

Figure 212: Inter-AS MS-PW example topology



In this example, only one Epipe is configured (Epipe 1, using MS-PW BGP routing). The T-PEs are PE-1, PE-5, and PE-6; the S-PEs are PE-7, PE-8, and PE-4.

A/S pseudowire redundancy together with MC-LAG at one end will be used, as shown in Figure 212: Inter-AS MS-PW example topology. Inter-Chassis Backup (ICB) spoke-SDPs between PE-5 and PE-6 are required in order to forward the in-flight packets while MC-LAG and A/S pseudowire are converging, in case of network failures. Those ICBs will also be signaled following the MS-PW routing procedures.

The example topology in Figure 212: Inter-AS MS-PW example topology is pre-configured with the following settings:

- There are two ASs (65536 and 65537) which are connected by two ASBR pairs (PE-7/PE-4 and PE-8/PE-5) running eBGP between them. These eBGP sessions will be used to exchange IPv4-labels (to set up the transport BGP-LBL tunnel, according to the RFC 3107) and MS-PW NLRIs.
- Within AS65536, PE-3 is used as an RR to reflect the MS-PW routes. In AS65537, there is a full mesh of iBGP sessions to distribute the MS-PW routes.
- IS-IS is used within each AS.
- LDP is used as a transport MPLS signaling protocol within each AS and a BGP tunnel will be used between the ASBRs (MS-PW routing supports LDP or BGP tunnels as transport).
- A redundant MC-LAG access to PE-6 and PE-5 is configured.

The next section will go through the configuration required to set up a redundant Epipe between CE-1 and CE-2, by combining A/S pseudowire in the network and MC-LAG at the access.

MS-PW using BGP routing

Epipe 1 will be configured at the end of this section, including the active and redundant pseudowires from PE-1 to PE-5/PE-6, as well as the required ICBs and SAPs at the access.

As discussed, the first step is the provisioning of the pw-routing context. Again, the **spe-address** must be provisioned in all T-PEs and S-PEs whereas prefixes are mandatory only on the T-PEs involved in the service. The following shows the prefixes configured on PE-1, PE-5, and PE-6. Two prefixes are needed in PE-1 in order to make sure that active and standby pseudowires follow disjoint paths.

```
# on PE-1:
configure
    service
        pw-routing
            spe-address 65536:192.0.2.1
            local-prefix 65536:192.0.2.11 create
                advertise-bgp route-distinguisher 65536:11 community 65535:11
            exit
            local-prefix 65536:192.0.2.12 create
                advertise-bgp route-distinguisher 65536:12 community 65535:12
            exit
```

```
# on PE-5:
configure
    service
        pw-routing
            spe-address 65537:192.0.2.5
            local-prefix 65537:192.0.2.5 create
                advertise-bgp route-distinguisher 65537:5 community 65535:5
            exit
```

```
# on PE-6:
configure
    service
        pw-routing
            spe-address 65537:192.0.2.6
            local-prefix 65537:192.0.2.6 create
                advertise-bgp route-distinguisher 65537:6 community 65535:60
            exit
```

Note:
0xFFFF006 and 0xFFFF007 are the values that have been assigned by IANA to the well-known communities for Long-Lived Graceful Restart (LLGR): “LLGR_STALE” and “NO_LLGR”. Therefore, on PE-6, the community value must not be 65535:6, because 65535:6 would be mapped to community “LLGR_STALE”. The value 65535:60 is used instead.

Once the S-PE-addresses and prefixes have been provisioned, BGP must be configured accordingly. A simple BGP export policy is used to export all the local MS-PW prefixes. The configuration on PE-1 is as follows:

```
# on PE-1:
configure
    router
        autonomous-system 65536
        policy-options
            begin
            policy-statement "export_ms-pw"
                entry 10
                    from
                        family ms-pw
                    exit
                    action accept
                        origin igp
                    exit
                exit
            exit
            commit
        exit
        bgp
            enable-peer-tracking
            rapid-withdrawal
            group "intra-AS"
                family ms-pw
                export "export_ms-pw"
                peer-as 65536
                neighbor 192.0.2.3
                exit
            exit
            no shutdown
        exit
    exit
```

The configuration on PE-6 is as follows:

```
# on PE-6:
configure
    router
        autonomous-system 65537
        policy-options
            begin
            policy-statement "export_ms-pw"
                entry 10
                    from
                        family ms-pw
                    exit
                    action accept
                        origin igp
                    exit
                exit
            exit
            commit
        exit
        bgp
            enable-peer-tracking
            rapid-withdrawal
            group "intra-AS"
                family ms-pw
                export "export_ms-pw"
                peer-as 65537
                neighbor 192.0.2.5
                exit
                neighbor 192.0.2.8
                exit
            exit
            no shutdown
        exit
    exit
```

At the ASBR, the BGP policies are more complex because the following tasks must be accomplished:

- ASBR IPv4 system addresses must be exported to the peer ASBR to establish the RFC 3107 BGP tunnel between ASBRs.
- BGP export policies must be used so that MS-PW NLRI exchange can be controlled and attributes like Multi Exit Discriminator (MED) toward the remote AS and/or local-preference (LP) toward the local AS can be modified.
- Finally, BGP import policies must also be used to modify the MS-PW route NH because the T-LDP next signaling hop must match a peer T-LDP system address.

The prefixes 65536:192.0.2.11 and 65537:192.0.2.6 must be preferred in the PE-7/PE-8 pair whereas the prefixes 65536:192.0.2.12 and 65537:192.0.2.5 must be preferred in the PE-4/PE-5 pair, so that the PWs are established as shown in Figure 212: Inter-AS MS-PW example topology. The preference can be propagated by using the BGP MED; the LP is used within the AS, but not relevant to eBGP. The following CLI excerpt shows an example of how to modify MED and LP, as well as changing the NH with an import policy. The configuration on ASBR PE-4 is as follows:

```
*A:PE-4#
configure
    router
        autonomous-system 65536
        policy-options
            begin
            prefix-list "system"
                prefix 192.0.2.4/32 exact
            exit
            community "65535:5"
                members "65535:5"
            exit
            community "65535:11"
                members "65535:11"
            exit
            community "65535:12"
                members "65535:12"
            exit
            community "65535:60"
                members "65535:60"
            exit
            policy-statement "ASBR to ASBR"
                entry 10
                    from
                        protocol bgp
                        community "65535:12"
                        family ms-pw
                    exit
                    action accept
                        origin igp
                        metric set 50
                    exit
                exit
                entry 20
                    from
                        protocol bgp
                        community "65535:11"
                        family ms-pw
                    exit
                    action accept
                        origin igp
                        metric set 100
                    exit
                exit
            exit
            policy-statement "ASBR to region"
                entry 10
                    from
                        protocol bgp
                        community "65535:5"
                        family ms-pw
                    exit
                    action accept
                        origin igp
                        local-preference 150
                        next-hop-self
                    exit
                exit
                entry 20
                    from
                        protocol bgp
                        community "65535:60"
                        family ms-pw
                    exit
                    action accept
                        origin igp
                        next-hop-self
                    exit
                exit
            exit
            policy-statement "export_ipv4_system"
                entry 10
                    from
                        prefix-list "system"
                    exit
                    action accept
                        origin igp
                    exit
                exit
            exit
            policy-statement "import ms-pw NH change"
                entry 10
                    from
                        protocol bgp
                        family ms-pw
                    exit
                    action accept
                        next-hop 192.0.2.5
                    exit
                exit
            exit
            commit
        exit
        bgp
            enable-peer-tracking
            rapid-withdrawal
            group "inter-AS"
                family ms-pw label-ipv4
                import "import ms-pw NH change"
                export "export_ipv4_system" "ASBR to ASBR"
                local-as 65536
                peer-as 65537
                neighbor 192.168.45.2
                exit
            exit
            group "intra-AS"
                family ms-pw
                export "ASBR to region"
                peer-as 65536
                neighbor 192.0.2.3
                exit
            exit
            no shutdown
        exit
```

The configuration on ASBR PE-7 is as follows:

```
# on PE-7:
configure
    router
        autonomous-system 65536
        policy-options
            begin
            prefix-list "system"
                prefix 192.0.2.7/32 exact
            exit
            community "65535:5"
                members "65535:5"
            exit
            community "65535:11"
                members "65535:11"
            exit
            community "65535:12"
                members "65535:12"
            exit
            community "65535:60"
                members "65535:60"
            exit
            policy-statement "ASBR to ASBR"
                entry 10
                    from
                        protocol bgp
                        community "65535:11"
                        family ms-pw
                    exit
                    action accept
                        origin igp
                        metric set 50
                    exit
                exit
                entry 20
                    from
                        protocol bgp
                        community "65535:12"
                        family ms-pw
                    exit
                    action accept
                        origin igp
                        metric set 100
                    exit
                exit
            exit
            policy-statement "ASBR to region"
                entry 10
                    from
                        protocol bgp
                        community "65535:60"
                        family ms-pw
                    exit
                    action accept
                        origin igp
                        local-preference 150
                        next-hop-self
                    exit
                exit
                entry 20
                    from
                        protocol bgp
                        community "65535:5"
                        family ms-pw
                    exit
                    action accept
                        origin igp
                        next-hop-self
                    exit
                exit
            exit
            policy-statement "export_ipv4_system"
                entry 10
                    from
                        prefix-list "system"
                    exit
                    action accept
                        origin igp
                    exit
                exit
            exit
            policy-statement "import ms-pw NH change"
                entry 10
                    from
                        protocol bgp
                        family ms-pw
                    exit
                    action accept
                        next-hop 192.0.2.8
                    exit
                exit
            exit
            commit
        exit
        bgp
            enable-peer-tracking
            rapid-withdrawal
            group "inter-AS"
                family ms-pw label-ipv4
                import "import ms-pw NH change"
                export "export_ipv4_system" "ASBR to ASBR"
                local-as 65536
                peer-as 65537
                neighbor 192.168.78.2
                exit
            exit
            group "intra-AS"
                family ms-pw
                export "ASBR to region"
                peer-as 65536
                neighbor 192.0.2.3
                exit
            exit
            no shutdown
        exit
```

PE-5 and PE-8 have similar configurations to the ones shown. However, PE-5 is a T-PE as well as an ASBR, therefore a local MS-PW prefix must be exported as opposed to only remote prefixes (that is, some export entries for the local MS-PW routes will not contain **protocol bgp** in the matching criteria).

After BGP is properly configured and the updates get exchanged, the RIBs are populated and the prefixes uploaded onto the MS-PW routing table as shown for PE-1 in the following output:

```
*A:PE-1# show router bgp routes ms-pw
===============================================================================
 BGP Router ID:192.0.2.1        AS:65536       Local AS:65536
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP MSPW Routes
===============================================================================
Flag  Network                 RD
      Nexthop                 AII-Type2/Preflen
      As-Path
-------------------------------------------------------------------------------
---snip---
u*>i  65537:192.0.2.5         65537:5
      192.0.2.4               65537:192.0.2.5:0/64
      65537
u*>i  65537:192.0.2.6         65537:6
      192.0.2.7               65537:192.0.2.6:0/64
      65537
-------------------------------------------------------------------------------
Routes : 4
===============================================================================
```

```
*A:PE-1# show service pw-routing route-table

===============================================================================
Service PW L2 Routing Information
===============================================================================
AII-Type2/Prefix-Len                           Next-Hop       Owner  Age
 Route-Distinguisher                            Community      Best
-------------------------------------------------------------------------------
65536:192.0.2.11:0/64                          192.0.2.1      local  00h09m19s
 0:0                                            0:0            yes
65536:192.0.2.11:0/64                          192.0.2.1      local  00h09m19s
 65536:11                                       65535:11       yes
65536:192.0.2.12:0/64                          192.0.2.1      local  00h09m19s
 0:0                                            0:0            yes
65536:192.0.2.12:0/64                          192.0.2.1      local  00h09m19s
 65536:12                                       65535:12       yes
65537:192.0.2.5:0/64                           192.0.2.4      bgp    00h02m34s
 65537:5                                        65535:5        yes
65537:192.0.2.6:0/64                           192.0.2.7      bgp    00h02m01s
 65537:6                                        65535:60       yes
-------------------------------------------------------------------------------
Entries found: 6
===============================================================================
```

For PE-6:

```
*A:PE-6# show router bgp routes ms-pw
===============================================================================
 BGP Router ID:192.0.2.6        AS:65537       Local AS:65537
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
```

```
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP MSPW Routes
===============================================================================
Flag  Network                 RD
      Nexthop                 AII-Type2/Preflen
      As-Path
-------------------------------------------------------------------------------
u*>i  65536:192.0.2.11        65536:11
      192.0.2.8               65536:192.0.2.11:0/64
      65536
u*>i  65536:192.0.2.12        65536:12
      192.0.2.5               65536:192.0.2.12:0/64
      65536
u*>i  65537:192.0.2.5         65537:5
      192.0.2.5               65537:192.0.2.5:0/64
      No As-Path
-------------------------------------------------------------------------------
Routes : 3
===============================================================================
```

```
*A:PE-6# show service pw-routing route-table

===============================================================================
Service PW L2 Routing Information
===============================================================================
AII-Type2/Prefix-Len                            Next-Hop       Owner  Age
 Route-Distinguisher                             Community      Best
-------------------------------------------------------------------------------
65536:192.0.2.11:0/64                           192.0.2.8      bgp    00h02m32s
 65536:11                                        65535:11       yes
65536:192.0.2.12:0/64                           192.0.2.5      bgp    00h02m52s
 65536:12                                        65535:12       yes
65537:192.0.2.5:0/64                            192.0.2.5      bgp    00h02m52s
 65537:5                                         65535:5        yes
65537:192.0.2.6:0/64                            192.0.2.6      local  00h08m43s
 0:0                                             0:0            yes
65537:192.0.2.6:0/64                            192.0.2.6      local  00h08m43s
 65537:6                                         65535:60       yes
-------------------------------------------------------------------------------
Entries found: 5
===============================================================================
```

As can be seen in the preceding **show** commands on PE-6, the two PE-1 prefixes are learned on (PE-5 and) PE-6 through different and disjoint paths. On PE-1, the PE-5 and PE-6 prefixes are learned through two different and disjoint paths.

After configuring the PW routing context and configuring BGP, the last step is the service configuration on the three T-PEs, as follows. T-LDP sessions must have been previously and explicitly configured between the T-PEs and S-PEs (between PE-1 and PE-4/7, between PE-4 and PE-5, PE-7 and PE-8, and between PE-6 and PE-5/8).

```
# on PE-1:
configure
    router Base
        ldp
            targeted-session
                peer 192.0.2.4
                exit
                peer 192.0.2.7
                exit
            exit
        exit
    exit
    service
        pw-template 1 name "PW1" create
            controlword
        exit
        epipe 1 name "Epipe1" customer 1 create
            description "ms-pw epipe with bgp, inter-AS, MC-LAG redundancy"
            endpoint "CORE" create
                description "endpoint for epipe A/S PW redundancy"
            exit
            sap 1/1/4:1 create
            exit
            spoke-sdp-fec 11 fec 129 aii-type 2 create endpoint CORE
                precedence primary
                pw-template-bind 1
                saii-type2 65536:192.0.2.11:1
                taii-type2 65537:192.0.2.6:1
                no shutdown
            exit
            spoke-sdp-fec 12 fec 129 aii-type 2 create endpoint CORE
                pw-template-bind 1
                saii-type2 65536:192.0.2.12:1
                taii-type2 65537:192.0.2.5:1
                no shutdown
            exit
            no shutdown
        exit
```

```
# on PE-5:
configure
    service
        pw-template 1 name "PW1" create
            controlword
        exit
        epipe 1 name "Epipe1" customer 1 create
            description "ms-pw epipe with bgp, inter-AS, MC-LAG redundancy"
            endpoint "CORE" create
                description "endpoint for epipe A/S PW redundancy"
            exit
            endpoint "ACCESS" create
            exit
            sap lag-1:1 endpoint "ACCESS" create
            exit
            spoke-sdp-fec 11 fec 129 aii-type 2 create endpoint CORE
                pw-template-bind 1
                saii-type2 65537:192.0.2.5:1
                taii-type2 65536:192.0.2.12:1
                no shutdown
            exit
            spoke-sdp-fec 12 fec 129 aii-type 2 create endpoint CORE icb
                pw-template-bind 1
                saii-type2 65537:192.0.2.5:2
                taii-type2 65537:192.0.2.6:2
                no shutdown
            exit
            spoke-sdp-fec 13 fec 129 aii-type 2 create endpoint ACCESS icb
                pw-template-bind 1
                saii-type2 65537:192.0.2.5:3
                taii-type2 65537:192.0.2.6:3
                no shutdown
            exit
            no shutdown
        exit
```

```
# on PE-6:
configure
    service
        pw-template 1 name "PW1" create
            controlword
        exit
        epipe 1 name "Epipe1" customer 1 create
            description "ms-pw epipe with bgp, inter-AS, MC-LAG redundancy"
            endpoint "CORE" create
                description "endpoint for epipe A/S PW redundancy"
            exit
            endpoint "ACCESS" create
            exit
            sap lag-1:1 endpoint "ACCESS" create
            exit
            spoke-sdp-fec 11 fec 129 aii-type 2 create endpoint CORE
                pw-template-bind 1
                saii-type2 65537:192.0.2.6:1
                taii-type2 65536:192.0.2.11:1
                no shutdown
            exit
            spoke-sdp-fec 12 fec 129 aii-type 2 create endpoint CORE icb
                pw-template-bind 1
                saii-type2 65537:192.0.2.6:3
                taii-type2 65537:192.0.2.5:3
                no shutdown
            exit
            spoke-sdp-fec 13 fec 129 aii-type 2 create endpoint ACCESS icb
                pw-template-bind 1
                saii-type2 65537:192.0.2.6:2
                taii-type2 65537:192.0.2.5:2
                no shutdown
            exit
            no shutdown
        exit
```

The following **show** commands can be executed to check the status of the Epipe 1 and the pseudowire status signaling received:

```
*A:PE-1# show service id 1 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : Epipe
---snip---

Admin State       : Up                  Oper State        : Up
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/4:1                              q-tag        1518    1518    Up   Up
sdp:32766:4294967294 SB(192.0.2.7)       MS-PW        0       9190    Up   Up
```

```
sdp:32767:4294967295 SB(192.0.2.4)       MS-PW        0       9190    Up   Up
===============================================================================
```

```
*A:PE-1# show service id 1 endpoint

===============================================================================
Service 1 endpoints
===============================================================================
Endpoint name                : CORE
Description                  : endpoint for epipe A/S PW redundancy
Creation Origin              : manual
Revert time                  : 0
Act Hold Delay               : 0
Standby Signaling Master     : false
Standby Signaling Slave      : false
Tx Active (SDP-FEC)          : 11
Tx Active Up Time            : 0d 00:00:35
Revert Time Count Down       : never
Tx Active Change Count       : 2
Last Tx Active Change        : 03/03/2021 10:49:02
-------------------------------------------------------------------------------
Members
-------------------------------------------------------------------------------
Sdp-fec: 11 Prec:0                                Oper Status: Up
Sdp-fec: 12 Prec:4                                Oper Status: Up
===============================================================================
===============================================================================
```

PE-5 will have the MC-LAG standby interface and as such the SAP will be operationally down and will drive the standby signaling to the remote T-PEs:

```
*A:PE-5# show service id 1 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                     Vpn Id            : 0
Service Type      : Epipe
---snip---

Admin State       : Up                    Oper State        : Up
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:lag-1:1                              q-tag        1518    1518    Up   Down
sdp:32766:4294967293 SB(192.0.2.6)       MS-PW        0       9190    Up   Up
sdp:32766:4294967294 SB(192.0.2.6)       MS-PW        0       9190    Up   Up
sdp:32767:4294967295 SB(192.0.2.4)       MS-PW        0       9190    Up   Up
===============================================================================
```

```
*A:PE-5# show service id 1 all | match Flags
Flags              : None
Flags              : None
Flags              : None
Flags              : PortOperDown StandByForMcProtocol
```

The following commands are useful on the S-PEs in order to find the PWs automatically created as well as the SDPs automatically used for those PWs.

```
*A:PE-7# show service sdp-using

===============================================================================
SDP Using
===============================================================================
SvcId      SdpId              Type   Far End              Opr   I.Label E.Label
                                                          State
-------------------------------------------------------------------------------
2147483647 32766:4294967294   MS-PW  192.0.2.1            Up    524280  524282
2147483647 32767:4294967295   MS-PW  192.0.2.8            Up    524281  524281
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
```

As it can be seen in the preceding output, two PWs (type MS-PW) have been automatically created over two also automatically created SDPs: 32766 and 32767. SDP 32766 is built over an LDP tunnel whereas SDP 32767 runs over a BGP tunnel.

```
*A:PE-7# show router tunnel-table

===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination           Owner     Encap TunnelId  Pref   Nexthop        Metric
   Color
-------------------------------------------------------------------------------
192.0.2.1/32          sdp       MPLS  32766     5      192.0.2.1      0
192.0.2.1/32          ldp       MPLS  65539     9      192.168.37.1   20
192.0.2.3/32          ldp       MPLS  65538     9      192.168.37.1   10
192.0.2.4/32          ldp       MPLS  65537     9      192.168.47.1   10
192.0.2.8/32          sdp       MPLS  32767     5      192.0.2.8      0
192.0.2.8/32          bgp       MPLS  262145    12     192.168.78.2   1000
-------------------------------------------------------------------------------
Flags: B = BGP or MPLS backup hop available
       L = Loop-Free Alternate (LFA) hop available
       E = Inactive best-external BGP route
       k = RIB-API or Forwarding Policy backup hop
===============================================================================
```

```
*A:PE-7# show service sdp 32766 detail | match "Active LSP"
Mixed LSP Mode       : Enabled               Active LSP Type    : LDP
```

```
*A:PE-7# show service sdp 32767 detail | match "Active LSP"
Mixed LSP Mode       : Enabled               Active LSP Type    : BGP
```

In addition to all of the recommended show commands, **vccv-ping** and **vccv-trace** are two extremely useful commands in this environment. **vccv-trace** can even help to trace the traffic going through the ICBs under failure situations.

Conclusion

Service Providers are always seeking highly scalable VLL services that can be deployed with the lowest operational cost. The SR OS supports MS-PW routing according to the draft-ietf-pwe3-dynamic-ms-pw. MS-PW routing allows the Service Provider to deploy Epipe services without having to provision services in the core of the network. In other words, MS-PW enables end-point provisioning in highly scalable seamless MPLS networks, through the use of BGP. Alternatively, static MS-PW routes or explicit paths can also be used.

The examples used in this chapter illustrate the configuration of MS-PW routing in intra-AS and inter-AS scenarios. Show and OAM commands have also been suggested so that the operator can verify and troubleshoot the MS-PW routing paths and procedures.

Operational Groups for EVPN-VXLAN VPWS Services

This chapter describes the Operational Groups for EVPN-VXLAN VPWS Services.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written based on SR OS Release 16.0.R5, but the CLI in the current edition corresponds to SR OS Release 21.5.R2. EVPN-VXLAN VPWS and service-level operational groups for VPWS services are supported in SR OS Release 16.0.R1, or later.

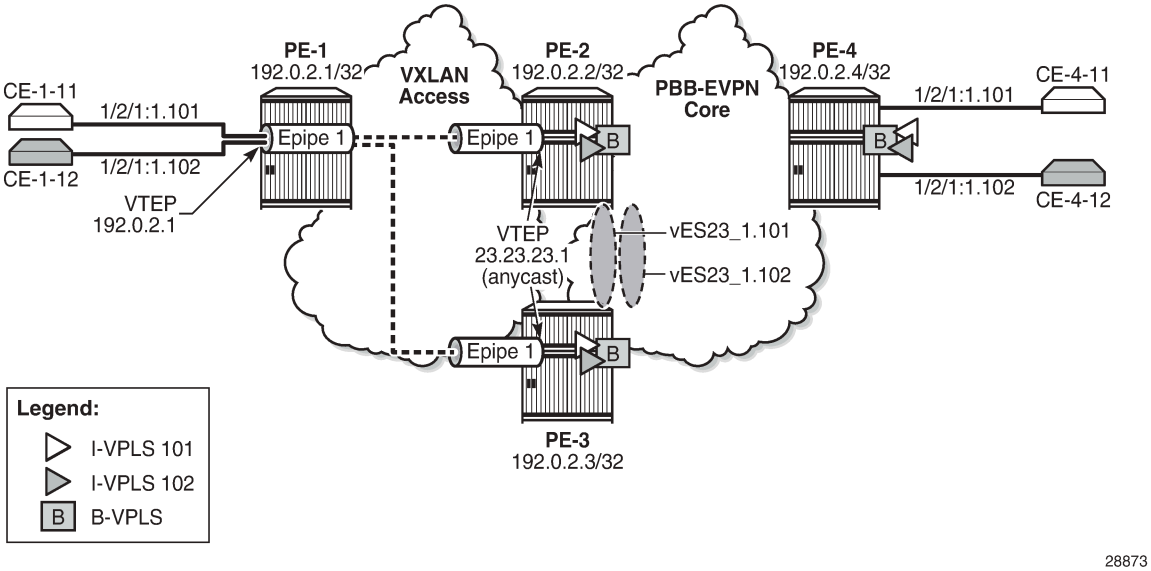
Overview

Operational groups on Epipe services are used for fault propagation to other services, such as I-VPLS or R-VPLS services. Epipes with VXLAN destinations are used in some edge PE applications along with port cross-connect (PXC) so that VXLAN networks can be terminated in other VPLS or VPRN services. In such cases, the operational status of the Epipe services terminating VXLAN must override the operational status of the SAPs of the VPLS or VPRN where the Epipe is stitched to.

Operational group on egress VTEP in Epipes with static VXLAN bindings

The Static VXLAN Termination in Epipe Services chapter describes how Epipes with static VXLAN termination are stitched to I-VPLS services. In Epipes with static VXLAN bindings, operational groups can be configured in the egress VTEP context. Figure 213: Epipe with static VXLAN termination shows the example topology with a static VXLAN tunnel between PE-1 and an anycast address on PE-2 and PE-3. The All-Active Multi-Homing Ethernet Segments (AA MH ESs) "vES23_1.101" and "vES23_1.102" are used by the I-VPLSs 101 and 102, which are both stitched to Epipe 1 in PE-2 and PE-3. The SAPs in these I-VPLSs monitor the operational group configured in the egress VTEP context of the Epipe service, so the SAPs will go operationally down when the operational group of the VTEP goes operationally down.

Figure 213: Epipe with static VXLAN termination



On PE-2 and PE-3, Epipe 1 is configured with static VXLAN bindings, as follows. The egress VTEP is 192.0.2.1, which is the system IP address of PE-1. Operational group "op-grp-1" in configured for this egress VTEP. LAG 2 combines PXC ports and is used to stitch Epipe 1 to the I-VPLS services 101 and 102. For a detailed description of the configuration, see the Static VXLAN Termination in Epipe Services chapter.

```
# on PE-2, PE-3:
configure
    service
        oper-group "op-grp-1" create
        exit
        epipe 1 name "Epipe 1" customer 1 create
            description "Epipe 1 with static VXLAN bindings"
            vxlan-src-vtep 23.23.23.1
            vxlan instance 1 vni 1 create
                egr-vtep 192.0.2.1
                    oper-group "op-grp-1"
                exit
            exit
            sap lag-2:1.* create
                no shutdown
            exit
            no shutdown
        exit
```

For failure propagation to the stitched I-VPLSs, the SAPs in the I-VPLSs can monitor the operational group "op-grp-1", for I-VPLS 101 on PE-2 and PE-3, as follows:

```
# on PE-2, PE-3:
configure
    service
        vpls 101 name "I-VPLS 101" customer 1 i-vpls create
            pbb
                backbone-vpls 100
                exit
            exit
            sap lag-1:1.101 create
                monitor-oper-group "op-grp-1"
                no shutdown
            exit
            no shutdown
        exit
```

When the egress VTEP prefix 192.0.2.1 disappears from the global route-table on PE-2, the VXLAN binding goes down, as follows:

```
*A:PE-2# show service id 1 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address                            Egress VNI          Oper      Vxlan
                                                            State     Type
-------------------------------------------------------------------------------
192.0.2.1                               1                   Down      static
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

When the egress VTEP 192.0.2.1 goes down, the operational group "op-grp-1" goes down too, as follows:

```
*A:PE-2# show service oper-group "op-grp-1"

===============================================================================
Service Oper Group Information
===============================================================================
Oper Group       : op-grp-1
Creation Origin  : manual                           Oper Status: down
Hold DownTime    : 0 secs                           Hold UpTime: 4 secs
Members          : 1                                Monitoring : 2
===============================================================================
```

When the operational group "op-grp-1" goes down, the monitoring SAP in I-VPLS 101 goes operationally down with flag OperGroupDown, as follows:

```
*A:PE-2# show service id 101 sap lag-1:1.101 detail
                                             | match expression "Flags | Oper State"
Admin State        : Up                       Oper State        : Down
Flags              : OperGroupDown
Stp Admin State    : Up                       Stp Oper State    : Down
```

When this SAP goes down, the entire I-VPLS 101 service goes down on PE-2, as follows:

```
*A:PE-2# show service id 101 base | match "State"
Admin State       : Up                  Oper State        : Down
```

Epipes with static VXLAN bindings impose the following restrictions, which cannot be overcome unless a control plane protocol such as BGP-EVPN is used for the VXLAN bindings.

- When anycast VTEPs on the PEs are used, a change in the vES preference on the DF PE triggers a DF switchover for the I-VPLS service. However, the access PE (PE-1 in Figure 213: Epipe with static VXLAN termination) is unaware and keeps sending the VXLAN traffic to the same PE, unless a change in DF comes with an automatic change in the underlay IGP metrics, which cannot be easily accomplished.

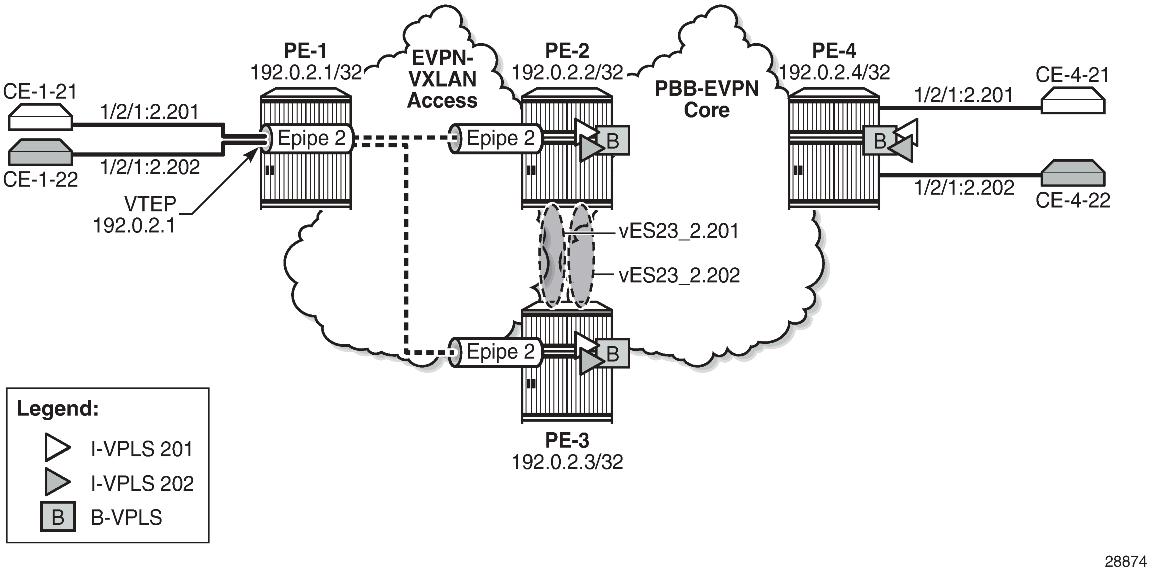
- Without a control plane, Eth-CFM must be used between PEs and access PEs to detect end-to-end service-level failures.
- Traffic from the access PE is forwarded to the anycast VTEP, based on underlay IGP metrics. There is no control on a per-service basis.
- The architecture does not support AA MH for the Epipe service, so the access PEs always send the traffic to one single PE.

The preceding challenges can be addressed by using different VTEPs on the PEs and adding a BGP-EVPN control plane on the Epipe, as described in the next section.

Operational groups in EVPN-VXLAN Epipes

Figure 214: Epipe 2 with EVPN-VXLAN and all-active multi-homing shows EVPN-VXLAN Epipe 2 stitched to I-VPLSs 201 and 202. AA MH ESs "vES23_2.201" and "vES23_2.202" are used by the I-VPLSs 201 and 202 respectively.

Figure 214: Epipe 2 with EVPN-VXLAN and all-active multi-homing



The EVPN-VXLAN VPWS chapter describes the configuration of Epipes with EVPN-VXLAN bindings instead of static VXLAN bindings. The egress VTEP is not configured manually, but dynamically learned through BGP-EVPN. Therefore, the operational group cannot be configured in the egress VTEP context. However, it is possible to configure an operational group in an Epipe at the service level, as follows:

```
# on PE-2:
configure
    service
        oper-group "op-grp-2" create
        exit
        epipe 2 name "Epipe 2" customer 1 create
            description "Epipe 2 with EVPN-VXLAN"
            oper-group "op-grp-2"
            vxlan instance 1 vni 2 create
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-23 create
                    eth-tag 123
                exit
                remote-attachment-circuit AC-1 create
                    eth-tag 101
                exit
                evi 2
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            sap lag-2:2.* create
                no shutdown
            exit
            no shutdown
```

The following shows the error raised when attempting to configure the egress VTEP manually in an Epipe service with BGP-EVPN enabled:

```
*A:PE-2>configure>service>epipe>vxlan# egr-vtep 192.0.2.1
MINOR: SVCMGR #7894 Cannot configure egr-vtep - service has bgp-evpn
```

An operational group can be associated with the entire Epipe or with specific objects, such as SAPs or spoke-SDPs, but not simultaneously. The following error is raised when attempting to associate the operational group "op-grp-2"—that is already associated with the Epipe with the SAP on PE-2:

```
*A:PE-2>config>service>epipe>sap# oper-group "op-grp-2"
MINOR: SVCMGR #1003 Inconsistent value - oper-group already in use as service oper group
```

The service-level operational group status is derived from the service operational status: when Epipe 2 is operationally down, the operational group "op-grp-2" will be down.

For fault propagation to the stitched I-VPLSs 201 and 202, the SAPs in the I-VPLSs monitor the operational group "op-grp-2", for I-VPLS 201 on PE-2 and PE-3, as follows:

```
# on PE-2, PE-3:
configure
    service
        vpls 201 name "I-VPLS 201" customer 1 i-vpls create
            pbb
                backbone-vpls 100
                exit
            exit
            sap lag-1:2.201 create
                monitor-oper-group "op-grp-2"
                no shutdown
            exit
            no shutdown
        exit
```

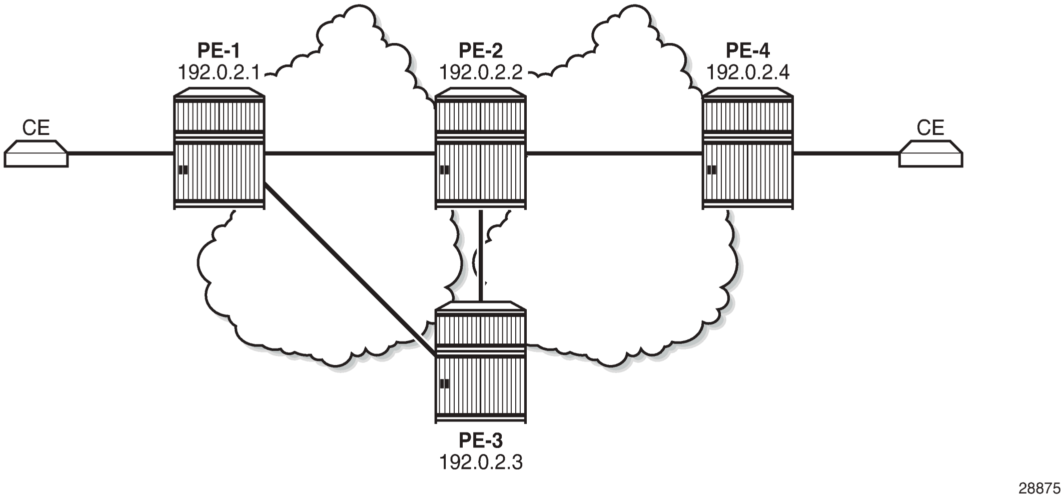
Configuration

In this section, the following use cases are described:

- operational group on egress VTEP in Epipes with static VXLAN bindings stitched to I-VPLSs using AA MH ESs
- service-level operational group in EVPN-VXLAN Epipes stitched to I-VPLSs using AA MH ESs
- service-level operational group in EVPN-VXLAN Epipes stitched to I-VPLSs using Single-Active (SA) MH ESs

Figure 215: Example topology shows the example topology with four PEs in an autonomous system.

Figure 215: Example topology



The initial configuration includes:

- Cards, MDAs, ports
- Router interfaces
- IS-IS as IGP (level capability 2 in the core between PE-2, PE-3, and PE-4; level capability 1 in the access toward PE-1)
- LDP between the core routers PE-2, PE-3, and PE-4

Oper group on egress VTEP in Epipes with static VXLAN bindings stitched to I-VPLSs using AA MH ESs

When static VXLAN bindings are used, no BGP-EVPN is required in the access network to and from PE-1; BGP is only configured in the core network. When PE-2 acts as the route reflector, its BGP configuration is as follows:

```
# on RR PE-2:
configure
    router Base
```

```
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            rapid-update evpn
            group "CORE"
                family evpn
                type internal
                cluster 192.0.2.2
                split-horizon
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
            exit
```

Figure 213: Epipe with static VXLAN termination shows that Epipe 1 is configured in the access network: on PE-1, the VTEP is the system address 192.0.2.1 (default), and on PE-2 and PE-3, the VTEP is a unicast address 23.23.23.1.

On PE-1, Epipe 1 is configured with egress VTEP 23.23.23.1, as follows:

```
# on PE-1:
configure
    service
        epipe 1 name "Epipe 1" customer 1 create
            description "Epipe 1 with static VXLAN bindings"
            vxlan instance 1 vni 1 create
                egr-vtep 23.23.23.1
                exit
            exit
            sap 1/2/1:1.* create
                no shutdown
            exit
            no shutdown
        exit
```

On PE-2 and PE-3, the following unicast address is configured:

```
# on PE-2, PE-3:
configure
    router
        interface "lo23"
            address 23.23.23.0/31
            loopback
        exit
```

On PE-2 and PE-3, three ports are configured as PXC. PXC 1 is used as Forwarding Path Extension (FPE) and the VXLAN tunnel termination 23.23.23.1 is configured with this FPE, as follows:

```
# on PE-2, PE-3:
configure
    service
        system
            vxlan
                tunnel-termination 23.23.23.1 fpe 1 create
            exit
```

PXCs 2 and 3 are used in the internal LAGs that are used to stitch Epipe 1 to I-VPLSs 101 and 102, as follows. LAG 1 will be used in the I-VPLS services; LAG 2 in the Epipe services.

```
# on PE-2, PE-3:
configure
    lag 1
        mode hybrid
        encap-type qinq
        port pxc-2.a
        port pxc-3.a
        no shutdown
    exit
    lag 2
        mode hybrid
        encap-type qinq
        port pxc-2.b
        port pxc-3.b
        no shutdown
    exit
```

On PE-2 and PE-3, Epipe 1 is configured with source VTEP 23.23.23.1 and egress VTEP 192.0.2.1, as follows. The operational group "op-grp-1" is associated with the egress VTEP. The SAP stitches Epipe 1 to the I-VPLSs 101 and 102.

```
# on PE-2, PE-3:
configure
    service
        oper-group "op-grp-1" create
        exit
        epipe 1 name "Epipe 1" customer 1 create
            description "Epipe 1 with static VXLAN bindings"
            vxlan-src-vtep 23.23.23.1
            vxlan instance 1 vni 1 create
                egr-vtep 192.0.2.1
                    oper-group "op-grp-1"
                exit
            exit
            sap lag-2:1.* create
                no shutdown
            exit
            no shutdown
        exit
```

On PE-2, B-VPLS 100 is configured as follows. The configuration is similar on PE-3 and PE-4.

```
# on PE-2:
configure
    service
        vpls 100 name "B-VPLS-100" customer 1 b-vpls create
            service-mtu 1532
            pbb
                source-bmac 00:00:00:00:00:02
                use-es-bmac
            exit
            bgp
            exit
            bgp-evpn
                evi 100
                mpls bgp 1
                    ingress-replication-bum-label
                    auto-bind-tunnel
```

```
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

On PE-2, I-VPLS 101 is configured as follows. The SAP monitors the operational group "op-grp-1" that is configured in Epipe 1. The AA MH ES "vES23_1.101" is used. The configuration of I-VPLS 102 is similar, but it uses AA MH ES "vES23_1.102" with preference value 50 instead. On PE-3, the preference values are reversed: preference value 50 for "vES23_1.101" and 100 for "vES23_1.102".

```
# on PE-2:
configure
    service
        system
            bgp-evpn
                ethernet-segment "vES23_1.101" virtual create
                    esi 01:00:00:00:00:23:00:00:01:11
                    source-bmac-lsb 23-11 es-bmac-table-size 8
                    es-activation-timer 3
                    service-carving
                        mode manual
                        manual
                            preference create
                                value 100
                            exit
                        exit
                    exit
                    multi-homing all-active
                    lag 1
                    qinq
                        s-tag 1 c-tag-range 101
                    exit
                    no shutdown
                exit
            exit
        exit
        vpls 101 name "I-VPLS 101" customer 1 i-vpls create
            pbb
                backbone-vpls 100
                exit
            exit
            sap lag-1:1.101 create
                monitor-oper-group "op-grp-1"
                no shutdown
            exit
            no shutdown
        exit
```

On PE-4, I-VPLS 101 is configured as follows:

```
# on PE-4:
configure
    service
        vpls 101 name "I-VPLS 101" customer 1 i-vpls create
            pbb
                backbone-vpls 100
                exit
            exit
            sap 1/2/1:1.101 create
```

```
                no shutdown
            exit
            no shutdown
        exit
```

To emulate a failure that affects the operational state of the egress VTEP (and also of the Epipe service), the SAP of Epipe 1 on PE-2 is disabled, as follows:

```
# on PE-2:
configure
    service
        epipe "Epipe 1"
            sap lag-2:1.*
                shutdown
```

When the SAP is operationally down, Epipe 1 goes down, as follows:

```
*A:PE-2# show service id 1 base | match "State"
Admin State       : Up                  Oper State        : Down
```

The egress VTEP 192.0.2.1 is operationally down, as follows:

```
*A:PE-2# show service id 1 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address                            Egress VNI          Oper      Vxlan
                                                            State     Type
-------------------------------------------------------------------------------
192.0.2.1                               1                   Down      static
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

The operational group "op-grp-1" is associated with the egress VTEP, so it goes operationally down, as follows:

```
*A:PE-2# show service oper-group "op-grp-1"

===============================================================================
Service Oper Group Information
===============================================================================
Oper Group       : op-grp-1
Creation Origin  : manual                            Oper Status: down
Hold DownTime    : 0 secs                            Hold UpTime: 4 secs
Members          : 1                                 Monitoring : 2
===============================================================================
```

This operational group is monitored by the SAPs in I-VPLSs 101 and 102, so these SAPs go down with flag OperGroupDown; for example, for I-VPLS 101 on PE-2:

```
*A:PE-2# show service id 101 sap lag-1:1.101 | match "Flags"
Flags              : OperGroupDown
```

When the SAP goes down, the I-VPLS service goes down, as follows:

```
*A:PE-2# show service id 101 base | match "State"
Admin State       : Up                  Oper State        : Down
```

Even though Epipe 1 on PE-2 is operationally down while Epipe 1 on PE-3 is up, PE-1 is unaware because the VXLAN destination in Epipe 1 remains up, as follows:

```
*A:PE-1# show service id 1 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address                            Egress VNI          Oper      Vxlan
                                                            State     Type
-------------------------------------------------------------------------------
23.23.23.1                              1                   Up        static
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

With ECMP=1, all traffic from PE-1 is directed to PE-2, regardless of the state of the Epipe on PE-2. The following route table on PE-1 shows that destination prefix 23.23.23.0/31 has next-hop 192.168.12.2, which is an interface address on PE-2.

```
*A:PE-1# show router route-table 23.23.23.0/31

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
23.23.23.0/31                                 Remote  ISIS      00h06m41s  15
       192.168.12.2                                                 10
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

Traffic from the CEs attached to PE-1 is forwarded by PE-1 to PE-2, where it is dropped.

Service-level operational group in EVPN-VXLAN Epipes stitched to I-VPLS using AA MH ESs

BGP must be enabled on all nodes for the EVPN address family, also in the access to and from PE-1. The BGP configuration on PE-1 is as follows:

```
# on PE-1:
configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            rapid-update evpn
            group "ACCESS"
                family evpn
                type internal
                split-horizon
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
```

On PE-1, the following EVPN-VXLAN Epipe 2 is configured with local Ethernet tag 101 and remote Ethernet tag 123.

```
# on PE-1:
configure
    service
        epipe 2 name "Epipe 2" customer 1 create
            description "Epipe 2 with EVPN-VXLAN"
            vxlan instance 1 vni 2 create
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-1 create
                    eth-tag 101
                exit
                remote-attachment-circuit AC-23 create
                    eth-tag 123
                exit
                evi 2
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            sap 1/2/1:2.* create
                no shutdown
            exit
            no shutdown
        exit
```

On PE-2, the following EVPN-VXLAN Epipe 2 is configured with local Ethernet tag 123 and remote Ethernet tag 101. The operational group "op-grp-2" is associated with Epipe 2. The configuration on PE-3 is identical.

```
# on PE-2:
configure
    service
        oper-group "op-grp-2" create
        exit
        epipe 2 name "Epipe 2" customer 1 create
            description "Epipe 2 with EVPN-VXLAN"
            oper-group "op-grp-2"
            vxlan instance 1 vni 2 create
            exit
            bgp
            exit
            bgp-evpn
                local-attachment-circuit AC-23 create
```

```
                    eth-tag 123
                exit
                remote-attachment-circuit AC-1 create
                    eth-tag 101
                exit
                evi 2
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            sap lag-2:2.* create
                no shutdown
            exit
            no shutdown
        exit
```

The configuration of B-VPLS 100 remains unchanged and the configuration of the I-VPLSs 201 and 202 resembles the configuration of VPLSs 101.

When there is no failure, the egress VTEP for Epipe 2 on PE-1 is 192.0.2.2, which is the system IP address of PE-2, as follows:

```
*A:PE-1# show service id 2 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address                            Egress VNI          Oper      Vxlan
                                                            State     Type
-------------------------------------------------------------------------------
192.0.2.2                               2                   Up        evpn
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

To emulate a failure that affects the operational state of the Epipe service, the SAP in Epipe 2 is disabled, as follows:

```
# on PE-2:
configure
    service
        epipe "Epipe 2"
            sap lag-2:2.*
                shutdown
```

When the SAP goes down, the Epipe goes down, as follows:

```
*A:PE-2# show service id 2 base | match "State"
Admin State       : Up                  Oper State        : Down
```

On PE-1, the egress VTEP for Epipe 2 is 192.0.2.3, which is the system IP address of PE-3, as follows:

```
*A:PE-1# show service id 2 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address                            Egress VNI          Oper      Vxlan
```

```
                                                            State     Type
-------------------------------------------------------------------------------
192.0.2.3                               2                   Up        evpn
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

The operational group "op-grp-2" follows the state of Epipe 2, so it goes down, as follows. As a consequence, the monitoring SAPs for this operational group also go down.

```
*A:PE-2# show service oper-group "op-grp-2" detail

===============================================================================
Service Oper Group Information
===============================================================================
Oper Group       : op-grp-2
Creation Origin  : manual                            Oper Status: down
Hold DownTime    : 0 secs                            Hold UpTime: 4 secs
Members          : 1                                 Monitoring : 2
===============================================================================

===============================================================================
Member Services for OperGroup: op-grp-2
===============================================================================
Svc Id
-------------------------------------------------------------------------------
2
-------------------------------------------------------------------------------
Service Entries found: 1
===============================================================================

===============================================================================
Monitoring SAPs for OperGroup: op-grp-2
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
lag-1:2.201                     201        1     none    1     none   Up   Down
lag-1:2.202                     202        1     none    1     none   Up   Down
-------------------------------------------------------------------------------
SAP Entries found: 2
===============================================================================
```

The SAPs in I-VPLSs 201 and 202 go down with the OperGroupDown flag, as follows:

```
*A:PE-2# show service id 201 sap lag-1:2.201 detail | match "Flags" context all
Flags              : OperGroupDown
*A:PE-2# show service id 202 sap lag-1:2.202 detail | match "Flags" context all
Flags              : OperGroupDown
```

When the SAPs go down, the I-VPLSs 201 and 202 also go down, as follows:

```
*A:PE-2# show service id 201 base | match "State"
Admin State       : Up                    Oper State        : Down
*A:PE-2# show service id 202 base | match "State"
Admin State       : Up                    Oper State        : Down
```

Even with this failure on PE-2, traffic can still flow between the CEs, as follows:

```
*A:PE-4# ping router 21 172.16.21.11 rapid
PING 172.16.21.11 56 data bytes
!!!!!
---- 172.16.21.11 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 3.51ms, avg = 4.84ms, max = 9.30ms, stddev = 2.24ms.
```

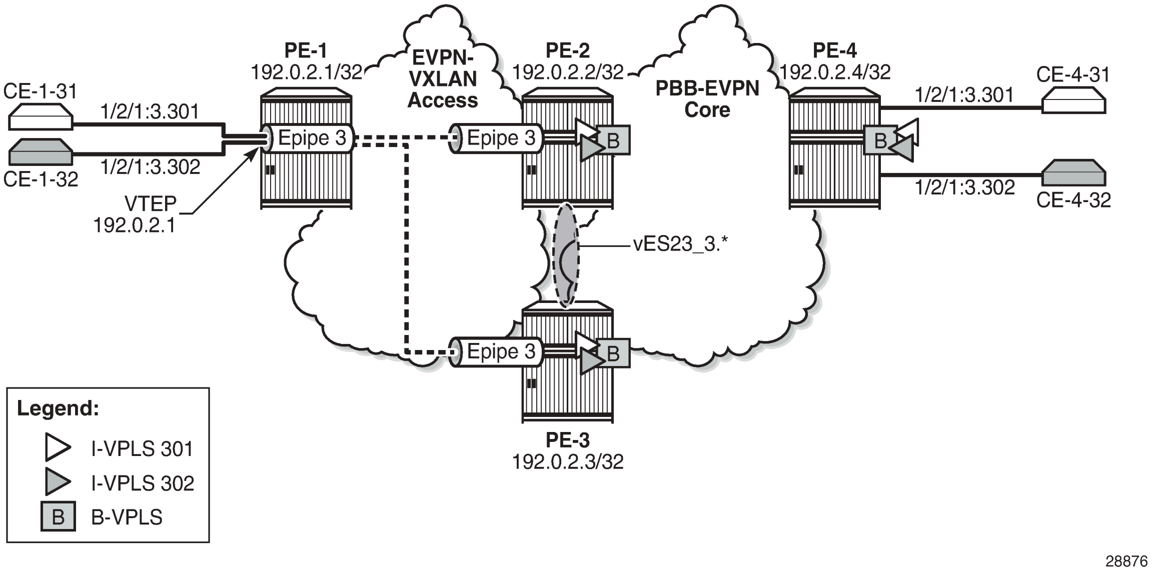
The following FDB for I-VPLS 201 on PE-4 shows that MAC address 00:ca:fe:00:21:11 of CE-1-21 is reachable via AA MH ES with ES-BMAC 00:00:00:00:23:21:

```
*A:PE-4# show service id 201 fdb detail

===============================================================================
Forwarding Database, Service 201
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
201        00:ca:fe:00:21:11 eES-BMAC:               L/90     07/15/21 13:50:28
                             00:00:00:00:23:21
201        00:ca:fe:00:21:41 sap:1/2/1:2.201         L/90     07/15/21 13:56:09
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

Service-level operational group in EVPN-VXLAN Epipes stitched to I-VPLS using SA MH ESs

Figure 216: Epipe 3 with EVPN-VXLAN and SA MH ES shows the example topology with an SA MH ES used by the I-VPLSs.

Figure 216: Epipe 3 with EVPN-VXLAN and SA MH ES



The configuration of Epipe 3 resembles the configuration of Epipe 2: the same Ethernet tags are used, only the VNI, EVI, and SAPs are different.

On PE-2 and PE-3, PXC 4 is configured to stitch Epipe 3 to I-VPLSs 301 and 302. The PXC port will be used in the SA MH ES.

On PE-2 and PE-3, Epipe 3 is configured as follows:

```
# on PE-2, PE-3:
configure
    service
        oper-group "op-grp-3" create
        exit
        epipe 3 name "Epipe 3" customer 1 create
            description "EVPN-VXLAN Epipe 3"
            oper-group "op-grp-3"
            vxlan instance 1 vni 3 create
            exit
            bgp-evpn
                local-attachment-circuit AC-23 create
                    eth-tag 123
                exit
                remote-attachment-circuit AC-1 create
                    eth-tag 101
                exit
                evi 3
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            sap pxc-4.b:3.* create
                no shutdown
            exit
            no shutdown
        exit
```

On PE-2, I-VPLS 301 uses SA MH ES "vES23_3.*", and is configured as follows.

```
# on PE-2:
configure
    service
        system
            bgp-evpn
                ethernet-segment "vES23_3.*" virtual create
                    esi 01:00:00:00:00:23:00:00:03:01
                    source-bmac-lsb 23-34 es-bmac-table-size 8
                    es-activation-timer 3
                    service-carving
                        mode manual
                        manual
                            preference create
                                value 100
                            exit
                        exit
                    exit
                    multi-homing single-active
                    port pxc-4.a
                    qinq
                        s-tag-range 3
                    exit
                    no shutdown
                exit
            exit
        exit
        vpls 301 name "I-VPLS 301" customer 1 i-vpls create
            pbb
                backbone-vpls 100
                exit
            exit
            stp
                shutdown
            exit
            sap pxc-4.a:3.301 create
                monitor-oper-group "op-grp-3"
                no shutdown
            exit
            no shutdown
        exit
```

The configuration is similar on PE-3, but with source-bmac-lsb 23-35 and preference 50.

When Epipe 3 on PE-2 is operationally up, the egress VTEP for Epipe 3 on PE-1 is 192.0.2.2, as follows:

```
*A:PE-1# show service id 3 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address                            Egress VNI          Oper      Vxlan
                                                            State     Type
-------------------------------------------------------------------------------
192.0.2.2                               3                   Up        evpn
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

To emulate a failure on PE-2 that affects the operational state of the Epipe service, the SAP in Epipe 3 is disabled, as follows:

```
# on PE-2:
configure
    service
        epipe "Epipe 3"
            sap pxc-4.b:3.*
                shutdown
```

When the SAP goes down, Epipe 3 goes down on PE-2, as follows:

```
*A:PE-2# show service id 3 base | match "State"
Admin State       : Up                  Oper State        : Down
```

When Epipe 3 on PE-2 goes operationally down, the egress VTEP for Epipe 3 on PE-1 is 192.0.2.3, as follows:

```
*A:PE-1# show service id 3 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address                            Egress VNI          Oper      Vxlan
                                                            State     Type
-------------------------------------------------------------------------------
192.0.2.3                               3                   Up        evpn
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

The operational group "op-grp-3" follows the state of Epipe 3 on PE-2, so it goes down. Also, the monitoring SAPs for this operational group go down.

```
*A:PE-2# show service oper-group "op-grp-3" detail

===============================================================================
Service Oper Group Information
===============================================================================
Oper Group       : op-grp-3
Creation Origin  : manual                         Oper Status: down
Hold DownTime    : 0 secs                         Hold UpTime: 4 secs
Members          : 1                              Monitoring : 2
===============================================================================

===============================================================================
Member Services for OperGroup: op-grp-3
===============================================================================
Svc Id
-------------------------------------------------------------------------------
3
-------------------------------------------------------------------------------
Service Entries found: 1
===============================================================================

===============================================================================
Monitoring SAPs for OperGroup: op-grp-3
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
pxc-4.a:3.301                   301        1     none    1     none   Up   Down
pxc-4.a:3.302                   302        1     none    1     none   Up   Down
-------------------------------------------------------------------------------
SAP Entries found: 2
===============================================================================
```

The SAPs in I-VPLSs 301 and 302 on PE-2 go down with the OperGroupDown flag, as follows:

```
*A:PE-2# show service id 301 sap pxc-4.a:3.301 | match "Flags" context all
Flags              : StandByForMHProtocol
                     OperGroupDown
*A:PE-2# show service id 302 sap pxc-4.a:3.302 | match "Flags" context all
Flags              : StandByForMHProtocol
                     OperGroupDown
```

When the SAPs go down, the I-VPLSs go down on PE-2, as follows:

```
*A:PE-2# show service id 301 base | match "State"
Admin State       : Up                   Oper State        : Down
*A:PE-2# show service id 302 base | match "State"
Admin State       : Up                   Oper State        : Down
```

When the initial DF PE-2 goes down for the I-VPLSs 301 and 302, PE-3 becomes the new DF. The connectivity between the CEs is preserved, as follows:

```
*A:PE-4# ping router 31 172.16.31.11 rapid
PING 172.16.31.11 56 data bytes
!!!!!
---- 172.16.31.11 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 3.68ms, avg = 4.15ms, max = 4.40ms, stddev = 0.264ms
```

The following FDB for I-VPLS 301 on PE-4 shows that the frames toward MAC address 00:ca:fe:00:31:11 of CE-1-31 are sent via PE-3 (192.0.2.3):

```
*A:PE-4# show service id 301 fdb detail

===============================================================================
Forwarding Database, Service 301
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
301        00:ca:fe:00:31:11 b-mpls:                 L/120    07/15/21 14:16:25
                             192.0.2.3:524279
           ldp:65537
301        00:ca:fe:00:31:41 sap:1/2/1:3.301         L/120    07/15/21 14:12:33
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

PE-3 is now the DF for I-VPLS 301, as follows:

```
*A:PE-3# show service id 301 ethernet-segment
```

```
===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
pxc-4.a:3.301         vES23_3.*                        DF
===============================================================================
No sdp entries
No vxlan instance entries
```

Conclusion

Some service providers use VXLAN as a next-generation access technology used between the MSANs (or access PEs) and core PE routers. EVPN-VXLAN Epipes can be stitched using PXC to other services, such as I-VPLS. Operational groups can be defined in the Epipe for fault propagation to the SAPs of the services where the Epipe is stitched to.

Operational Groups in EVPN Services

This chapter provides information about Operational Groups in EVPN Services.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

The information and configuration in this chapter are based on SR OS Release 21.10.R1. EVPN operational groups are supported in EVPN-VXLAN and EVPN-MPLS VPLS and R-VPLS services in SR OS Release 19.10.R2 and later; in EVPN-MPLS Epipes in SR OS Release 19.5.R1 and later.

Overview

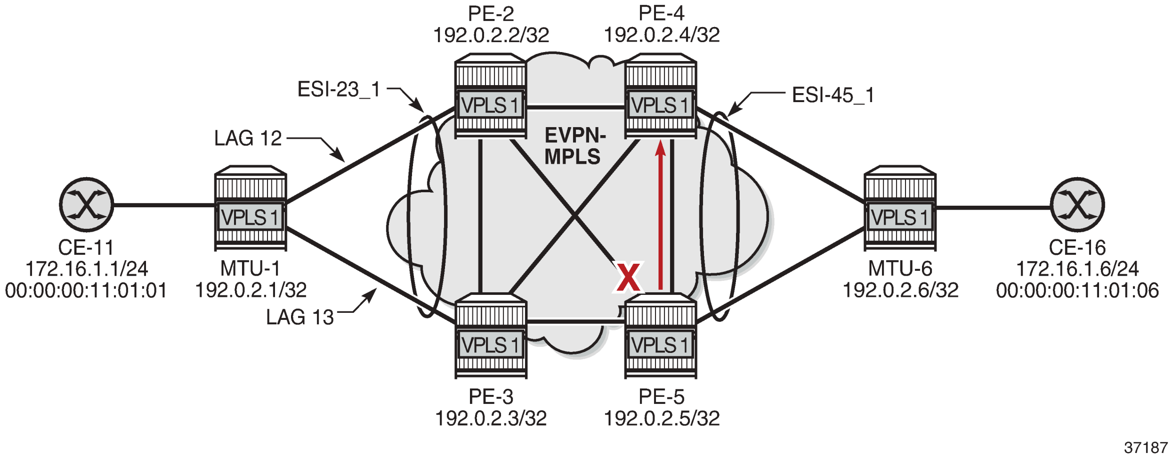
An operational group includes objects and drives the status of service endpoints (such as pseudowires, SAPs, IP interfaces) located in the same or in different service instances. The operational group status is derived from the status of the individual components. Other service objects can monitor the operational group status. The status of the operational group influences the status of the monitoring objects.

If the operational group goes down, the monitoring objects are also brought operationally down. When one of the objects included in the operational group comes up, the entire operational group comes up, as well as the monitoring objects.

Operational groups for EVPN destinations

Figure 217: EVPN mesh going down triggers DF switchover from PE-5 to PE-4 shows a sample topology with VPLS 1 configured on all nodes. PE-4 and PE-5 share a single-active Ethernet Segment (ES) "ESI-45_1" where PE-5 is the Designated Forwarder (DF).

Figure 217: EVPN mesh going down triggers DF switchover from PE-5 to PE-4



When the EVPN-VPLS service becomes isolated from the rest of the EVPN network (for example, all EVPN destinations are removed on DF PE-5), an operational group for EVPN destinations is required to trigger a DF switchover and bring the monitoring access SAP (or spoke SDP) down. EVPN single-active multi-homing PEs that are elected as NDF must notify their attached access nodes to prevent these from sending traffic to the NDF. Ethernet Connectivity Fault Management (ETH-CFM) is enabled on a down Maintenance Endpoint (MEP) configured on the SAP to detect SAP failure. After the remote MEP on MTU-6 detects the failure, MTU-6 redirects its traffic to PE-4. This avoids blackholes when PE-5 is disconnected from the EVPN core.

On PE-5, VPLS 1 is configured with operational group "vpls-1_45" in EVPN-MPLS and SAP 1/1/2:1 monitoring this operational group. The operational group configured under a BGP-EVPN instance cannot be configured under any other object, such as SAPs or SDP-bindings.

```
# on PE-5:
configure
    service
        oper-group "vpls-1_45" create
            hold-time
                group down 0
                group up 0
            exit
        exit
        vpls 1 name "VPLS 1" customer 1 create
            bgp
            exit
            bgp-evpn
                cfm-mac-advertisement
                evi 1
                mpls bgp 1
                    oper-group "vpls-1_45"
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/1/2:1 create
                description "to MTU-6"
                eth-cfm
                    mep 56 domain 1 association 11 direction down
                        fault-propagation-enable suspend-ccm
                        ccm-enable
                        mac-address 00:00:00:00:56:05
                        no shutdown
                    exit
                exit
                monitor-oper-group "vpls-1_45"
                no shutdown
            exit
            no shutdown
        exit
```

Using operational groups in the EVPN service, it is possible to monitor if the PE is isolated and, if it is, trigger a Designated Forwarder switchover. The operational group associated to the EVPN-MPLS instance goes down in the following cases:

- bgp-evpn mpls is disabled (shutdown)
- VPLS is disabled (shutdown)
- all EVPN destinations associated to the instance are removed, for example, when:
    - no tunnels are available for auto-bind-tunnel resolution
    - the network ports facing the EVPN ports are down
    - the BGP sessions to the route reflector or PEs are down

Operational groups for Ethernet Segments (Port-active multi-homing)

Operational groups can be configured on single-active ESs that need to function as port-active multi-homing Ethernet Segments. 'Port-active' refers to a special single-active mode where the PE is DF or non-DF for all the services attached to the ES. The configuration of a port-active ES is as follows:

```
# on PE-2:
configure
    service
        oper-group "vpls-1_23" create
            hold-time
                group down 0
                group up 0
            exit
        exit
        system
            bgp-evpn
                ethernet-segment "ESI-23_1" create
                    esi 01:23:00:00:00:00:01:00:00:00
                    es-activation-timer 3
                    service-carving
                        mode manual
                        manual
                            preference create
                                value 150                # on PE-3: value 100
                            exit
                        exit
                    exit
                    multi-homing single-active
                    lag 12                               # on PE-3: lag 13
                    oper-group "vpls-1_23"
                    no shutdown
                exit
```

This ES operational group can be monitored on the LAG:

```
# on PE-2:
configure
    lag 12 name "lag-12"
        description "to MTU-1"
        mode access
        encap-type dot1q
        monitor-oper-group "vpls-1_23"
        port 1/1/2
        lacp active administrative-key 1 system-id 00:00:00:01:02:01
                                         system-priority 1
        standby-signaling lacp                           # default
        no shutdown
    exit
```

When the operational group is configured on the ES and monitored on the associated LAG:

- The status of the ES operational group is driven by the ES DF status.
  - When a node becomes NDF, the ES operational group goes down and all the SAPs in the ES go down.
- The ES operational group goes down when all the SAPs in the ES go down.
  - When all SAPs in the ES go down, the operational group goes down and the node becomes NDF.

The monitoring LAG goes down when the ES operational group is down. The LAG signals the LAG standby state to the access node. The LAG standby signaling can be configured as **lacp** or **power-off**.

```
*A:PE-2>config>lag# standby-signaling
  - no standby-signaling
  - standby-signaling {lacp|power-off}
```

- **standby-signaling lacp** signals LACP out-of-sync to the CE when the application layer instructs the LAG to become standby
- **standby-signaling power-off** brings the LAG members down, and hence the access SAPs down

The ES and AD routes for the ES are not withdrawn because the router recognizes that the LAG becomes standby due to the ES operational group.

Some restrictions:

- Multi-chassis LAG and ES are mutually exclusive:

  ```
  *A:PE-3>config>redundancy>mc>peer>mc-lag# lag 13
  MINOR: LAGMGR #1321 lag associated with ethernet segment
  ```

- LAG sub-groups are blocked:

  ```
  *A:PE-3>config>lag# port 1/1/1 sub-group 2
  MINOR: CLI Could not set subgroup for port "1/1/1".
  MINOR: LAGMGR #1031 Port settings incompatible - invalid combination port sub-group <->
   monitor-oper-group
  ```

- Only LAGs in access mode can monitor operational groups:

```
*A:PE-3>config>lag$ monitor-oper-group "vpls-1_23"
MINOR: LAGMGR #1031 Port settings incompatible - monitor-oper-group not allowed when lag is
 not access
```

- Operational groups cannot be assigned to virtual ESs:

```
*A:PE-3>config>service>system>bgp-evpn>eth-seg# oper-group "vpls-1_23"
MINOR: SVCMGR #8050 Ethernet segment config cannot be modified - oper-group not supported
 with virtual ethernet-segments
```

- Operational groups cannot be assigned to all-active ESs:

```
*A:PE-3>config>service>system>bgp-evpn>eth-seg$ oper-group "vpls-1_23"
MINOR: SVCMGR #8050 Ethernet segment config cannot be modified - oper-group not supported
 with all-active ethernet-segment
```

- Operational groups cannot be assigned to ESs with service-carving auto:

```
*A:PE-3>config>service>system>bgp-evpn>eth-seg$ oper-group "vpls-1_23"
MINOR: SVCMGR #8050 Ethernet segment config cannot be modified - oper-group not supported on
 ethernet-segments with service carving auto
```

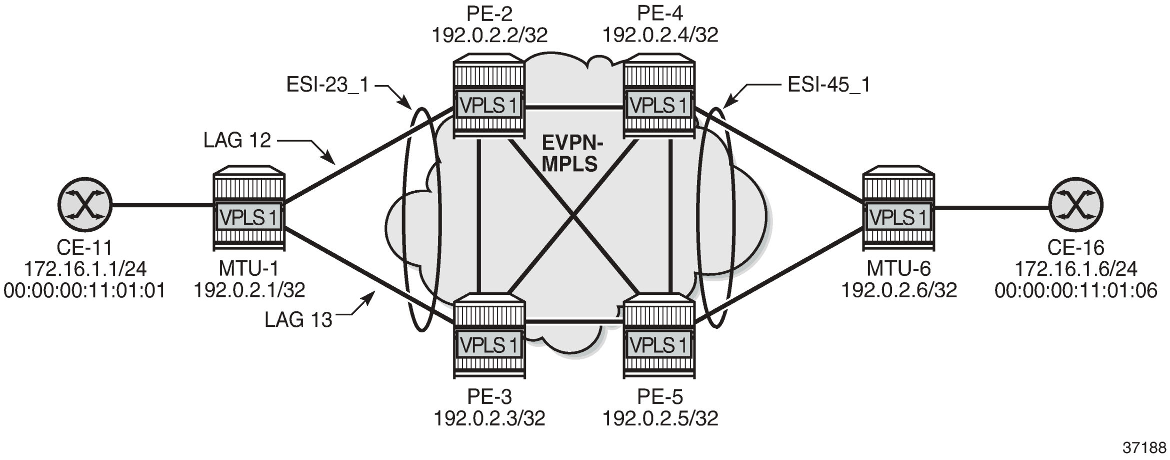
Link Loss Forwarding in EVPN-VPWS

Fault propagation in EVPN-VPWS services is supported using ETH-CFM. However, not all access nodes support ETH-CFM and, in that case, LAG **standby-signaling lacp** or **power-off** can be used instead.

Configuration

Figure 218: Sample topology with VPLS 1 shows the sample topology with VPLS 1 configured on all nodes.

Figure 218: Sample topology with VPLS 1



The initial configuration includes:

- Cards, MDAs, ports
- LAG 12 between PE-1 and PE-2; LAG 13 between PE-1 and PE-3
- Router interfaces between PE-2, PE-3, PE-4, and PE-5
- IS-IS on all router interfaces
- LDP between PE-2, PE-3, PE-4, and PE-5
- BGP between PE-2, PE-3, PE-4, and PE-5

For BGP, PE-2 acts as route reflector and the configuration is as follows:

```
# on PE-2:
configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family evpn
                cluster 192.0.2.2
                peer-as 64500
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
        exit
```

Operational groups for EVPN destinations

On PE-4, single-active ES "ESI-45_1" is configured with service carving auto. Operational group "vpls-1_45" is associated with EVPN-MPLS in VPLS 1 and SAP 1/1/1:1 is monitoring that operational group. ETH-CFM is enabled on a down MEP configured on the SAP to detect SAP failures. The service configuration is as follows:

```
# on PE-4:
configure
    service
        oper-group "vpls-1_45" create
            hold-time
                group down 0
                group up 0
            exit
        exit
        system
            bgp-evpn
                ethernet-segment "ESI-45_1" create
                    esi 01:45:00:00:00:00:01:00:00:00
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    port 1/1/1
                    no shutdown
                exit
            exit
        exit
        vpls 1 name "VPLS 1" customer 1 create
            bgp
            exit
            bgp-evpn
                cfm-mac-advertisement
                evi 1
                mpls bgp 1
                    oper-group "vpls-1_45"
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/1/1:1 create
                description "to MTU-6"
                eth-cfm
                    mep 46 domain 1 association 10 direction down
                        ccm-enable
                        mac-address 00:00:00:00:46:04
                        no shutdown
                    exit
                exit
                monitor-oper-group "vpls-1_45"
                no shutdown
            exit
            no shutdown
        exit
```

The configuration on PE-5 is similar.

On MTU-6, VPLS 1 is configured with three SAPs: SAP 1/1/2:1 toward PE-4, SAP 1/1/1:1 toward PE-5, and SAP 1/2/1:1 toward CE-16. ETH-CFM MEPs are configured on SAP 1/1/1:1 and SAP 1/1/2:1. The service configuration is as follows:

```
# on MTU-6:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            stp
                shutdown
            exit
            sap 1/1/1:1 create
                description "to PE-5"
                eth-cfm
                    mep 65 domain 1 association 11
                        ccm-enable
                        mac-address 00:00:00:00:65:06
                        no shutdown
                    exit
                exit
                no shutdown
            exit
            sap 1/1/2:1 create
                description "to PE-4"
                eth-cfm
                    mep 64 domain 1 association 10
                        ccm-enable
                        mac-address 00:00:00:00:64:06
                        no shutdown
                    exit
                exit
                no shutdown
            exit
            sap 1/2/1:1 create
                description "to CE-16"
                no shutdown
            exit
            no shutdown
        exit
```

Initial situation without failure

On MTU-6, ETH-CFM MEP 65 receives Continuity Check (CC) messages from its remote peer 56 on PE-5:

```
*A:MTU-6# show eth-cfm mep 65 domain 1 association 11 all-remote-mepids

=============================================================================
Eth-CFM Remote-Mep Table
=============================================================================
R-mepId AD Rx CC RxRdi Port-Tlv If-Tlv Peer Mac Addr     CCM status since
-----------------------------------------------------------------------------
56         True  False Absent   Absent 00:00:00:00:56:05 12/23/2021 13:51:58
=============================================================================
Entries marked with a 'T' under the 'AD' column have been auto-discovered.
```

The following command shows that PE-5 is DF for VPLS 1:

```
*A:PE-5# show service id 1 ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
1/1/2:1               ESI-45_1                         DF
===============================================================================
No sdp entries
No vxlan instance entries
```

PE-5 has full mesh with all EVPN destinations in VPLS 1:

```
*A:PE-5# show service id 1 evpn-mpls

===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address                    Egr Label     Num.   Mcast Last Change
                               Transport:Tnl MACs         Sup BCast Domain
-------------------------------------------------------------------------------
192.0.2.2                      524282        0      bum   12/23/2021 13:51:47
                               ldp:65539                  No
192.0.2.3                      524282        0      bum   12/23/2021 13:51:47
                               ldp:65537                  No
192.0.2.4                      524282        0      bum   12/23/2021 13:51:47
                               ldp:65538                  No
-------------------------------------------------------------------------------
Number of entries : 3
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                       Num. Macs               Last Change
-------------------------------------------------------------------------------
01:23:00:00:00:00:01:00:00:00   1                       12/23/2021 13:52:28
-------------------------------------------------------------------------------
Number of entries: 1
-------------------------------------------------------------------------------
===============================================================================
```

Avoiding blackholes when EVPN destinations are removed

On PE-5, a failure is simulated by disabling LDP:

```
# on PE-5:
configure
    router Base
        ldp
            shutdown
```

With LDP disabled, PE-5 has no tunnels available for auto-bind-tunnel in VPLS 1 and all EVPN destinations are removed, as follows:

```
*A:PE-5# show service id 1 evpn-mpls

===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address                    Egr Label     Num.   Mcast Last Change
                               Transport:Tnl MACs         Sup BCast Domain
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================

===============================================================================
BGP EVPN-MPLS Ethernet Segment Dest
===============================================================================
Eth SegId                       Num. Macs               Last Change
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

Log 99 on PE-5 shows that the operational group "vpls-45_1" goes down and PE-5 becomes NDF in "ESI-45_1":

```
73 2021/12/23 13:53:52.244 UTC MINOR: SVCMGR #2094 Base
"Ethernet Segment:ESI-45_1, EVI:1, Designated Forwarding state changed to:false"

72 2021/12/23 13:53:52.243 UTC MINOR: SVCMGR #2542 Base
"Oper-group vpls-1_45 changed status to down"
```

The following command on PE-5 shows that the operational status of oper-group "vpls-45_1" is down, the EVPN-MPLS destinations are down, and the monitoring SAP 1/1/2:1 is down:

```
*A:PE-5# show service oper-group "vpls-1_45" detail

===============================================================================
Service Oper Group Information
===============================================================================
Oper Group       : vpls-1_45
Creation Origin  : manual                          Oper Status: down
Hold DownTime    : 0 secs                          Hold UpTime: 0 secs
Members          : 1                               Monitoring : 1
===============================================================================

===============================================================================
Member BGP-EVPN for OperGroup: vpls-1_45
===============================================================================
SvcId:Instance (Type)                   Status
-------------------------------------------------------------------------------
1:1 (mpls)                              Inactive
-------------------------------------------------------------------------------
BGP-EVPN Entries found: 1
===============================================================================

===============================================================================
Monitoring SAPs for OperGroup: vpls-1_45
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
```

```
1/1/2:1                        1          1     none    1     none   Up   Down
-------------------------------------------------------------------------------
SAP Entries found: 1
===============================================================================
```

The following command shows that SAP 1/1/2:1 is operationally down with flags StandByForMHProtocol and OperGroupDown:

```
*A:PE-5# show service id 1 sap 1/1/2:1

===============================================================================
Service Access Points(SAP)
===============================================================================
Service Id         : 1
SAP                : 1/1/2:1                  Encap             : q-tag
Description        : to MTU-6
Admin State        : Up                       Oper State        : Down
Flags              : StandByForMHProtocol
                     OperGroupDown
Multi Svc Site     : None
Last Status Change : 12/23/2021 13:53:52
Last Mgmt Change   : 12/23/2021 13:51:45
===============================================================================
```

With ETH-CFM enabled, log 99 on MTU-6 shows that local MEP 65 did not receive a Continuity Check Message (CCM) from the remote MEP:

```
58 2021/12/23 13:53:56.388 UTC MINOR: ETH_CFM #2001 Base
"MEP 1/11/65 highest defect is now defRemoteCCM"
```

PE-4 receives the following BGP-EVPN withdrawal messages:

```
27 2021/12/23 13:53:52.225 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 129
    Flag: 0x90 Type: 15 Len: 125 Multiprotocol Unreachable NLRI:
        Address Family EVPN
        Type: EVPN-ETH-SEG Len: 23 RD: 192.0.2.5:0
              ESI: 01:45:00:00:00:00:01:00:00:00, IP-Len: 4 Orig-IP-Addr: 192.0.2.5
        Type: EVPN-AD Len: 25 RD: 192.0.2.5:1 ESI: 01:45:00:00:00:00:01:00:00:00,
              tag: 0 Label: 0 (Raw Label: 0x0) PathId:
        Type: EVPN-MAC Len: 33 RD: 192.0.2.5:1 ESI: ESI-0, tag: 0, mac len: 48
              mac: 00:00:00:11:01:06, IP len: 0, IP: NULL, label1: 0
        Type: EVPN-MAC Len: 33 RD: 192.0.2.5:1 ESI: ESI-0, tag: 0, mac len: 48
              mac: 00:00:00:00:65:06, IP len: 0, IP: NULL, label1: 0
"
```

The following command on PE-4 shows that PE-4 is the DF and the only DF candidate in "ESI-45_1" for VPLS 1:

```
*A:PE-4# show service system bgp-evpn ethernet-segment name "ESI-45_1" evi 1

===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
1             1             0                   yes 12/23/2021 13:53:55
```

```
===============================================================================

===============================================================================
DF Candidates                          Time Added           Oper Pref  Do Not
                                                              Value    Preempt
-------------------------------------------------------------------------------
192.0.2.4                              12/23/2021 13:51:38  0          Disabl*
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

Finally, the failure is restored by re-enabling LDP on PE-5:

```
# on PE-5:
configure
    router Base
        ldp
            no shutdown
```

Operational groups for ES (Port-Active Multi-Homing)

On PE-2 and PE-3, operational group vpls-1_23 is configured and associated with ES " ESI-23_1", but not configured or monitored in VPLS 1. The service configuration on PE-3 is as follows:

```
# on PE-3:
configure
    service
        oper-group "vpls-1_23" create
            hold-time
                group down 0
                group up 0
            exit
        exit
        system
            bgp-evpn
                ethernet-segment "ESI-23_1" create
                    esi 01:23:00:00:00:00:01:00:00:00
                    es-activation-timer 3
                    service-carving
                        mode manual
                        manual
                            preference create
                                value 100                # on PE-2: value 150
                            exit
                        exit
                    exit
                    multi-homing single-active
                    ac-df-capability exclude
                    lag 13
                    oper-group "vpls-1_23"
                    no shutdown
                exit
            exit
        exit
        vpls 1 name "VPLS 1" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap lag-13:1 create                          # on PE-2: lag-12:1
                description "to MTU-1"
                no shutdown
            exit
            no shutdown
        exit
```

LAG 12 on PE-2 and LAG 13 on PE-3 monitor operational group "vpls-1_23". The **monitor-oper-group** command can be added to the LAG without the need to disable (shutdown) the LAG:

```
# on PE-3:
configure
    lag 13 name "lag-13"
        description "to MTU-1"
        mode access
        encap-type dot1q
        monitor-oper-group "vpls-1_23"
        port 1/1/1
        lacp active administrative-key 1 system-id 00:00:00:01:03:01
                                         system-priority 1
        standby-signaling lacp                     # default
        no shutdown
```

Note:
In this example, MTU-1 is connected to PE-2 and PE-3 through two different LAGs, however, this port-active multi-homing mode also supports the use of a single LAG on MTU-1. If a single LAG was used on MTU-1, the LAG ports on PE-2 and PE-3 must be configured with the same LACP parameters (administrative-key, system-id and system-priority) to ensure that PE-2 and PE-3 show themselves as a single system to MTU-1.

EVPN single-active multi-homing PEs that are elected as NDF must notify their attached access nodes to prevent these from sending traffic to the NDF. In this port-active multi-homing mode, ETH-CFM is not used, and other notification mechanisms are needed, such as LAG standby signaling (**lacp** or **power-off**). When the EVPN application layer instructs the LAG to become standby as a result of the NDF status, the behavior is as follows:

- the **lacp** option signals LACP out-of-sync to MTU-1
- the power-off option brings down the LAG ports connected to MTU-1

MTU-1 is connected to PE-2 and PE-3 using two different access LAGs with encapsulation dot1q and at least one port in each LAG. Any encapsulation type is supported in the LAGs. The LAG configuration is as follows:

```
# on MTU-1:
configure
    lag 12 name "lag-12"
        description "to PE-2"
        mode access
        encap-type dot1q
        port 1/1/1
        lacp active administrative-key 32768
        no shutdown
    exit
    lag 13 name "lag-13"
        description "to PE-3"
        mode access
        encap-type dot1q
        port 1/1/2
        lacp active administrative-key 32769
        no shutdown
    exit
```

On MTU-1, VPLS 1 is configured as follows:

```
# on MTU-1:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            stp
                shutdown
            exit
            sap 1/2/1:1 create
                description "to CE-11"
                no shutdown
            exit
            sap lag-12:1 create
                description "to PE-2"
                no shutdown
            exit
            sap lag-13:1 create
                description "to PE-3"
                no shutdown
            exit
            no shutdown
        exit
```

Initial situation without failures

PE-2 is DF for VPLS 1:

```
*A:PE-2# show service id 1 ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                           Status
-------------------------------------------------------------------------------
lag-12:1              ESI-23_1                          DF
===============================================================================
No sdp entries
No vxlan instance entries
```

```
*A:PE-3# show service id 1 ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                           Status
```

```
-------------------------------------------------------------------------------
lag-13:1              ESI-23_1                         NDF
===============================================================================
No sdp entries
No vxlan instance entries
```

On NDF PE-3, operational group "vpls-1_23" is operationally down, which has an impact on the operational status of the monitoring LAG, as follows:

```
*A:PE-3# show service oper-group "vpls-1_23" detail

===============================================================================
Service Oper Group Information
===============================================================================
Oper Group       : vpls-1_23
Creation Origin  : manual                          Oper Status: down
Hold DownTime    : 0 secs                          Hold UpTime: 0 secs
Members          : 1                               Monitoring : 1
===============================================================================

===============================================================================
Member Ethernet-Segment for OperGroup: vpls-1_23
===============================================================================
Ethernet-Segment                        Status
-------------------------------------------------------------------------------
ESI-23_1                                Inactive
-------------------------------------------------------------------------------
Ethernet-Segment Entries found: 1
===============================================================================

===============================================================================
Monitoring LAG for OperGroup: vpls-1_23
===============================================================================
Lag-id         Adm       Opr       Weighted  Threshold Up-Count  Act/Stdby
    name
-------------------------------------------------------------------------------
13             up        down      No        0         0         N/A
    lag-13
-------------------------------------------------------------------------------
LAG Entries found: 1
===============================================================================
```

The following command shows that SAP lag-13:1 is operationally down on PE-3 with flags PortOperDown and StandByForMHProtocol:

```
*A:PE-3# show service id 1 sap lag-13:1

===============================================================================
Service Access Points(SAP)
===============================================================================
Service Id         : 1
SAP                : lag-13:1                Encap             : q-tag
Description        : to MTU-1
Admin State        : Up                      Oper State        : Down
Flags              : PortOperDown StandByForMHProtocol
Multi Svc Site     : None
Last Status Change : 12/23/2021 13:46:51
Last Mgmt Change   : 12/23/2021 13:51:29
===============================================================================
```

The following command on PE-3 shows that LAG 13 has LACP standby signaling enabled to the MTU-1. LAG 13 is operationally down because the operational group is down.

```
*A:PE-3# show lag 13 detail

===============================================================================
LAG Details
===============================================================================
Description        : N/A
-------------------------------------------------------------------------------
Details
-------------------------------------------------------------------------------
Lag-id              : 13                    Mode                 : access
Lag-name            : lag-13
Adm                 : up                    Opr                  : down
---snip---

Standby Signaling   : lacp
---snip---

Monitor oper group  : vpls-1_23
Oper group status   : down
Adaptive loadbal.   : disabled              Tolerance            : N/A

-------------------------------------------------------------------------------
Port-id        Adm     Act/Stdby Opr     Primary   Sub-group     Forced  Prio
-------------------------------------------------------------------------------
1/1/1          up      active    down    yes       1             -       32768

-------------------------------------------------------------------------------
Port-id        Role      Exp   Def   Dist  Col   Syn   Aggr  Timeout  Activity
-------------------------------------------------------------------------------
1/1/1          actor     No    No    No    No    No    Yes   Yes      Yes
1/1/1          partner   No    No    No    No    Yes   Yes   Yes      Yes
===============================================================================
```

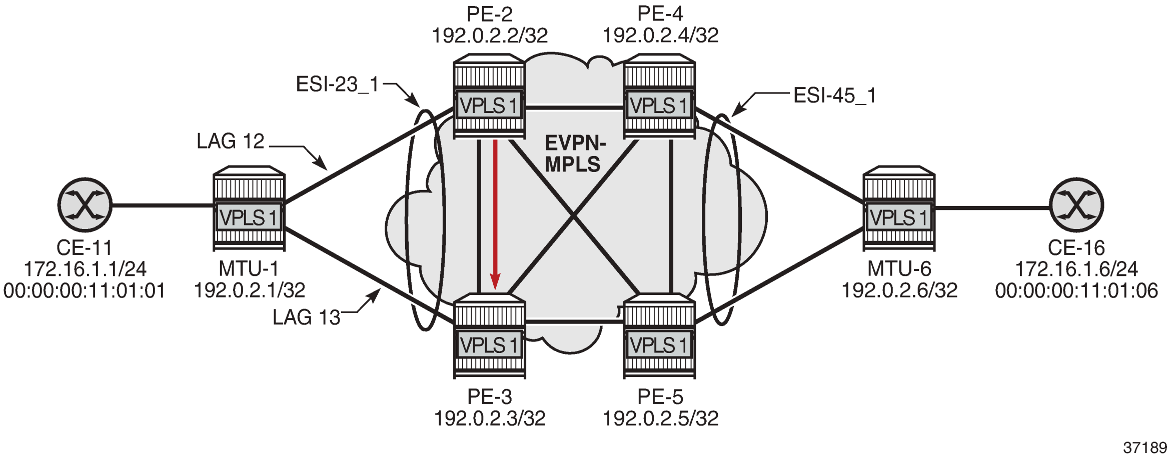
DF switchover

To trigger a DF switchover, the preference value is modified on PE-2, as follows:

```
# on PE-2:
configure
    service
        system
            bgp-evpn
                ethernet-segment "ESI-23_1"  create
                    service-carving
                        manual
                            preference create
                                value 50
                            exit
```

Figure 219: DF switchover in single-active ESI-23_1 shows a DF switchover from PE-2 to PE-3. PE-2 becomes the NDF and LAG 12 is in standby.

Figure 219: DF switchover in single-active ESI-23_1



Log 99 on PE-2 shows that SAP lag-12:1 goes down, the ES operational group goes down, the monitoring LAG 12 goes down, port 1/1/2 goes down, and subsequently an LACP out-of-sync message is sent:

```
93 2021/12/23 14:04:40.062 UTC WARNING: LAG #2007 Base LAG
"LAG lag-12 : partner oper state bits changed on member 1/1/2 : [sync FALSE -> TRUE] [expired
 TRUE -> FALSE] [defaulted TRUE -> FALSE]"

92 2021/12/23 14:04:40.062 UTC WARNING: LAG #2007 Base LAG
"LAG lag-12 : LACP RX state machine entered current state on member 1/1/2"

91 2021/12/23 14:04:40.058 UTC MAJOR: SVCMGR #2210 Base
"Processing of an access port state change event is finished and the status of all affected
 SAPs on port lag-12 has been updated."

90 2021/12/23 14:04:40.058 UTC WARNING: SNMP #2004 Base lag-12
"Interface lag-12 is not operational"

89 2021/12/23 14:04:40.058 UTC WARNING: SNMP #2004 Base 1/1/2
"Interface 1/1/2 is not operational"

88 2021/12/23 14:04:40.058 UTC MINOR: SVCMGR #2203 Base
"Status of SAP lag-12:1 in service 1 (customer 1) changed to admin=up oper=down flags=Mh
Standby"

87 2021/12/23 14:04:40.058 UTC WARNING: LAG #2006 Base LAG
"LAG lag-12 : initializing LACP, all members will be brought down"

86 2021/12/23 14:04:40.058 UTC MINOR: SVCMGR #2094 Base
"Ethernet Segment:ESI-23_1, EVI:1, Designated Forwarding state changed to:false"

85 2021/12/23 14:04:40.058 UTC MINOR: SVCMGR #2542 Base
"Oper-group vpls-1_23 changed status to down"
```

On PE-3, log 99 shows that PE-3 becomes DF for "ESI-23_1" and operational group "vpls-1_23", interface 1/1/1, and LAG 13 are operationally up.

```
107 2021/12/23 14:04:43.313 UTC WARNING: LAG #2007 Base LAG
"LAG lag-13 : partner oper state bits changed on member 1/1/1 :  [collecting FALSE -> TRUE]"

106 2021/12/23 14:04:43.306 UTC MAJOR: SVCMGR #2210 Base
```

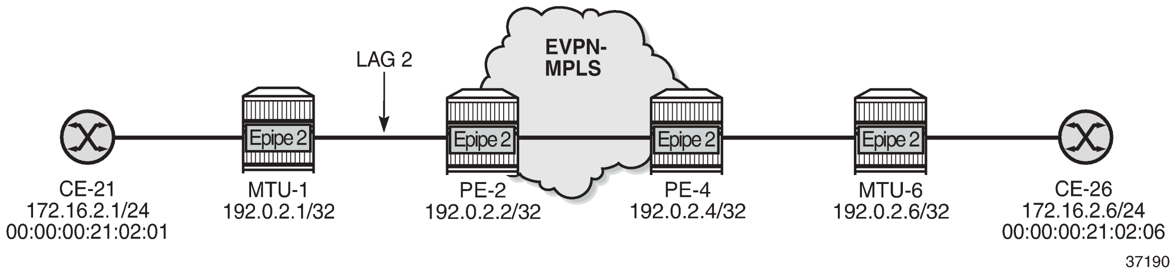
```
"Processing of an access port state change event is finished and the status of all affected
 SAPs on port lag-13 has been updated."

105 2021/12/23 14:04:43.305 UTC WARNING: SNMP #2005 Base lag-13
"Interface lag-13 is operational"

104 2021/12/23 14:04:43.305 UTC WARNING: SNMP #2005 Base 1/1/1
"Interface 1/1/1 is operational"

103 2021/12/23 14:04:43.105 UTC MAJOR: SVCMGR #2210 Base
"Processing of an access port state change event is finished and the status of all affected
 SAPs on port lag-13 has been updated."

102 2021/12/23 14:04:43.085 UTC MINOR: SVCMGR #2094 Base
"Ethernet Segment:ESI-23_1, EVI:1, Designated Forwarding state changed to:true"

101 2021/12/23 14:04:43.085 UTC MINOR: SVCMGR #2542 Base
"Oper-group vpls-1_23 changed status to up"
```

Link Loss Forwarding in EVPN-VPWS

Fault propagation in EVPN-VPWS services is supported using ETH-CFM, but also using LAG **standby-signaling lacp** or **power-off**.

Figure 220: Sample topology with Epipe 2 shows the sample topology with Epipe 2.

Figure 220: Sample topology with Epipe 2



The configuration on MTU-1 is as follows:

```
# on MTU-1:
configure
    lag 2
        mode access
        encap-type dot1q
        port 1/1/5
        lacp passive administrative-key 32769
        no shutdown
    exit
    service
        epipe 2 name "Epipe 2" customer 1 create
            sap 1/2/1:2 create
                no shutdown
            exit
            sap lag-2:2 create
                no shutdown
            exit
            no shutdown
        exit
```

On PE-2, operational group "llf-1" is configured and associated to EVPN-MPLS. LAG 2 monitors this operational group.

```
# on PE-2:
configure
    service
        oper-group "llf-1" create
            hold-time
                group down 0
                group up 0
            exit
        exit
    exit
    lag 2
        mode access
        encap-type dot1q
        monitor-oper-group "llf-1"
        port 1/1/5
        lacp active administrative-key 2 system-id 00:00:00:00:12:01
                                         system-priority 1
        standby-signaling lacp                          # default
        no shutdown
    exit
    service
        epipe 2 name "Epipe 2" customer 1 create
            bgp
            exit
            bgp-evpn
                local-attachment-circuit "ac-1_2" create
                    eth-tag 12
                exit
                remote-attachment-circuit "ac-6_2" create
                    eth-tag 62
                exit
                evi 2
                mpls bgp 1
                    oper-group "llf-1"
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            sap lag-2:2 create
                no shutdown
            exit
            no shutdown
        exit
```

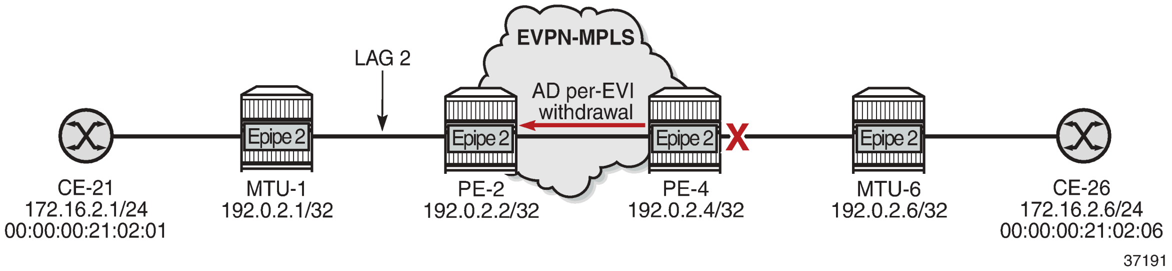
The configuration on PE-4 is as follows:

```
# on PE-4:
configure
    service
        epipe 2 name "Epipe 2" customer 1 create
            bgp
            exit
            bgp-evpn
                local-attachment-circuit "ac-6_2" create
                    eth-tag 62
```

```
                 exit
                 remote-attachment-circuit "ac-1_2" create
                     eth-tag 12
                 exit
                 evi 2
                 mpls bgp 1
                     auto-bind-tunnel
                         resolution any
                     exit
                     no shutdown
                 exit
             exit
             sap 1/1/5:2 create
                 no shutdown
             exit
             no shutdown
         exit
```

Figure 221: LLF in Epipe 2 - PE-4 failure shows when a failure occurs on PE-4.

Figure 221: LLF in Epipe 2 - PE-4 failure



The failure is simulated on PE-4 by disabling port 1/1/5 toward MTU-6.

```
# on PE-4:
configure
    port 1/1/5
        shutdown
```

When the link between PE-4 and MTU-6 fails, PE-4 withdraws the AD per-EVI route for Epipe 2. PE-2 receives the following AD per-EVI withdrawal from PE-4:

```
159 2021/12/23 14:17:25.843 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 34
    Flag: 0x90 Type: 15 Len: 30 Multiprotocol Unreachable NLRI:
        Address Family EVPN
        Type: EVPN-AD Len: 25 RD: 192.0.2.4:2 ESI: ESI-0, tag: 62
                      Label: 0 (Raw Label: 0x0) PathId:
"
```

Upon receiving this AD per-EVI route, Epipe 2 goes operationally down on PE-2:

```
*A:PE-2# show service id 2 base | match "Oper State"
Admin State       : Up                  Oper State        : Down
```

Operational group "llf-1" goes down when the Epipe is operationally down:

```
*A:PE-2# show lag 2 detail | match "per group"
Monitor oper group  : llf-1
Oper group status   : down
```

On PE-2, the detailed information for operational group "llf-1" shows that the operational group and the monitoring LAG are down.

```
*A:PE-2# show service oper-group "llf-1" detail

===============================================================================
Service Oper Group Information
===============================================================================
Oper Group       : llf-1
Creation Origin  : manual                          Oper Status: down
Hold DownTime    : 0 secs                          Hold UpTime: 0 secs
Members          : 1                               Monitoring : 1
===============================================================================

===============================================================================
Member BGP-EVPN for OperGroup: llf-1
===============================================================================
SvcId:Instance (Type)                   Status
-------------------------------------------------------------------------------
2:1 (mpls)                              Inactive
-------------------------------------------------------------------------------
BGP-EVPN Entries found: 1
===============================================================================

===============================================================================
Monitoring LAG for OperGroup: llf-1
===============================================================================
Lag-id         Adm       Opr       Weighted  Threshold Up-Count  Act/Stdby
    name
-------------------------------------------------------------------------------
2              up        down      No        0         0         N/A
    lag-2
-------------------------------------------------------------------------------
LAG Entries found: 1
===============================================================================
```

PE-2 signals the fault based on the configuration of the LAG standby signaling:

- If the LAG standby signaling is power-off, PE-2 brings down the ports in the LAG.
- If the LACP standby signaling is configured, PE-2 signals an LACP out-of-sync on the LAG ports.

In either case, MTU-1 stops forwarding traffic to PE-2.

The following debug message in log 99 on MTU-1 shows that MTU-1 received an LACP out-of-sync message for port 1/1/5 of LAG 2:

```
181 2021/12/23 14:17:25.845 UTC WARNING: LAG #2007 Base LAG
"LAG lag-2 : partner oper state bits changed on member 1/1/5 : [sync TRUE -> FALSE] [collecting
 TRUE -> FALSE]"
```

The following debug messages in log 99 on MTU-1 show that LAG 2 and interface 1/1/5 are not operational:

```
183 2021/12/23 14:17:25.845 UTC WARNING: SNMP #2004 Base lag-2
```

```
"Interface lag-2 is not operational"

182 2021/12/23 14:17:25.845 UTC WARNING: SNMP #2004 Base 1/1/5
"Interface 1/1/5 is not operational"
```

On MTU-1, LAG 2 is operationally down:

```
*A:MTU-1# show lag 2
===============================================================================
Lag Data
===============================================================================
Lag-id         Adm     Opr     Weighted Threshold Up-Count MC Act/Stdby
    name
-------------------------------------------------------------------------------
2              up      down    No       0         0        N/A
    lag-2
===============================================================================
```

Conclusion

Operational groups can be useful in EVPN services to avoid blackholes when a PE is disconnected from the EVPN core. Failures can be propagated by the PEs to access nodes, either by ETH-CFM or LAG standby signaling.

P2MP mLDP FEC Resolution for BGP-LU in EVPN

This chapter provides information about P2MP mLDP FEC Resolution for BGP-LU in EVPN.

Topics in this chapter include:

Applicability

This chapter was initially written for SR OS Release 16.0.R3, but the CLI in the current edition is based on SR OS Release 21.5.R1. Recursive and non-recursive multicast Label Distribution Protocol (mLDP) Forwarding Equivalence Class (FEC) resolution for BGP Labeled Unicast (BGP-LU) is supported in SR OS Release 15.0.R1 or later; see the P2MP mLDP Inter-AS Model C for EVPN-MPLS Services chapter.

In SR OS Release 15.0.R4, and later, a leaf node in an MVPN can generate non-recursive mLDP mapping messages even if the root IP address is resolved using BGP-LU, without the need to leak BGP routes to IGP and LDP. In SR OS Release 16.0.R1, this is also supported for EVPN-MPLS services.

Overview

In inter-AS and intra-AS scenarios, recursive and non-recursive FEC label mapping messages can be used to set up the mLDP tree. In the P2MP mLDP Inter-AS Model C for EVPN-MPLS Services chapter, recursive and non-recursive mLDP FEC resolution is documented for inter-AS model C.

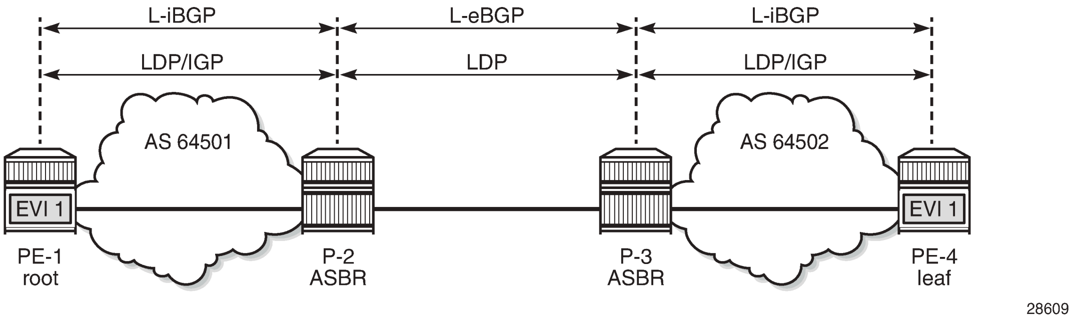
This chapter describes the following use cases for recursive and non-recursive mLDP FEC resolution for BGP-LU:

- P2MP mLDP FEC resolution for inter-AS model C
- P2MP mLDP FEC resolution for seamless MPLS

Some routers do not support recursive mLDP FEC, so basic non-recursive mLDP FEC is used instead. The non-recursive mLDP FEC resolution does not require the root IP address to be leaked from BGP to IGP and LDP. This is different from the configuration in the P2MP mLDP Inter-AS Model C for EVPN-MPLS Services chapter.

Figure 222: Example topology for inter-AS model C shows the example topology for inter-AS model C with the configured protocols (IGP, LDP, BGP). Root node PE-1 is situated in AS 64501 and leaf node PE-4 in AS 64502. P-2 and P-3 are AS Border Routers (ASBRs) that are configured with next-hop-self (NHS). VPLS 1 is configured on root node PE-1 and leaf node PE-4, and is EVPN-MPLS enabled. The example topology for seamless MPLS is similar, but P-2 and P-3 will then act as Area Border Routers (ABRs) and IGP instance 0 is configured between them.

Figure 222: Example topology for inter-AS model C

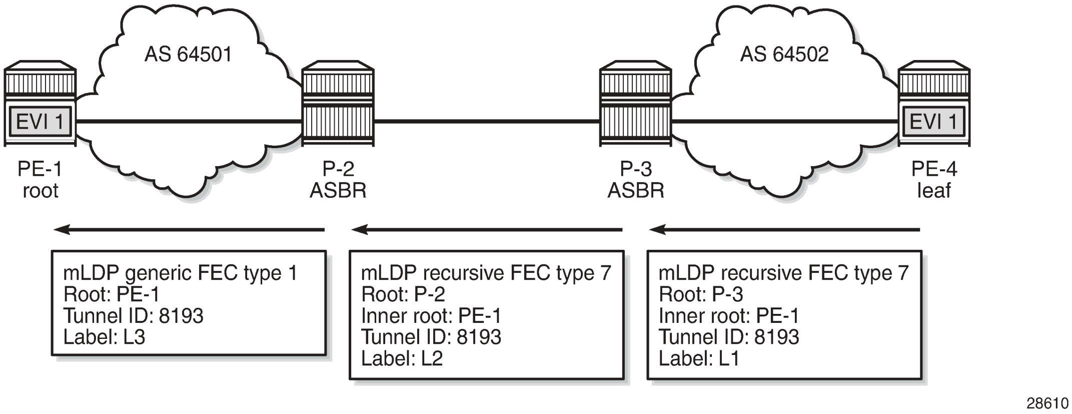


Recursive mLDP FEC resolution requires the nodes in a remote AS (or remote area in case of seamless MPLS) to support GRT recursive FEC type 7 to join the root node.

- PE-4 has a labeled BGP route to root PE-1 with next-hop P-3 in its route table. If PE-4 supports it, it sends a GRT recursive FEC type 7 label mapping message with inner root PE-1 and root P-3.
- P-3 has a labeled BGP route to PE-1 with next-hop P-2. When P-3 receives the mLDP label mapping message from PE-4, it generates its own GRT recursive FEC type 7 message with inner root PE-1 and root P-2.
- P-2 has an IGP route to root PE-1. When P-2 receives the mLDP label mapping message from P-3, it generates a non-recursive FEC type 1 message with root PE-1.

Figure 223: mLDP FEC label mapping messages for inter-AS model C shows the mLDP label mapping messages for inter-AS model C.

Figure 223: mLDP FEC label mapping messages for inter-AS model C



However, if the leaf node PE-4 does not support GRT recursive FEC type 7, it is possible to generate a non-recursive FEC type 1 label mapping message with root PE-1 to the local ASBR that supports GRT recursive FEC type 7. The following command generates only generic FEC type 1 label mapping messages with PE-1 as the root, on the leaf node PE-4:

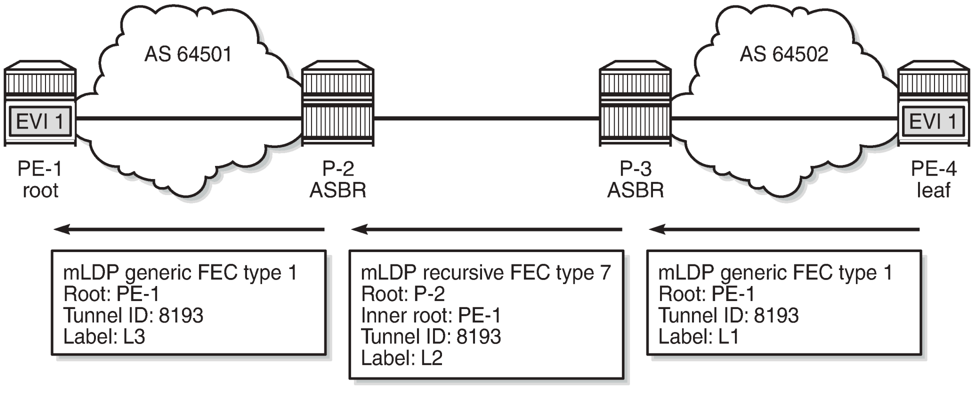
```
# on PE-4:
configure
    router Base
        ldp
```

```
            generate-basic-fec-only
```

Note:
SR OS always generates a recursive FEC if the root node is resolved via BGP; if the root node is resolved via IGP, basic FEC is generated instead. The only way to not generate a recursive FEC when the root is resolved via BGP is by configuring the **generate-basic-fec-only** command.

Figure 224: Non-recursive mLDP FEC for inter-AS model C shows the non-recursive mLDP label mapping messages for inter-AS model C.

Figure 224: Non-recursive mLDP FEC for inter-AS model C

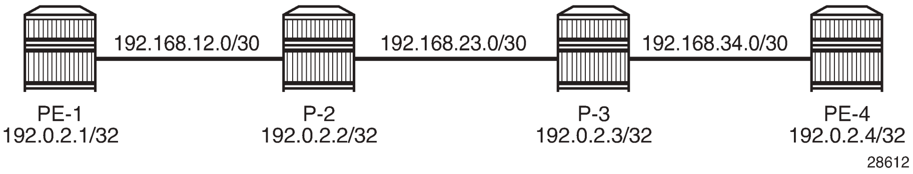


It is also possible that the ASBR routers do not support GRT recursive FEC either. The same **generate-basic-fec-only** command can be configured on all these nodes, which will then generate basic FEC type 1 label mapping messages with root address 192.0.2.1 to the next-hop.

Configuration

Figure 225: Example topology shows the example topology with four nodes.

Figure 225: Example topology



The initial configuration includes the following:

- Cards, MDAs, ports
- Router interfaces

Inter-AS model C

Figure 222: Example topology for inter-AS model C showed the example topology for inter-AS model C. The following is configured for that topology. For a detailed explanation of the configuration, see the P2MP mLDP Inter-AS Model C for EVPN-MPLS Services chapter.

- Within each AS, OSPF is configured as IGP (alternatively, IS-IS can be used).
- LDP is enabled within each AS.
- LDP is enabled between the ASBRs using the interface IP addresses 192.168.23.x.
- On the ASBRs, a static route 192.168.23.y/32 for the interface IP address on the ASBR peer is configured (with mask /32 instead of /30). When a label mapping message is received for an LDP FEC prefix, the next-hop for a FEC prefix is resolved using the routing table. The FEC is installed in the Label Information Base (LIB) if the next-hop matches a /32 route entry.
- BGP is configured on all nodes for the labeled IPv4 address family. An export policy exports the system IP addresses of the root and leaf nodes PE-1 and PE-4.
- A multi-hop BGP session is established between PE-1 and PE-4 for the EVPN address family, allowing inclusive multicast EVPN routes to be exchanged.
- EVPN-MPLS VPLS 1 is configured on PE-1 and PE-4 with mLDP enabled. PE-1 is configured as root node.

The BGP configuration on PE-1 is as follows. The BGP configuration on PE-4 is similar, but with different neighbors and AS numbers. The export policy is identical.

```
# on PE-1:
configure
    router Base
        policy-options
            begin
            prefix-list "sysPE"
                prefix 192.0.2.0/24 prefix-length-range 32-32
            exit
            policy-statement "PE-sys-to-labeled-BGP"
                entry 10
                    from
                        protocol direct
                        prefix-list "sysPE"
                    exit
                    to
                        protocol bgp-label
                    exit
                    action accept
                    exit
                exit
            exit
            commit
        exit
        bgp
            split-horizon
            group "eBGP"
                family evpn
                type external
                multihop 10
                local-as 64501
                peer-as 64502
                neighbor 192.0.2.4
                exit
            exit
            group "iBGP"
                type internal
                neighbor 192.0.2.2
                    family label-ipv4
                    export "PE-sys-to-labeled-BGP"
                exit
            exit
            no shutdown
        exit
```

On PE-1, VPLS 1 is configured as follows. The service configuration on PE-4 is similar, but with different RT values and without the **root-and-leaf** parameter.

```
# on PE-1:
configure
    service
        vpls 1 name "EVI-1" customer 1 create
            bgp
                route-target export target:64501:1 import target:64502:1
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    ingress-replication-bum-label
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            provider-tunnel
                inclusive
                    owner bgp-evpn-mpls
                    root-and-leaf        # PE-1 is configured as root node
                    mldp
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/2/1:1 create
                no shutdown
            exit
            no shutdown
        exit
```

On P-2, the following static route with mask /32 is configured for the interface IP address of the peer ASBR. The configuration on P-3 is similar.

```
# on P-2:
configure
    router Base
        static-route-entry 192.168.23.2/32
            next-hop 192.168.23.2
                no shutdown
            exit
        exit
```

On P-2, the LDP and BGP configuration is as follows. The configuration on P-3 is similar.

```
# on P-2:
configure
    router Base
        ldp
            interface-parameters
                interface "int-P-2-PE-1" dual-stack
                    ipv4
                        no shutdown
                    exit
                    no shutdown
                exit
                interface "int-P-2-P-3" dual-stack
                    ipv4
                        local-lsr-id interface
                        no shutdown
                    exit
                    no shutdown
                exit
            exit
        exit
        bgp
            split-horizon
            group "eBGP"
                type external
                neighbor 192.168.23.2
                    family label-ipv4
                    next-hop-self
                    local-as 64501
                    peer-as 64502
                    advertise-inactive
                exit
            exit
            group "iBGP"
                type internal
                neighbor 192.0.2.1
                    family label-ipv4
                    cluster 192.0.2.2
                exit
            exit
            no shutdown
        exit
```

Leaf node PE-4 has a labeled BGP route toward root node PE-1 using next-hop 192.0.2.3, as follows:

```
*A:PE-4# show router route-table

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age         Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                  Remote  BGP_LABEL 00h00m08s  170
       192.0.2.3 (tunneled)                                         10
192.0.2.3/32                                  Remote  OSPF      00h02m33s  10
       192.168.34.1                                                 10
192.0.2.4/32                                  Local   Local     00h02m34s  0
       system                                                       0
192.168.34.0/30                               Local   Local     00h02m34s  0
       int-PE-4-P-3                                                 0
-------------------------------------------------------------------------------
```

```
No. of Routes: 4
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

Likewise, ASBR P-3 has a labeled BGP route toward root node PE-1 using next-hop 192.168.23.1, as follows:

```
*A:P-3# show router route-table 192.0.2.1/32

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                  Remote  BGP_LABEL 00h01m35s  170
       192.168.23.1                                                 0
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

P-2 has an IGP route toward root node PE-1, as follows:

```
*A:P-2# show router route-table 192.0.2.1/32

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                  Remote  OSPF      00h03m49s  10
       192.168.12.1                                                 10
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

Recursive mLDP FEC resolution for inter-AS model C

With the preceding configuration, leaf node PE-4 sends a recursive mLDP FEC label mapping message with PE-1 as inner root and P-3 as root. On PE-4, the number of GRT recursive mLDP bindings is 1, as follows:

```
*A:PE-4# show router ldp bindings active p2mp summary ipv4
 No. of Generic IPv4 P2MP Active Bindings: 0
 No. of In-Band-SSM IPv4 P2MP Active Bindings: 0
 No. of In-Band-VPN-SSM IPv4 P2MP Active Bindings: 0
 No. of In-Band-SSM IPv4 P2MP Active Bindings: 0
 No. of VPN Recursive with Generic IPv4 P2MP Active Bindings: 0
 No. of GRT Recursive with Generic IPv4 P2MP Active Bindings: 1
```

```
*A:PE-4# show router ldp bindings p2mp opaque-type grt-recursive ipv4 detail

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.4)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
==============================================================================
LDP GRT Recursive with Generic IPv4 P2MP Bindings
==============================================================================
-------------------------------------------------------------------------------
P2MP Type      : 7                 P2MP-Id    : 8193
Root-Addr      : 192.0.2.3
InnerRoot-Addr : 192.0.2.1
-------------------------------------------------------------------------------
Peer           : 192.0.2.3:0
Ing Lbl        : 524282U
Egr Lbl        :   --
Egr Int/LspId  :   --
EgrNextHop     :   --
Egr. Flags     : None              Ing. Flags : None
===============================================================================
No. of GRT Recursive with Generic IPv4 P2MP Bindings: 1
==============================================================================
```

On P-3, there are two GRT recursive mLDP bindings with PE-1 as inner root, as follows:

```
*A:P-3# show router ldp bindings active p2mp summary ipv4
 No. of Generic IPv4 P2MP Active Bindings: 0
 No. of In-Band-SSM IPv4 P2MP Active Bindings: 0
 No. of In-Band-VPN-SSM IPv4 P2MP Active Bindings: 0
 No. of In-Band-SSM IPv4 P2MP Active Bindings: 0
 No. of VPN Recursive with Generic IPv4 P2MP Active Bindings: 0
 No. of GRT Recursive with Generic IPv4 P2MP Active Bindings: 2
```

The first GRT recursive mLDP binding has root 192.0.2.3 (P-3), which is the Lower FEC (LF) toward its peer PE-4; the second GRT recursive mLDP binding has root 192.168.23.1 (P-2), which is the Upper FEC (UF) toward the inner root PE-1, as follows:

```
*A:P-3# show router ldp bindings p2mp opaque-type grt-recursive ipv4 detail

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
```

```
===============================================================================
LDP GRT Recursive with Generic IPv4 P2MP Bindings
===============================================================================
-------------------------------------------------------------------------------
P2MP Type      : 7                 P2MP-Id    : 8193
Root-Addr      : 192.0.2.3 (LF)
InnerRoot-Addr : 192.0.2.1
-------------------------------------------------------------------------------
Peer           : 192.0.2.4:0
Ing Lbl        :   --
Egr Lbl        : 524282
Egr Int/LspId  : 1/1/1
EgrNextHop     : 192.168.34.2
Egr. Flags     : None              Ing. Flags : None
Egr If Name    : int-P-3-PE-4
Metric         : 1                 Mtu        : 8986
-------------------------------------------------------------------------------
P2MP Type      : 7                 P2MP-Id    : 8193
Root-Addr      : 192.168.23.1 (UF)
InnerRoot-Addr : 192.0.2.1
-------------------------------------------------------------------------------
Peer           : 192.168.23.1:0
Ing Lbl        : 524281U
Egr Lbl        :   --
Egr Int/LspId  :   --
EgrNextHop     :   --
Egr. Flags     : None              Ing. Flags : None
===============================================================================
No. of GRT Recursive with Generic IPv4 P2MP Bindings: 2
===============================================================================
```

On P-2, there is one GRT recursive mLDP binding with PE-1 as inner root and a non-recursive mLDP binding with root PE-1, as follows:

```
*A:P-2# show router ldp bindings active p2mp summary ipv4
 No. of Generic IPv4 P2MP Active Bindings: 1
 No. of In-Band-SSM IPv4 P2MP Active Bindings: 0
 No. of In-Band-VPN-SSM IPv4 P2MP Active Bindings: 0
 No. of In-Band-SSM IPv4 P2MP Active Bindings: 0
 No. of VPN Recursive with Generic IPv4 P2MP Active Bindings: 0
 No. of GRT Recursive with Generic IPv4 P2MP Active Bindings: 1
```

On P-2, the following GRT recursive mLDP binding with PE-1 as inner root has LF 192.168.23.1, which is an interface address of P-2. The peer is 192.168.23.2, which is an interface address of P-3.

```
*A:P-2# show router ldp bindings p2mp opaque-type grt-recursive ipv4 detail

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
===============================================================================
LDP GRT Recursive with Generic IPv4 P2MP Bindings
===============================================================================
-------------------------------------------------------------------------------
```

```
P2MP Type      : 7                 P2MP-Id    : 8193
Root-Addr      : 192.168.23.1 (LF)
InnerRoot-Addr : 192.0.2.1
-------------------------------------------------------------------------------
Peer           : 192.168.23.2:0
Ing Lbl        :   --
Egr Lbl        : 524281
Egr Int/LspId  : 1/1/1
EgrNextHop     : 192.168.23.2
Egr. Flags     : None              Ing. Flags : None
Egr If Name    : int-P-2-P-3
Metric         : 1                 Mtu        : 8986
===============================================================================
No. of GRT Recursive with Generic IPv4 P2MP Bindings: 1
==============================================================================
```

On P-2, the following non-recursive mLDP binding to root PE-1 has root address 192.0.2.1 as UF:

```
*A:P-2# show router ldp bindings p2mp opaque-type generic ipv4 detail

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
==============================================================================
LDP Generic IPv4 P2MP Bindings
==============================================================================
-------------------------------------------------------------------------------
P2MP Type      : 1                 P2MP-Id    : 8193
Root-Addr      : 192.0.2.1 (UF)
-------------------------------------------------------------------------------
Peer           : 192.0.2.1:0
Ing Lbl        : 524281U
Egr Lbl        :   --
Egr Int/LspId  :   --
EgrNextHop     :   --
Egr. Flags     : None              Ing. Flags : None
===============================================================================
No. of Generic IPv4 P2MP Bindings: 1
==============================================================================
```

On PE-1, there is only a non-recursive mLDP binding with root PE-1, as follows:

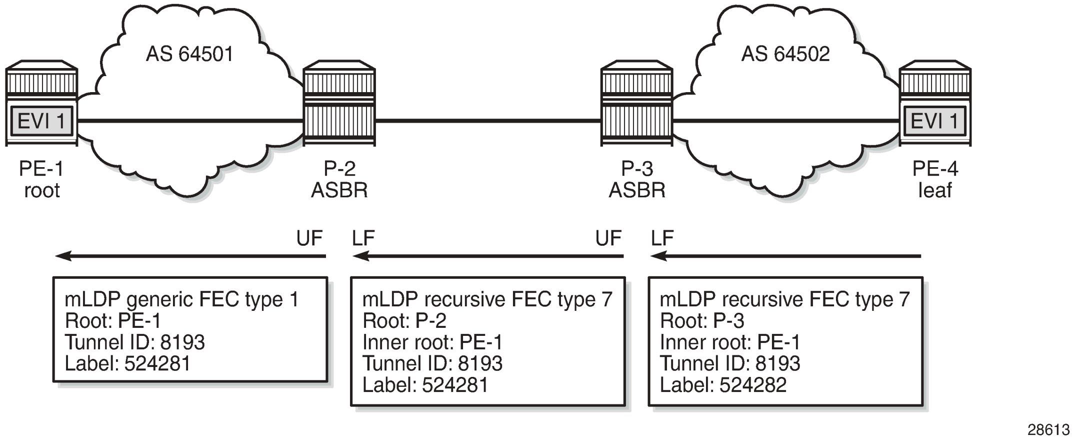
```
*A:PE-1# show router ldp bindings active p2mp summary ipv4
 No. of Generic IPv4 P2MP Active Bindings: 1
 No. of In-Band-SSM IPv4 P2MP Active Bindings: 0
 No. of In-Band-VPN-SSM IPv4 P2MP Active Bindings: 0
 No. of In-Band-SSM IPv4 P2MP Active Bindings: 0
 No. of VPN Recursive with Generic IPv4 P2MP Active Bindings: 0
 No. of GRT Recursive with Generic IPv4 P2MP Active Bindings: 0
```

On PE-1, the following non-recursive mLDP binding with root PE-1 has peer 192.0.2.2 (P-2):

```
*A:PE-1# show router ldp bindings p2mp opaque-type generic ipv4 detail
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
==============================================================================
LDP Generic IPv4 P2MP Bindings
==============================================================================
-------------------------------------------------------------------------------
P2MP Type      : 1                 P2MP-Id    : 8193
Root-Addr      : 192.0.2.1
-------------------------------------------------------------------------------
Peer           : 192.0.2.2:0
Ing Lbl        :   --
Egr Lbl        : 524281
Egr Int/LspId  : 1/1/1
EgrNextHop     : 192.168.12.2
Egr. Flags     : None              Ing. Flags : None
Egr If Name    : int-PE-1-P-2
Metric         : 1                 Mtu        : 8986
===============================================================================
No. of Generic IPv4 P2MP Bindings: 1
==============================================================================
```

Figure 226: Recursive mLDP FEC for inter-AS model C shows the mLDP label mapping messages with the corresponding labels: label 524281 is used between PE-1 and P-2; label 524281 is used between P-2 and P-3; label 524282 is used between P-3 and PE-4.

Figure 226: Recursive mLDP FEC for inter-AS model C



Non-recursive mLDP FEC resolution for inter-AS model C

Some routers may not support GRT recursive FEC type 7. In that case, the router generates a non-recursive FEC type 1 with root PE-1 to the next-hop P-3. In this example, leaf node PE-4 does not support GRT recursive FEC type 7 and is configured to only send basic FEC type 1 messages. ASBR P-3 supports

GRT recursive type 7 and sends similar messages as in the preceding scenario. However, it is possible that none of the routers supports GRT recursive FEC type 7. In that case, the **generate-basic-fec-only** command is configured on all nodes.

The following command is configured on leaf node PE-4 to make the system send only basic FEC type 1 messages:

```
# on PE-4:
configure
    router Base
        ldp
            generate-basic-fec-only
        exit
```

When PE-4 is configured to only generate basic FEC type 1, PE-4 withdraws the GRT recursive type 7 (T:7) label mapping message with PE-1 as inner root and P-3 as root and sends a non-recursive generic type 1 (T:1) label mapping message with PE-1 as root instead. When debugging is enabled on PE-4 for LDP label mapping messages between P-3 and PE-4, the following messages are logged:

```
# on PE-4:
1 2021/06/17 08:44:52.342 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Label Withdraw packet (msgId 96) to 192.0.2.3:0
Protocol version = 1
Label 524282 withdrawn for the following FECs
P2MP: root = 192.0.2.3, T: 7, L: 17 (InnerRoot: 192.0.2.1 T: 1, L: 4, TunnelId: 8193)
"

2 2021/06/17 08:44:52.342 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Label Mapping packet (msgId 97) to 192.0.2.3:0
Protocol version = 1
Label 524281 advertised for the following FECs
P2MP: root = 192.0.2.1, T: 1, L: 4, TunnelId: 8193
"

3 2021/06/17 08:44:52.344 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Recv Label Release packet (msgId 95) from 192.0.2.3:0
Protocol version = 1
Label 524282 released for the following FECs
P2MP: root = 192.0.2.3, T: 7, L: 17 (InnerRoot: 192.0.2.1 T: 1, L: 4, TunnelId: 8193)
"
```

On PE-4, there is one non-recursive generic mLDP binding, as follows:

```
*A:PE-4# show router ldp bindings p2mp summary ipv4
 No. of Generic IPv4 P2MP Bindings: 1
 No. of In-Band-SSM IPv4 P2MP Bindings: 0
 No. of In-Band-VPN-SSM IPv4 P2MP Bindings: 0
 No. of Recursive with In-Band-SSM IPv4 P2MP Bindings: 0
 No. of VPN Recursive with Generic IPv4 P2MP Bindings: 0
 No. of GRT Recursive with Generic IPv4 P2MP Bindings: 0
```

On PE-4, the following non-recursive generic mLDP binding has root PE-1 and peer P-3:

```
*A:PE-4# show router ldp bindings p2mp opaque-type generic detail ipv4

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.4)
```

```
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
==============================================================================
LDP Generic IPv4 P2MP Bindings
==============================================================================
-------------------------------------------------------------------------------
P2MP Type      : 1                  P2MP-Id    : 8193
Root-Addr      : 192.0.2.1
-------------------------------------------------------------------------------
Peer           : 192.0.2.3:0
Ing Lbl        : 524281U
Egr Lbl        :   --
Egr Int/LspId  :   --
EgrNextHop     :   --
Egr. Flags     : None               Ing. Flags : None
===============================================================================
No. of Generic IPv4 P2MP Bindings: 1
==============================================================================
```

On P-3, there is one generic mLDP binding and one recursive mLDP binding, as follows:

```
*A:P-3# show router ldp bindings p2mp summary ipv4
 No. of Generic IPv4 P2MP Bindings: 1
 No. of In-Band-SSM IPv4 P2MP Bindings: 0
 No. of In-Band-VPN-SSM IPv4 P2MP Bindings: 0
 No. of Recursive with In-Band-SSM IPv4 P2MP Bindings: 0
 No. of VPN Recursive with Generic IPv4 P2MP Bindings: 0
 No. of GRT Recursive with Generic IPv4 P2MP Bindings: 1
```

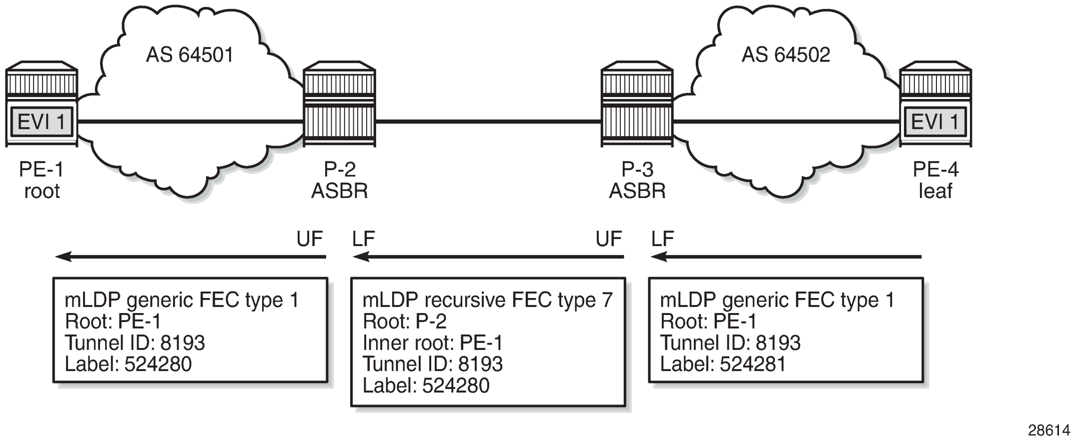
```
*A:P-3# show router ldp bindings p2mp opaque-type generic detail ipv4

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
==============================================================================
LDP Generic IPv4 P2MP Bindings
==============================================================================
-------------------------------------------------------------------------------
P2MP Type      : 1                  P2MP-Id    : 8193
Root-Addr      : 192.0.2.1 (LF)
-------------------------------------------------------------------------------
Peer           : 192.0.2.4:0
Ing Lbl        :   --
Egr Lbl        : 524281
Egr Int/LspId  : 1/1/1
EgrNextHop     : 192.168.34.2
Egr. Flags     : None               Ing. Flags : None
Egr If Name    : int-P-3-PE-4
```

```
Metric         : 1                 Mtu        : 8986
===============================================================================
No. of Generic IPv4 P2MP Bindings: 1
==============================================================================
```

```
*A:P-3# show router ldp bindings p2mp opaque-type grt-recursive detail ipv4

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
==============================================================================
LDP GRT Recursive with Generic IPv4 P2MP Bindings
==============================================================================
-------------------------------------------------------------------------------
P2MP Type      : 7                 P2MP-Id    : 8193
Root-Addr      : 192.168.23.1 (UF)
InnerRoot-Addr : 192.0.2.1
-------------------------------------------------------------------------------
Peer           : 192.168.23.1:0
Ing Lbl        : 524280U
Egr Lbl        :   --
Egr Int/LspId  :   --
EgrNextHop     :   --
Egr. Flags     : None              Ing. Flags : None
===============================================================================
No. of GRT Recursive with Generic IPv4 P2MP Bindings: 1
==============================================================================
```

On P-2 and PE-1, the mLDP bindings are similar to the preceding scenario, but the labels are different. Figure 227: Non-recursive mLDP FEC for inter-AS model C shows the label mapping messages with label 524280 between PE-1 and P-2 and label 524280 between P-2 and P-3; label 524281 is used between P-3 and PE-4.

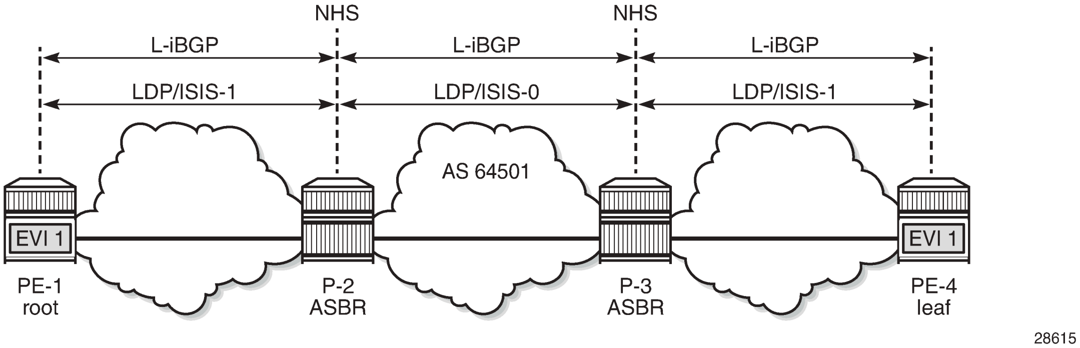
Figure 227: Non-recursive mLDP FEC for inter-AS model C



Seamless MPLS

Figure 228: Example topology for seamless MPLS shows the example topology for seamless MPLS.

Figure 228: Example topology for seamless MPLS



The configuration is according to the *Seamless MPLS: Isolated IGP/LDP Domains and Labeled BGP* chapter.

IS-IS is configured as IGP. IS-IS instance 0 is configured between P-2 and P-3, whereas IS-IS instance 1 is configured between P-2 and PE-1 and between P-3 and PE-4. On P-2, IS-IS is configured as follows:

```
# on P-2:
configure
    router
        isis 0
            level-capability level-2
            area 49.0001
            interface "system"
            exit
            interface "int-P-2-P-3"
                interface-type point-to-point
            exit
            no shutdown
        exit
        isis 1
            level-capability level-2
            area 49.0001
            interface "system"
            exit
            interface "int-P-2-PE-1"
                interface-type point-to-point
            exit
            no shutdown
        exit
```

Other characteristics of this example are as follows:

- Unlike the preceding use case for inter-AS model C, no static route is required between P-2 and P-3.
- LDP is configured on all interfaces.
- VPLS 1 is configured as before, but the route target is identical for import and export, and equal to 64501:1.
- All nodes are in AS 64501, so only iBGP is configured.

On PE-1, BGP is configured as follows, using the same policy as for inter-AS model C. The BGP configuration on PE-4 is similar, but the neighbors are different.

```
# on PE-1:
configure
    router Base
       autonomous-system 64501
       bgp
            split-horizon
            group "iBGP"
                type internal
                neighbor 192.0.2.2
                    family label-ipv4
                    export "PE-sys-to-labeled-BGP"
                exit
                neighbor 192.0.2.4
                    family evpn
                exit
            exit
        exit
```

On P-2, the BGP configuration is as follows. The ABRs are configured with **next-hop-self** in both directions. The BGP configuration is similar on P-3.

```
*A:P-2#
configure
    router Base
        autonomous-system 64501
        bgp
            split-horizon
            group "iBGP"
                type internal
                neighbor 192.0.2.1
                    family label-ipv4
                    next-hop-self
                    cluster 192.0.2.2
                exit
                neighbor 192.0.2.3
                    family label-ipv4
                    next-hop-self
                    advertise-inactive
                exit
            exit
```

The following route table on PE-4 shows a labeled BGP route to root node PE-1 with P-3 as the next-hop:

```
*A:PE-4# show router route-table 192.0.2.1

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                  Remote  BGP_LABEL 00h00m47s  170
       192.0.2.3 (tunneled)                                         10
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
```

```
===============================================================================
```

Likewise, P-3 has a labeled BGP route to root node PE-1 with P-2 as the next-hop, as follows:

```
*A:P-3# show router route-table 192.0.2.1

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                  Remote  BGP_LABEL 00h01m00s  170
       192.0.2.2 (tunneled)                                         10
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

P-2 has an IS-IS route to PE-1, using IS-IS instance 1, as follows:

```
*A:P-2# show router route-table 192.0.2.1

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                  Remote  ISIS(1)   00h01m51s  18
       192.168.12.1                                                 10
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

Recursive mLDP FEC resolution for seamless MPLS

When the leaf node PE-4 supports GRT recursive FEC type 7, it generates one GRT recursive FEC label mapping message with PE-1 as inner root and P-3 as root, as follows:

```
*A:PE-4# show router ldp bindings p2mp summary ipv4
 No. of Generic IPv4 P2MP Bindings: 0
 No. of In-Band-SSM IPv4 P2MP Bindings: 0
 No. of In-Band-VPN-SSM IPv4 P2MP Bindings: 0
 No. of Recursive with In-Band-SSM IPv4 P2MP Bindings: 0
 No. of VPN Recursive with Generic IPv4 P2MP Bindings: 0
 No. of GRT Recursive with Generic IPv4 P2MP Bindings: 1
```

```
*A:PE-4# show router ldp bindings p2mp opaque-type grt-recursive detail ipv4

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.4)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
==============================================================================
LDP GRT Recursive with Generic IPv4 P2MP Bindings
==============================================================================
-------------------------------------------------------------------------------
P2MP Type      : 7                 P2MP-Id    : 8193
Root-Addr      : 192.0.2.3
InnerRoot-Addr : 192.0.2.1
-------------------------------------------------------------------------------
Peer           : 192.0.2.3:0
Ing Lbl        : 524281U
Egr Lbl        :   --
Egr Int/LspId  :   --
EgrNextHop     :   --
Egr. Flags     : None              Ing. Flags : None
===============================================================================
No. of GRT Recursive with Generic IPv4 P2MP Bindings: 1
==============================================================================
```

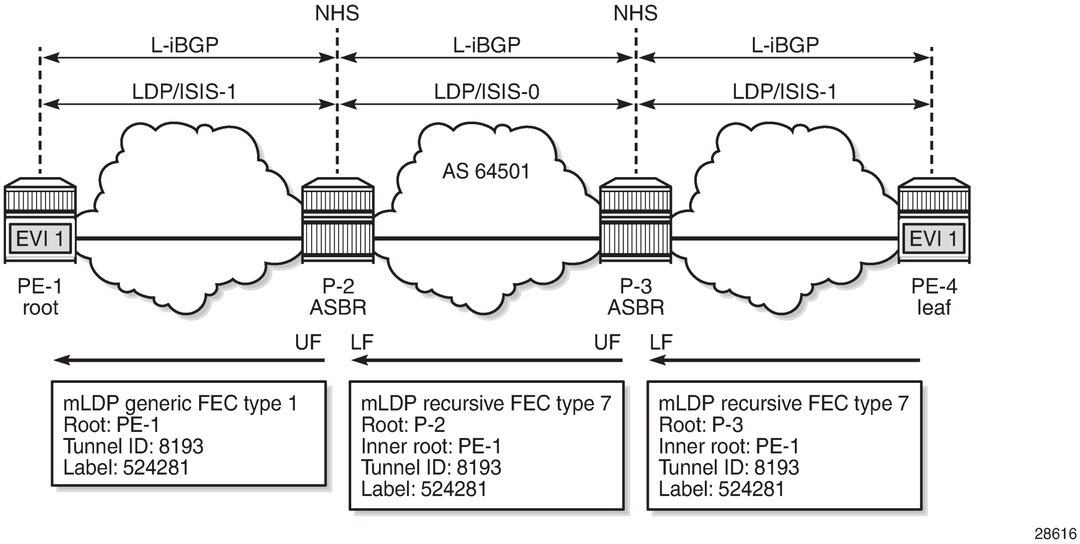
P-3 has two GRT recursive FEC bindings with inner root 192.0.2.1: one with UF 192.0.2.2 and another with LF 192.0.2.3, as follows:

```
*A:P-3# show router ldp bindings p2mp opaque-type grt-recursive detail ipv4

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
==============================================================================
LDP GRT Recursive with Generic IPv4 P2MP Bindings
==============================================================================
-------------------------------------------------------------------------------
P2MP Type      : 7                 P2MP-Id    : 8193
Root-Addr      : 192.0.2.2 (UF)
InnerRoot-Addr : 192.0.2.1
-------------------------------------------------------------------------------
Peer           : 192.0.2.2:0
Ing Lbl        : 524281U
Egr Lbl        :   --
Egr Int/LspId  :   --
EgrNextHop     :   --
Egr. Flags     : None              Ing. Flags : None
-------------------------------------------------------------------------------
P2MP Type      : 7                 P2MP-Id    : 8193
Root-Addr      : 192.0.2.3 (LF)
InnerRoot-Addr : 192.0.2.1
-------------------------------------------------------------------------------
```

```
Peer            : 192.0.2.4:0
Ing Lbl         :   --
Egr Lbl         : 524281
Egr Int/LspId   : 1/1/1
EgrNextHop      : 192.168.34.2
Egr. Flags      : None               Ing. Flags : None
Egr If Name     : int-P-3-PE-4
Metric          : 1                  Mtu        : 8986
===============================================================================
No. of GRT Recursive with Generic IPv4 P2MP Bindings: 2
==============================================================================
```

P-2 has one GRT recursive FEC binding with inner root PE-1 and root P-2 (LF). P-2 also has one non-recursive FEC binding with root PE-1 (UF). PE-1 only has a non-recursive FEC binding with root PE-1. Figure 229: Recursive mLDP FEC for seamless MPLS shows the mLDP label mapping messages that all have label 524281 in this example.

Figure 229: Recursive mLDP FEC for seamless MPLS



Non-recursive mLDP FEC resolution for seamless MPLS

For nodes that do not support GRT recursive mLDP FEC type 7, the following command ensures that only non-recursive mLDP type 1 label mapping messages will be sent. In this example, it is assumed that only PE-4 does not support GRT recursive mLDP FEC type 7.

```
# on PE-4:
configure
    router Base
        ldp
            generate-basic-fec-only
```

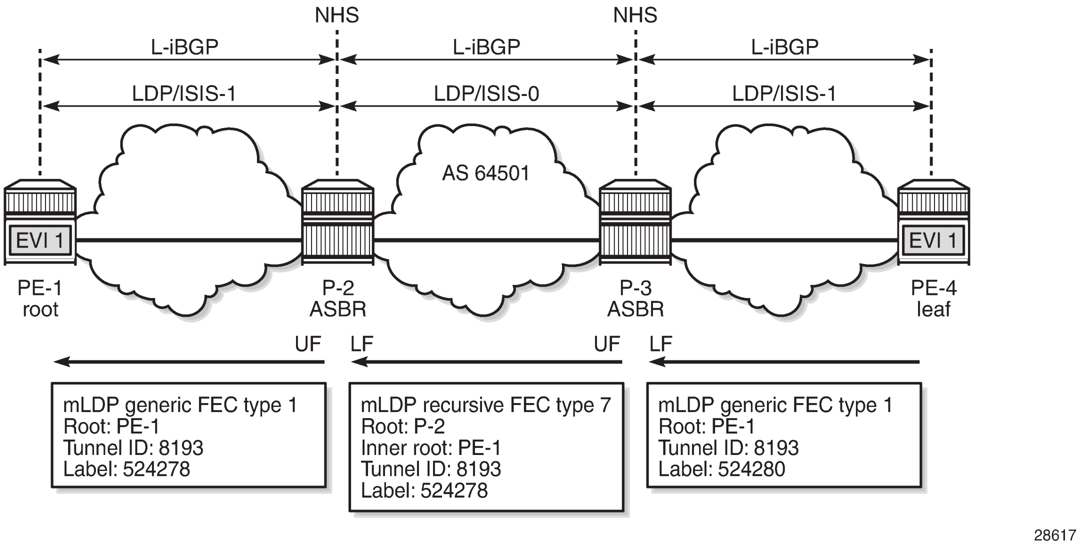
PE-4 sends a non-recursive mLDP label mapping message with PE-1 as the root to its peer P-3, as follows:

```
*A:PE-4# show router ldp bindings p2mp opaque-type generic detail ipv4
```

```
===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.4)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
==============================================================================
LDP Generic IPv4 P2MP Bindings
==============================================================================
-------------------------------------------------------------------------------
P2MP Type      : 1                 P2MP-Id    : 8193
Root-Addr      : 192.0.2.1
-------------------------------------------------------------------------------
Peer           : 192.0.2.3:0
Ing Lbl        : 524280U
Egr Lbl        :   --
Egr Int/LspId  :   --
EgrNextHop     :   --
Egr. Flags     : None              Ing. Flags : None
===============================================================================
No. of Generic IPv4 P2MP Bindings: 1
==============================================================================
```

Figure 230: Leaf node sends basic FEC in seamless MPLS shows the label mapping messages when leaf node PE-4 only generates basic FEC type 1 messages.

Figure 230: Leaf node sends basic FEC in seamless MPLS



It is possible that ABR routers do not support GRT recursive either. The same command is configured on P-2 and P-3, as follows:

```
# on P-2, P-3:
configure
```

```
    router Base
        ldp
            generate-basic-fec-only
```

When **generate-basic-fec-only** is enabled in the ABRs, P-2 and P-3 will only generate basic FEC messages. On P-3, there are no GRT recursive mLDP bindings anymore, as follows:

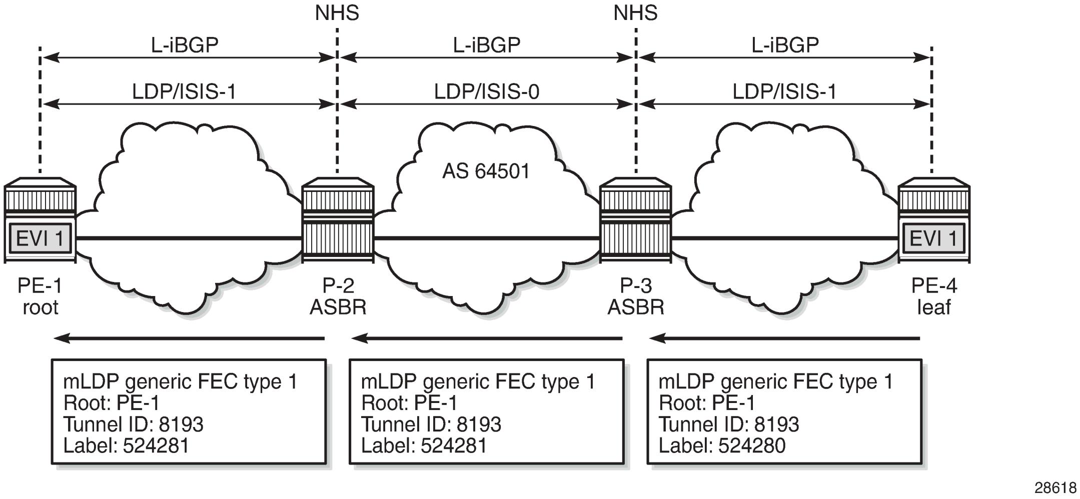
```
*A:P-3# show router ldp bindings p2mp summary ipv4
 No. of Generic IPv4 P2MP Bindings: 2
 No. of In-Band-SSM IPv4 P2MP Bindings: 0
 No. of In-Band-VPN-SSM IPv4 P2MP Bindings: 0
 No. of Recursive with In-Band-SSM IPv4 P2MP Bindings: 0
 No. of VPN Recursive with Generic IPv4 P2MP Bindings: 0
 No. of GRT Recursive with Generic IPv4 P2MP Bindings: 0
```

The two generic mLDP bindings on P-3 have root address 192.0.2.1, as follows. There is no UF or LF.

```
*A:P-3# show router ldp bindings p2mp opaque-type generic detail ipv4

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
==============================================================================
LDP Generic IPv4 P2MP Bindings
==============================================================================
-------------------------------------------------------------------------------
P2MP Type      : 1                  P2MP-Id    : 8193
Root-Addr      : 192.0.2.1
-------------------------------------------------------------------------------
Peer           : 192.0.2.2:0
Ing Lbl        : 524281U
Egr Lbl        :   --
Egr Int/LspId  :   --
EgrNextHop     :   --
Egr. Flags     : None               Ing. Flags : None
-------------------------------------------------------------------------------
P2MP Type      : 1                  P2MP-Id    : 8193
Root-Addr      : 192.0.2.1
-------------------------------------------------------------------------------
Peer           : 192.0.2.4:0
Ing Lbl        :   --
Egr Lbl        : 524280
Egr Int/LspId  : 1/1/1
EgrNextHop     : 192.168.34.2
Egr. Flags     : None               Ing. Flags : None
Egr If Name    : int-P-3-PE-4
Metric         : 1                  Mtu        : 8986
===============================================================================
No. of Generic IPv4 P2MP Bindings: 2
==============================================================================
```

The output on P-2 is similar. Figure 231: ABRs and leaf node send basic FEC in seamless MPLS shows the label mapping messages when all nodes only generate basic FEC type 1 messages.

Figure 231: ABRs and leaf node send basic FEC in seamless MPLS



Conclusion

In inter-AS and intra-AS scenarios, mLDP trees can be set up using recursive or non-recursive label mapping messages. Routers not supporting recursive FEC can generate only non-recursive FEC, even if the system address of the root node is resolved via BGP. This feature is supported in MVPN and in EVPN.

P2MP mLDP Inter-AS Model C for EVPN-MPLS Services

This chapter provides information about P2MP mLDP Inter-AS Model C for EVPN-MPLS Services.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS Release 15.0.R5, but the CLI in the current edition is based on SR OS Release 21.5.R1.

Point-to-Multipoint Multicast Label Distribution Protocol (P2MP mLDP) for Broadcast, Unknown Unicast, and Multicast (BUM) traffic in EVPN-MPLS networks is supported in SR OS Release 14.0.R1, and later. EVPN with P2MP mLDP LSPs is supported in a seamless MPLS or inter-AS model C scenario in SR OS Release 15.0.R1, and later. This chapter describes the inter-AS model C scenario, but the configuration for seamless MPLS is similar.

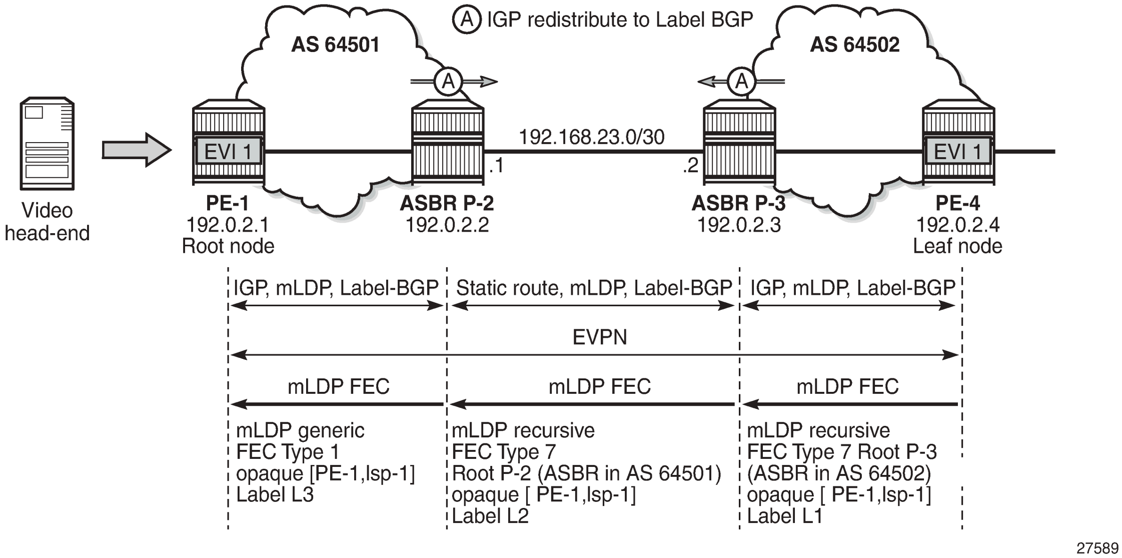
Overview

Chapter P2MP mLDP Tunnels for BUM Traffic in EVPN-MPLS Services describes P2MP mLDP within an Autonomous System (AS). PEs configured as root-and-leaf can send BUM traffic over P2MP mLDP tunnels; PEs configured as not root-and-leaf (that is, leaf-only) can only send BUM traffic over Ingress Replication (IR) tunnels. Both types of PEs (root-and-leaf and leaf-only) can receive BUM traffic over either P2MP mLDP tunnels or IR tunnels.

When **provider-tunnel inclusive mldp** is enabled in an EVPN-MPLS service in combination with **root-and-leaf** and **bgp-evpn>ingress-repl-inc-mcast-advertisement**, the system will send an Inclusive Multicast Ethernet Tag (IMET) route with a composite tunnel type (IMET-P2MP-IR) in the provider tunnel attributed.

Inter-AS VPN model C is described in chapters "Inter-AS VPRN Model C" in the *7450 ESS, 7750 SR, and 7950 XRS Layer 3 Services Advanced Configuration Guide for Classic CLI* and Inter-AS Model C for VLL. Labeled IPv4 unicast BGP is used to provide inter-AS connectivity. The system IP addresses within each AS are exported by the Autonomous System Border Routers (ASBRs) and a multi-hop BGP session is established between root node and leaf node for address family EVPN. The root node advertises a composite IMET-P2MP-IR route to the leaf nodes and the leaf nodes advertise an IMET-IR route to the root node. Figure 232: Inter-AS Model C for P2MP mLDP shows an example topology with root node PE-1 in AS 64501 and leaf node PE-4 in AS 64502. P-2 and P-3 are ASBRs.

Figure 232: Inter-AS Model C for P2MP mLDP



The composite IMET-P2MP-IR route received by leaf node PE-4 contains the root node (192.0.2.1) and the LSP ID (0x2001) that will be used by the nodes to set up a P2MP mLDP tree toward the root.

```
# on PE-4:
3 2021/06/02 08:31:13.913 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.1
"Peer 1: 192.0.2.1: UPDATE
Peer 1: 192.0.2.1 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 92
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.1
        Type: EVPN-INCL-MCAST Len: 17 RD: 192.0.2.1:1, tag: 0, orig_addr len: 32,
                              orig_addr: 192.0.2.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 6 AS Path:
        Type: 2 Len: 1 < 64501 >
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64501:1
        bgp-tunnel-encap:MPLS
    Flag: 0xc0 Type: 22 Len: 25 PMSI:
        Tunnel-type Composite LDP P2MP IR (130)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf: not required]
        MPLS Label1 Ag  0
        MPLS Label2 IR 8388544
        Root-Node 192.0.2.1, LSP-ID 0x2001
"
```

The Provider Multicast Service Interface (PMSI) tunnel attribute for tunnel type 130 (composite tunnel) has two MPLS labels, of which MPLS label 1 always equals zero in SR OS Release 21.5.R1, because SR OS does not support aggregated P2MP tunnels. MPLS label 2 is used by the downstream nodes to set up the EVPN-MPLS destination to the root node and add it to the default multicast list. The actual MPLS label only uses the high-order 20 bits out of the 24 bits advertised in the MPLS label. Therefore, the value 8388544 needs to be divided by 16 to get the MPLS label value: 8388544/16 = 524284. This is due to the debug message being shown before the router can parse the label field and see whether it corresponds to an

MPLS label (20 bits) or a VXLAN VNI (24 bits). The following command on PE-4 shows the EVPN-MPLS destination 192.0.2.1 with MPLS label 524284 using a BGP transport tunnel:

```
*A:PE-4# show service id 1 evpn-mpls

===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address     Egr Label     Num. MACs   Mcast           Last Change
                Transport:Tnl                             Sup BCast Domain
-------------------------------------------------------------------------------
192.0.2.1       524284        0           bum             06/02/2021 08:31:14
                bgp:262146                                No
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

The use of mLDP with recursive opaque values is specified in RFC 6512.

When the leaf node PE-4 receives the composite IMET-P2MP-IR route from the root node PE-1, a P2MP mLDP tree needs to be established from the leaf node to the root node. Leaf node PE-4 resolves the IP address of PE-1 to a labeled BGP route with next-hop ASBR P-3. PE-4 then sends an mLDP FEC with root node ASBR P-3 and an opaque value containing the root PE-1 and an LSP ID that was advertised in the IMET-P2MP-IR route, as follows:

```
# on PE-4:
4 2021/06/02 08:31:13.915 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Label Mapping packet (msgId 40) to 192.0.2.3:0
Protocol version = 1
Label 524283 advertised for the following FECs
P2MP: root = 192.0.2.3, T: 7, L: 17 (InnerRoot: 192.0.2.1 T: 1, L: 4, TunnelId: 8193)
"
```

T: 7 indicates the MLDP recursive FEC type 7. The tunnel ID 8193 corresponds to the hexadecimal value 0x2001 sent by the root node PE-1, which is the inner root 192.0.2.1 in the recursive opaque value.

When ASBR P-3 receives this mLDP FEC, it identifies itself as root node and resolves the recursive opaque value (PE-1, LSP ID) and creates a new mLDP FEC element with root node ASBR P-2 and an identical opaque value (PE-1, LSP ID). The following mLDP FEC is sent to ASBR P-2:

```
# on P-3:
12 2021/06/02 08:32:36.794 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Label Mapping packet (msgId 34) to 192.168.23.1:0
Protocol version = 1
Label 524279 advertised for the following FECs
P2MP: root = 192.168.23.1, T: 7, L: 17 (InnerRoot: 192.0.2.1 T: 1, L: 4, TunnelId: 8193)
"
```

ASBR P-2 receives the mLDP FEC and finds that it is the root node. P-2 creates a new mLDP FEC, but no recursion is required because P-2 knows the IP address of PE-1 through the IGP. P-2 sends the following mLDP FEC with root node PE-1, LSP ID 8193, and mLDP FEC type 1.

```
# on P-2:
12 2021/06/02 08:32:36.814 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
```

```
Send Label Mapping packet (msgId 50) to 192.0.2.1:0
Protocol version = 1
Label 524279 advertised for the following FECs
P2MP: root = 192.0.2.1, T: 1, L: 4, TunnelId: 8193
"
```

Configuration

The example topology was already shown in Figure 232: Inter-AS Model C for P2MP mLDP. The initial configuration includes the following:

- Cards, MDAs, ports
- Router interfaces
- OSPF as IGP within each AS (alternatively, IS-IS can be used)
- LDP enabled within each AS

The following two scenarios are configured:

- Inter-AS model C for mLDP
- Optimized inter-AS model C for mLDP

Inter-AS Model C for mLDP

The initial BGP configuration on the PEs only includes a label-IPv4 peering with the ASBRs. The BGP configuration on PE-1 is as follows:

```
# on PE-1:
configure
    router Base
        bgp
            group "iBGP"
                type internal
                neighbor 192.0.2.2
                    family label-ipv4
                exit
```

On the ASBRs, BGP is configured for address family label-IPv4, both internal to PE-1 and external to the peer ASBR. The BGP configuration on P-2 is as follows:

```
# on P-2:
configure
    router Base
        bgp
            group "eBGP"
                type external
                neighbor 192.168.23.2
                    family label-ipv4
                    export "PE-sys-to-labeled-BGP"
                    local-as 64501
                    peer-as 64502
                    split-horizon
                exit
            exit
            group "iBGP"
```

```
                type internal
                neighbor 192.0.2.1
                    family label-ipv4
                exit
            exit
```

The BGP configuration on ASBR P-3 is similar, but the IP addresses are different and the local AS and peer AS are swapped. The following export policy is identical on both ASBRs P-2 and P-3:

```
# on P-2, P-3:
configure
    router
        policy-options
            begin
            prefix-list "sysPE"
                prefix 192.0.2.0/24 longer
            exit
            policy-statement "PE-sys-to-labeled-BGP"
                entry 10
                    from
                        prefix-list "sysPE"
                    exit
                    to
                        protocol bgp-label
                    exit
                    action accept
                    exit
                exit
            exit
            commit
```

This policy exports the system prefixes as label-IPv4 routes to the eBGP peer.

When a P2MP mLDP tree must be established across ASs, LDP needs to be enabled on the interface between the ASBRs with **local-lsr-id interface** instead of the default value "system". The LDP configuration on P-2 is as follows:

```
# on P-2:
configure
    router
        ldp
            interface-parameters
                interface "int-P-2-P-3"
                    ipv4
                        local-lsr-id interface
                    exit
                exit
```

With this LDP configuration, a link adjacency will be established toward the interface IP address instead of the system address, as follows:

```
*A:P-2# show router ldp session ipv4

==============================================================================
LDP IPv4 Sessions
==============================================================================
Peer LDP Id         Adj Type  State         Msg Sent  Msg Recv  Up Time
------------------------------------------------------------------------------
192.0.2.1:0         Link      Established   229       229       0d 00:09:48
192.168.23.2:0      Link      Established   137       140       0d 00:05:46
------------------------------------------------------------------------------
```

```
No. of IPv4 Sessions: 2
===============================================================================
```

However, this LDP configuration is insufficient for the resolution of mLDP FEC as link LSR ID. LDP needs a /32 route instead of a /30 route, so the following /32 static route is configured on P-2:

```
# on P-2:
configure
    router
        static-route-entry 192.168.23.2/32
            next-hop 192.168.23.2
                no shutdown
            exit
        exit
```

The configuration on ASBR P-3 is similar for static route 192.168.23.1/32. When this static route is not configured, no mLDP label mapping message will be sent from P-3 to P-2, so the mLDP P2MP tree cannot be established.

On PE-1, VPLS 1 is configured with mLDP root-and-leaf, as follows:

```
# on PE-1:
configure
    service
        vpls 1 name "EVI-1" customer 1 create
            bgp
                route-target export target:64501:1 import target:64502:1
            exit
            bgp-evpn
                ingress-repl-inc-mcast-advertisement
                evi 1
                mpls bgp 1
                    ingress-replication-bum-label
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            provider-tunnel
                inclusive
                    owner bgp-evpn-mpls
                    root-and-leaf
                    mldp
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/2/1:1 create
                no shutdown
            exit
            no shutdown
```

On PE-4, VPLS 1 is configured with mLDP leaf-only (no root-and-leaf, which is default), as follows:

```
# on PE-4:
configure
    service
        vpls 1 name "EVI-1" customer 1 create
```

```
            bgp
                route-target export target:64502:1 import target:64501:1
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    ingress-replication-bum-label
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            provider-tunnel
                inclusive
                    owner bgp-evpn-mpls
                    mldp
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/2/1:1 create
                no shutdown
            exit
            no shutdown
```

The Route Distinguisher (RD) is auto-derived from EVI 1, but the route target (RT) should not be auto-derived, because the export RT on PE-1 must match the import RT on PE-4, and vice versa. It is an option to configure an identical RT on all PEs, such as 1:1, but in this example, the export RT on PE-1 is 64501:1, which equals the import RT on PE-4. When the RTs do not match, the BGP routes will be received at the PE in the peer AS, but they will not become active and no mLDP P2MP tree can be established.

Multi-hop BGP peering is configured between PE-1 and PE-4 for address family EVPN. The external BGP configuration on PE-1 is as follows:

```
# on PE-1:
configure
    router
        bgp
            split-horizon
            group eBGP
                family evpn
                multihop 10
                local-as 64501
                peer-as 64502
                neighbor 192.0.2.4
                exit
            exit
```

The external BGP configuration on PE-4 is similar, but the local AS and peer AS are swapped, and the neighbor IP address is different.

Inter-AS Model C for mLDP - Verification

The following BGP summary shows that P-2 has sent and received two prefixes with its eBGP peer P-3 and has advertised two prefixes to its iBGP peer PE-1:

```
*A:P-2# show router bgp summary all

===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
Description
ServiceId          AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
192.0.2.1
Def. Instance  64501       11    0 00h04m03s 0/0/2 (Lbl-IPv4)
                           13    0
192.168.23.2
Def. Instance  64502       12    0 00h03m54s 2/2/2 (Lbl-IPv4)
                           12    0

-------------------------------------------------------------------------------
```

ASBR P-2 advertised the following prefixes from AS 64501 to its neighbor P-3:

```
*A:P-2# show router bgp neighbor 192.168.23.2 advertised-routes label-ipv4
===============================================================================
 BGP Router ID:192.0.2.2        AS:64501       Local AS:64501
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP Routes
===============================================================================
Flag  Network                                            LocalPref   MED
      Nexthop (Router)                                   Path-Id     IGP Cost
      As-Path                                                        Label
-------------------------------------------------------------------------------
i     192.0.2.1/32                                       n/a         10
      192.168.23.1                                       None        n/a
      64501                                                          524284
i     192.0.2.2/32                                       n/a         None
      192.168.23.1                                       None        n/a
      64501                                                          524285
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

ASBR P-2 received the following prefixes from AS 64502 from its neighbor P-3. Both routes are used.

```
*A:P-2# show router bgp neighbor 192.168.23.2 received-routes label-ipv4
===============================================================================
 BGP Router ID:192.0.2.2        AS:64501       Local AS:64501
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP Routes
===============================================================================
Flag  Network                                            LocalPref   MED
      Nexthop (Router)                                   Path-Id     IGP Cost
      As-Path                                                        Label
-------------------------------------------------------------------------------
u*>i  192.0.2.3/32                                       n/a         None
      192.168.23.2                                       None        0
      64502                                                          524285
u*>i  192.0.2.4/32                                       n/a         10
      192.168.23.2                                       None        0
      64502                                                          524284
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

These routes are advertised by P-2 to its iBGP neighbor PE-1, so PE-1 will have the same label-IPv4 routes. The following command shows the route table on PE-1 that includes tunneled routes to P-3 and PE-4 in AS 64502.

```
*A:PE-1# show router route-table

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                  Local   Local     00h06m21s  0
       system                                                       0
192.0.2.2/32                                  Remote  OSPF      00h06m13s  10
       192.168.12.2                                                 10
192.0.2.3/32                                  Remote  BGP_LABEL 00h03m20s  170
       192.0.2.2 (tunneled)                                         10
192.0.2.4/32                                  Remote  BGP_LABEL 00h03m20s  170
       192.0.2.2 (tunneled)                                         10
192.168.12.0/30                               Local   Local     00h06m21s  0
       int-PE-1-P-2                                                 0
-------------------------------------------------------------------------------
No. of Routes: 5
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The following command shows the tunnel table on PE-1:

```
*A:PE-1# show router tunnel-table

===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination           Owner     Encap TunnelId  Pref   Nexthop        Metric
   Color
-------------------------------------------------------------------------------
```

```
127.0.128.0/32        sdp       MPLS  32767     5      127.0.128.0    0
192.0.2.2/32          ldp       MPLS  65537     9      192.168.12.2   10
192.0.2.3/32          bgp       MPLS  262145    12     192.0.2.2      1000
192.0.2.4/32          bgp       MPLS  262146    12     192.0.2.2      1000
-------------------------------------------------------------------------------
Flags: B = BGP or MPLS backup hop available
       L = Loop-Free Alternate (LFA) hop available
       E = Inactive best-external BGP route
       k = RIB-API or Forwarding Policy backup hop
===============================================================================
```

The tunnels toward P-3 and PE-4 are BGP tunnels. The SDP in the list is auto-created on the root node by mLDP. The output of these show commands on PE-4 is similar, but no SDP will be created on a leaf-only node.

The route-table on ASBR P-2 includes tunneled routes toward P-3 and PE-4 and a static route to 192.168.23.2/32, as follows:

```
*A:P-2# show router route-table

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                  Remote  OSPF      00h06m24s  10
       192.168.12.1                                                 10
192.0.2.2/32                                  Local   Local     00h06m25s  0
       system                                                       0
192.0.2.3/32                                  Remote  BGP_LABEL 00h03m50s  170
       192.168.23.2                                                 0
192.0.2.4/32                                  Remote  BGP_LABEL 00h03m50s  170
       192.168.23.2                                                 0
192.168.12.0/30                               Local   Local     00h06m25s  0
       int-P-2-PE-1                                                 0
192.168.23.0/30                               Local   Local     00h06m25s  0
       int-P-2-P-3                                                  0
192.168.23.2/32                               Remote  Static    00h00m28s  5
       192.168.23.2                                                 1
-------------------------------------------------------------------------------
No. of Routes: 7
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The tunnel table on P-2 has an LDP tunnel toward PE-1 and a BGP tunnel toward P-3 and PE-4, as follows:

```
*A:P-2# show router tunnel-table

===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination           Owner     Encap TunnelId  Pref   Nexthop        Metric
   Color
-------------------------------------------------------------------------------
192.0.2.1/32          ldp       MPLS  65537     9      192.168.12.1   10
192.0.2.3/32          bgp       MPLS  262146    12     192.168.23.2   1000
192.0.2.4/32          bgp       MPLS  262145    12     192.168.23.2   1000
```

```
-------------------------------------------------------------------------------
Flags: B = BGP or MPLS backup hop available
       L = Loop-Free Alternate (LFA) hop available
       E = Inactive best-external BGP route
       k = RIB-API or Forwarding Policy backup hop
===============================================================================
```

One BGP-EVPN IMET route is received and used on PE-1:

```
*A:PE-1# show router bgp routes evpn incl-mcast
===============================================================================
 BGP Router ID:192.0.2.1        AS:64501       Local AS:64501
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Flag  Route Dist.         OrigAddr
      Tag                 NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.4:1         192.0.2.4
      0                   192.0.2.4

-------------------------------------------------------------------------------
Routes : 1
===============================================================================
```

The preceding route is an IMET-IR route received from node PE-4, as follows:

```
*A:PE-1# show router bgp routes evpn incl-mcast detail
===============================================================================
 BGP Router ID:192.0.2.1        AS:64501       Local AS:64501
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Original Attributes

Network        : n/a
Nexthop        : 192.0.2.4
From           : 192.0.2.4
Res. Nexthop   : n/a
Local Pref.    : n/a                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 0
Connector      : None
Community      : target:64502:1 bgp-tunnel-encap:MPLS
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.4
Flags          : Used Valid Best IGP
Route Source   : External
AS-Path        : 64502
EVPN type      : INCL-MCAST
```

```
Tag            : 0
Originator IP  : 192.0.2.4
Route Dist.    : 192.0.2.4:1
Route Tag      : 0
Neighbor-AS    : 64502
Orig Validation: N/A
Source Class   : 0                       Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h01m57s
-------------------------------------------------------------------------------
PMSI Tunnel Attributes :
Tunnel-type    : Ingress Replication
Flags          : Type: RNVE(0) BM: 0 U: 0 Leaf: not required
MPLS Label     : LABEL 524284
Tunnel-Endpoint: 192.0.2.4
-------------------------------------------------------------------------------
---snip---
```

PE-4 has received an IMET-P2MP-IR route sent by root node PE-1, as follows:

```
*A:PE-4# show router bgp routes evpn incl-mcast detail
===============================================================================
 BGP Router ID:192.0.2.4        AS:64502       Local AS:64502
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Original Attributes

Network        : n/a
Nexthop        : 192.0.2.1
From           : 192.0.2.1
Res. Nexthop   : n/a
Local Pref.    : n/a                     Interface Name : NotAvailable
Aggregator AS  : None                    Aggregator     : None
Atomic Aggr.   : Not Atomic              MED            : None
AIGP Metric    : None                    IGP Cost       : 0
Connector      : None
Community      : target:64501:1 bgp-tunnel-encap:MPLS
Cluster        : No Cluster Members
Originator Id  : None                    Peer Router Id : 192.0.2.1
Flags          : Used Valid Best IGP
Route Source   : External
AS-Path        : 64501
EVPN type      : INCL-MCAST
Tag            : 0
Originator IP  : 192.0.2.1
Route Dist.    : 192.0.2.1:1
Route Tag      : 0
Neighbor-AS    : 64501
Orig Validation: N/A
Source Class   : 0                       Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h01m59s
-------------------------------------------------------------------------------
PMSI Tunnel Attributes :
Tunnel-type    : Composite LDP P2MP IR
Flags          : Type: RNVE(0) BM: 0 U: 0 Leaf: not required
```

```
MPLS Label1 Ag : LABEL 0
MPLS Label2 IR : LABEL 524284
Root-Node      : 192.0.2.1              LSP-ID         : 8193
-------------------------------------------------------------------------------
---snip---
```

When leaf node PE-4 receives this IMET-P2MP-IR route, a provider tunnel is established toward the root. One P2MP LDP binding of opaque type GRT recursive is active on PE-4:

```
*A:PE-4# show router ldp bindings active p2mp summary ipv4
 No. of Generic IPv4 P2MP Active Bindings: 0
 No. of In-Band-SSM IPv4 P2MP Active Bindings: 0
 No. of In-Band-VPN-SSM IPv4 P2MP Active Bindings: 0
 No. of In-Band-SSM IPv4 P2MP Active Bindings: 0
 No. of VPN Recursive with Generic IPv4 P2MP Active Bindings: 0
 No. of GRT Recursive with Generic IPv4 P2MP Active Bindings: 1
```

The following GRT recursive P2MP LDP binding with root P-3 and inner root PE-1 is active on PE-4:

```
*A:PE-4# show router ldp bindings active p2mp opaque-type grt-recursive ipv4

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.4)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
===============================================================================
LDP GRT Recursive with Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id
InnerRootAddr                               Interface
RootAddr                                    Op
IngLbl                                      EgrLbl
EgrNH                                       EgrIf/LspId
-------------------------------------------------------------------------------
8193
192.0.2.1                                   73728
192.0.2.3                                   Pop
524283                                        --
  --                                          --

-------------------------------------------------------------------------------
No. of GRT Recursive with Generic IPv4 P2MP Active Bindings: 1
===============================================================================
```

The following detailed output shows that the P2MP type is 7:

```
*A:PE-4# show router ldp bindings active p2mp opaque-type grt-recursive ipv4 detail

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.4)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
==============================================================================
LDP GRT Recursive with Generic IPv4 P2MP Bindings (Active)
==============================================================================
-------------------------------------------------------------------------------
P2MP Type      : 7                 P2MP-Id    : 8193
Root-Addr      : 192.0.2.3
InnerRoot-Addr : 192.0.2.1
-------------------------------------------------------------------------------
Op             : Pop
Ing Lbl        : 524283
Egr Lbl        :   --
Egr Int/LspId  :   --
EgrNextHop     :   --
Egr. Flags     : None               Ing. Flags : None
===============================================================================
No. of GRT Recursive with Generic IPv4 P2MP Active Bindings: 1
==============================================================================
```

P-3 has two P2MP LDP bindings active: one toward the—downstream—lower FEC (LF) PE-4 and another to the—upstream—upper FEC (UF) P-2, as follows. Both P2MP LDP bindings have inner root 192.0.2.1 and they are stitched to each other.

```
*A:P-3# show router ldp bindings active p2mp opaque-type grt-recursive ipv4

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
===============================================================================
LDP GRT Recursive with Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id
InnerRootAddr                               Interface
RootAddr                                    Op
IngLbl                                      EgrLbl
EgrNH                                       EgrIf/LspId
-------------------------------------------------------------------------------
8193
192.0.2.1                                   Unknw
192.0.2.3 (LF)                              Push
  --                                        524283
192.168.34.2                                1/1/1

8193
192.0.2.1                                   Unknw
192.168.23.1 (UF)                           Swap
524279                                      Stitched
  --                                          --

-------------------------------------------------------------------------------
No. of GRT Recursive with Generic IPv4 P2MP Active Bindings: 2
```

```
===============================================================================
```

P-2 has two P2MP LDP bindings active: one GRT recursive (type 7) and one generic (type 1), as follows:

```
*A:P-2# show router ldp bindings active p2mp summary ipv4
 No. of Generic IPv4 P2MP Active Bindings: 1
 No. of In-Band-SSM IPv4 P2MP Active Bindings: 0
 No. of In-Band-VPN-SSM IPv4 P2MP Active Bindings: 0
 No. of In-Band-SSM IPv4 P2MP Active Bindings: 0
 No. of VPN Recursive with Generic IPv4 P2MP Active Bindings: 0
 No. of GRT Recursive with Generic IPv4 P2MP Active Bindings: 1
```

On P-2, the GRT recursive P2MP LDP binding with inner root 192.0.2.1 is toward LF P-3, as follows:

```
*A:P-2# show router ldp bindings active p2mp opaque-type grt-recursive ipv4

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
===============================================================================
LDP GRT Recursive with Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id
InnerRootAddr                               Interface
RootAddr                                    Op
IngLbl                                      EgrLbl
EgrNH                                       EgrIf/LspId
-------------------------------------------------------------------------------
8193
192.0.2.1                                   Unknw
192.168.23.1 (LF)                           Push
  --                                        524279
192.168.23.2                                1/1/1

-------------------------------------------------------------------------------
No. of GRT Recursive with Generic IPv4 P2MP Active Bindings: 1
===============================================================================
```

On P-2, the generic P2MP LDP binding is toward UF PE-1, as follows. The UF has root address 192.0.2.1 and is stitched to the LF with inner root address 192.0.2.1.

```
*A:P-2# show router ldp bindings active p2mp opaque-type generic ipv4

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
```

```
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                    Interface
RootAddr                                   Op
IngLbl                                     EgrLbl
EgrNH                                      EgrIf/LspId
-------------------------------------------------------------------------------
8193                                       Unknw
192.0.2.1 (UF)                             Swap
524279                                     Stitched
  --                                         --

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 1
===============================================================================
```

PE-1 has one P2MP LDP active binding toward LF P-2 (type 1- generic):

```
*A:PE-1# show router ldp bindings active p2mp opaque-type generic ipv4

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                    Interface
RootAddr                                   Op
IngLbl                                     EgrLbl
EgrNH                                      EgrIf/LspId
-------------------------------------------------------------------------------
8193                                       73728
192.0.2.1                                  Push
  --                                       524279
192.168.12.2                               1/1/1

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 1
===============================================================================
```

The EVPN BUM traffic is forwarded from the root node PE-1 to the leaf node PE-4 over the P2MP tree. The following command on root node PE-1 shows that an EVPN destination (that uses a BGP tunnel) toward leaf node PE-4 is established, and can carry multicast traffic (BUM):

```
*A:PE-1# show service id 1 evpn-mpls

===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address     Egr Label     Num. MACs   Mcast           Last Change
                Transport:Tnl                             Sup BCast Domain
-------------------------------------------------------------------------------
192.0.2.4       524284        0           bum             06/02/2021 08:31:14
```

```
                bgp:262146                                 No
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

The provider tunnel in VPLS 1 is established using LDP and the operational state is up, as follows. The router will always use the provider tunnel and not the EVPN-MPLS destination, as long as the provider tunnel Oper State is up:

```
*A:PE-1# show service id 1 provider-tunnel

===============================================================================
Service Provider Tunnel Information
===============================================================================
Type               : inclusive          Root and Leaf      : enabled
Admin State        : enabled            Data Delay Intvl   : 15 secs
PMSI Type          : ldp                LSP Template       :
Remain Delay Intvl : 0 secs             LSP Name used      : 8193
PMSI Owner         : bgpEvpnMpls
Oper State         : up                 Root Bind Id       : 32767
===============================================================================
```

The following SDP of type VplsPmsi is auto-created in VPLS 1 on root node PE-1:

```
*A:PE-1# show service id 1 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
32767:4294967294 VplsPmsi not applicable  Up      Up        None      3
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

The following **tools dump** command shows the originating provider tunnels for VPLS 1 on root node PE-1:

```
*A:PE-1# tools dump service id 1 provider-tunnels type originating

===============================================================================
VPLS 1 Inclusive Provider Tunnels Originating
===============================================================================
ipmsi (LDP)                                  P2MP-ID  Root-Addr
-------------------------------------------------------------------------------
8193                                         8193    192.0.2.1

-------------------------------------------------------------------------------
```

The following command shows the terminating provider tunnels for VPLS 1 on leaf node PE-4:

```
*A:PE-4# tools dump service id 1 provider-tunnels type terminating

===============================================================================
VPLS 1 Inclusive Provider Tunnels Terminating
===============================================================================
ipmsi (LDP)                                  P2MP-ID  Root-Addr
```

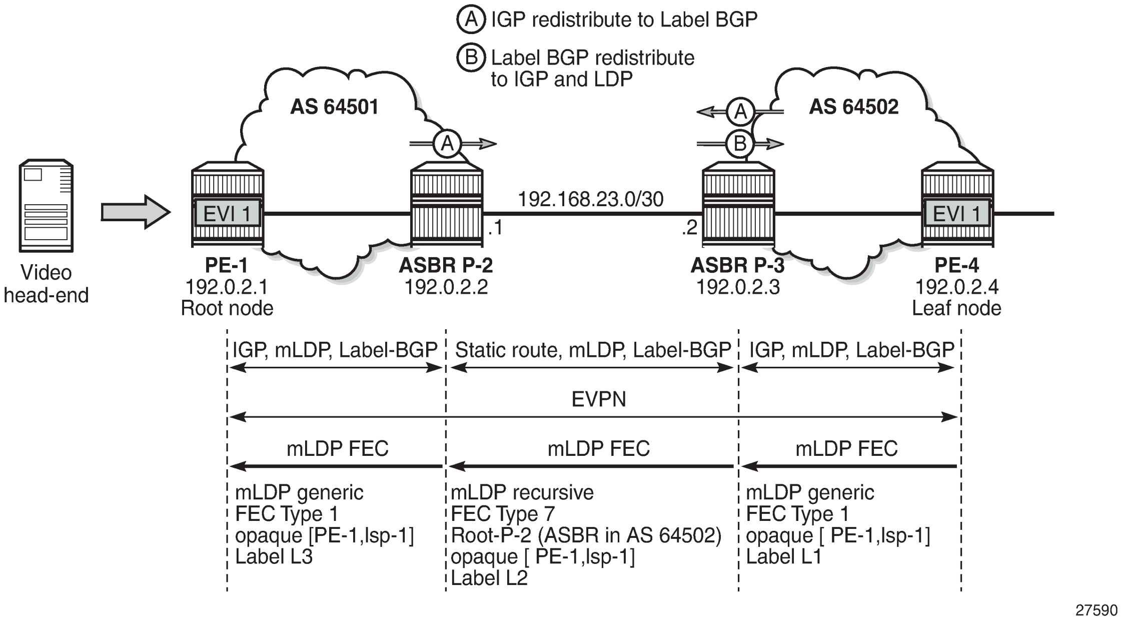
```
-------------------------------------------------------------------------------
                                                8193    192.0.2.1

-------------------------------------------------------------------------------
```

Optimized Inter-AS Model C for mLDP

When some leaf nodes do not support labeled BGP routes or recursive opaque mLDP label mapping, the ASBR in the AS where the leaf nodes are situated needs to leak the root IP address into the leaf PE IGP, which allows the leaf node PE-4 to send a generic FEC type 1 to join the root. The recursive opaque functionality is pushed to the local ASBR P-3.

Figure 233: Example topology for optimized Inter-AS Model C for mLDP shows the example topology for the optimized inter-AS model C for mLDP.

Figure 233: Example topology for optimized Inter-AS Model C for mLDP



The configuration starts with the configuration in the preceding section Inter-AS Model C for mLDP. The policy to export system prefixes from the ASs to labeled BGP is already configured and applied on both ASBRs. The following additional policies are defined on ASBR P-3 in the AS of the leaf node to export labeled BGP routes to OSPF and to LDP.

```
# on ASBR P-3:
configure
    router
        policy-options
            begin
            policy-statement "bgpToOspf"
                entry 10
                    from
                        protocol bgp-label
                    exit
                    to
```

```
                        protocol ospf
                    exit
                    action accept
                    exit
                exit
            exit
            policy-statement "bgpToLdp"
                entry 10
                    from
                        protocol bgp-label
                    exit
                    to
                        protocol ldp
                    exit
                    action accept
                    exit
                exit
            exit
            commit
```

Policy "bgpToOspf" is configured in the OSPF context and policy "bgpToLdp" in the **ldp** context, as follows:

```
# on ASBR P-3:
configure
    router Base
        ospf
            export "bgpToOspf"
        exit
        ldp
            export-tunnel-table "bgpToLdp"
        exit
```

Optimized Inter-AS Model C for mLDP - Verification

The prefixes from AS 64501 are now exported to OSPF and LDP in AS 64502; therefore, leaf node PE-4 will no longer use the labeled BGP routes to a node in AS 64501.

```
*A:PE-4# show router bgp routes label-ipv4
===============================================================================
 BGP Router ID:192.0.2.4        AS:64502       Local AS:64502
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP Routes
===============================================================================
Flag  Network                                            LocalPref   MED
      Nexthop (Router)                                   Path-Id     IGP Cost
      As-Path                                                        Label
-------------------------------------------------------------------------------
*i    192.0.2.1/32                                       100         10
      192.0.2.3                                          None        10
      64501                                                          524283
*i    192.0.2.2/32                                       100         None
      192.0.2.3                                          None        10
      64501                                                          524282
-------------------------------------------------------------------------------
```

```
Routes : 2
===============================================================================
```

The following route table in PE-4 shows that an OSPF route exists toward prefix 192.0.2.1:

```
*A:PE-4# show router route-table 192.0.2.1

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                  Remote  OSPF      00h00m21s  150
       192.168.34.1                                                 10
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

On PE-4, all tunnels are LDP tunnels; no BGP tunnels are established from PE-4 to PE-1 and P-2, as follows:

```
*A:PE-4# show router tunnel-table

===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination           Owner     Encap TunnelId  Pref   Nexthop        Metric
   Color
-------------------------------------------------------------------------------
192.0.2.1/32          ldp       MPLS  65538     9      192.168.34.1   10
192.0.2.2/32          ldp       MPLS  65539     9      192.168.34.1   1
192.0.2.3/32          ldp       MPLS  65537     9      192.168.34.1   10
-------------------------------------------------------------------------------
Flags: B = BGP or MPLS backup hop available
       L = Loop-Free Alternate (LFA) hop available
       E = Inactive best-external BGP route
       k = RIB-API or Forwarding Policy backup hop
===============================================================================
```

On all other nodes, the route table and tunnel table are the same as in the non-optimized scenario. The route table and the tunnel table for ASBR P-3 are as follows:

```
*A:P-3# show router route-table protocol bgp-label

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                  Remote  BGP_LABEL 00h10m48s  170
       192.168.23.1                                                0
192.0.2.2/32                                  Remote  BGP_LABEL 00h10m48s  170
       192.168.23.1                                                0
-------------------------------------------------------------------------------
No. of Routes: 2
```

```
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

```
*A:P-3# show router tunnel-table

===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination           Owner     Encap TunnelId  Pref   Nexthop        Metric
   Color
-------------------------------------------------------------------------------
192.0.2.1/32          bgp       MPLS  262145    12     192.168.23.1   1000
192.0.2.2/32          bgp       MPLS  262146    12     192.168.23.1   1000
192.0.2.4/32          ldp       MPLS  65537     9      192.168.34.2   10
-------------------------------------------------------------------------------
Flags: B = BGP or MPLS backup hop available
       L = Loop-Free Alternate (LFA) hop available
       E = Inactive best-external BGP route
       k = RIB-API or Forwarding Policy backup hop
===============================================================================
```

Root node PE-1 will send an IMET-P2MP-IR route to leaf node PE-4. PE-4 will send an mLDP label mapping message type 1 instead of type 7, because there is an LDP tunnel toward PE-1 instead of a BGP tunnel. The only P2MP mLDP binding on leaf node PE-4 is a generic P2MP binding, as follows:

```
*A:PE-4# show router ldp bindings p2mp summary ipv4
 No. of Generic IPv4 P2MP Bindings: 1
 No. of In-Band-SSM IPv4 P2MP Bindings: 0
 No. of In-Band-VPN-SSM IPv4 P2MP Bindings: 0
 No. of Recursive with In-Band-SSM IPv4 P2MP Bindings: 0
 No. of VPN Recursive with Generic IPv4 P2MP Bindings: 0
 No. of GRT Recursive with Generic IPv4 P2MP Bindings: 0
```

PE-4 sends the following mLDP label mapping message type 1 with root address 192.0.2.1 (PE-1) to its peer P-3.

```
15 2021/06/02 08:39:21.702 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Label Mapping packet (msgId 100) to 192.0.2.3:0
Protocol version = 1
Label 524279 advertised for the following FECs
P2MP: root = 192.0.2.1, T: 1, L: 4, TunnelId: 8193
"
```

The following generic P2MP mLDP binding for root address 192.0.2.1 is seen on PE-4:

```
*A:PE-4# show router ldp bindings p2mp opaque-type generic detail ipv4

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.4)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
==============================================================================
LDP Generic IPv4 P2MP Bindings
==============================================================================
-------------------------------------------------------------------------------
P2MP Type      : 1                 P2MP-Id    : 8193
Root-Addr      : 192.0.2.1
-------------------------------------------------------------------------------
Peer           : 192.0.2.3:0
Ing Lbl        : 524279U
Egr Lbl        :   --
Egr Int/LspId  :   --
EgrNextHop     :   --
Egr. Flags     : None              Ing. Flags : None
===============================================================================
No. of Generic IPv4 P2MP Bindings: 1
==============================================================================
```

ASBR P-3 receives the generic P2MP mLDP label mapping message from PE-4 (T: 1) and resolves the root node 192.0.2.1 to next-hop P-2. P-3 sends a GRT recursive P2MP mLDP label mapping message (T: 7) with inner root 192.0.2.1 to its peer P-2 (root 192.168.23.1) in AS 64501:

```
25 2021/06/02 08:39:21.696 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Recv Label Mapping packet (msgId 100) from 192.0.2.4:0
Protocol version = 1
Label 524279 advertised for the following FECs
P2MP: root = 192.0.2.1, T: 1, L: 4, TunnelId: 8193
"
```

```
26 2021/06/02 08:39:21.696 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Label Mapping packet (msgId 81) to 192.168.23.1:0
Protocol version = 1
Label 524278 advertised for the following FECs
P2MP: root = 192.168.23.1, T: 7, L: 17 (InnerRoot: 192.0.2.1 T: 1, L: 4, TunnelId: 8193)
"
```

```
*A:P-3# show router ldp bindings p2mp opaque-type generic detail ipv4

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
==============================================================================
LDP Generic IPv4 P2MP Bindings
==============================================================================
-------------------------------------------------------------------------------
P2MP Type      : 1                 P2MP-Id    : 8193
Root-Addr      : 192.0.2.1 (LF)
-------------------------------------------------------------------------------
Peer           : 192.0.2.4:0
Ing Lbl        :   --
```

```
Egr Lbl        : 524279
Egr Int/LspId  : 1/1/1
EgrNextHop     : 192.168.34.2
Egr. Flags     : None               Ing. Flags : None
Egr If Name    : int-P-3-PE-4
Metric         : 1                  Mtu        : 1564
===============================================================================
No. of Generic IPv4 P2MP Bindings: 1
==============================================================================
```

```
*A:P-3# show router ldp bindings p2mp opaque-type grt-recursive detail ipv4

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
==============================================================================
LDP GRT Recursive with Generic IPv4 P2MP Bindings
==============================================================================
-------------------------------------------------------------------------------
P2MP Type      : 7                  P2MP-Id    : 8193
Root-Addr      : 192.168.23.1 (UF)
InnerRoot-Addr : 192.0.2.1
-------------------------------------------------------------------------------
Peer           : 192.168.23.1:0
Ing Lbl        : 524278U
Egr Lbl        :   --
Egr Int/LspId  :   --
EgrNextHop     :   --
Egr. Flags     : None               Ing. Flags : None
===============================================================================
No. of GRT Recursive with Generic IPv4 P2MP Bindings: 1
==============================================================================
```

The P2MP mLDP bindings on P-2 and PE-1 are the same as in the previous non-optimized inter-AS model C for mLDP scenario. P-2 has one GRT recursive mLDP binding to P-3 and one generic mLDP binding to root node PE-1, whereas PE-1 only has a generic mLDP binding to P-2.

The following command on root node PE-1 shows that an EVPN-MPLS destination is created to the leaf node PE-4. This EVPN destination runs over a BGP tunnel and can transport multicast (BUM) traffic. However, as discussed in the preceding section, the EVPN destination is used for BUM traffic only in the case where the provider tunnel goes operationally down.

```
*A:PE-1# show service id 1 evpn-mpls

===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address     Egr Label     Num. MACs   Mcast           Last Change
                Transport:Tnl                             Sup BCast Domain
-------------------------------------------------------------------------------
192.0.2.4       524284        0           bum             06/02/2021 08:31:14
                bgp:262146                                No
-------------------------------------------------------------------------------
Number of entries : 1
```

```
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

The same command on the leaf node PE-4 shows an EVPN destination running on an LDP tunnel instead of a BGP tunnel. This destination is used whenever PE-4 needs to send BUM traffic to PE-1:

```
*A:PE-4# show service id 1 evpn-mpls

===============================================================================
BGP EVPN-MPLS Dest
===============================================================================
TEP Address     Egr Label     Num. MACs   Mcast           Last Change
                Transport:Tnl                             Sup BCast Domain
-------------------------------------------------------------------------------
192.0.2.1       524284        0           bum             06/02/2021 08:31:14
                ldp:65538                                 No
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

The other **show** commands in the Inter-AS Model C for mLDP section have an identical output for both scenarios.

Conclusion

P2MP mLDP is supported in inter-AS model C for EVPN-MPLS services with or without optimization. Optimization in this chapter refers to the ability to set up an end-to-end mLDP tunnel without the need for recursive opaque mLDP FECs on the leaf nodes. A similar configuration is applied in the case of seamless MPLS across different areas.

P2MP mLDP Tunnels for BUM Traffic in EVPN-MPLS Services

This chapter provides information about P2MP mLDP Tunnels for BUM Traffic in EVPN-MPLS Services.

Topics in this chapter include:

Applicability

This chapter was initially written for SR OS Release 14.0.R4, but the CLI in the current edition is based on SR OS Release 23.3.R3.

Point-to-Multipoint (P2MP) multicast Label Distribution Protocol (mLDP) tunnels for Broadcast, Unknown unicast, and Multicast (BUM) traffic in Ethernet Virtual Private Network Multiprotocol Label Switching (EVPN-MPLS) networks are supported in SR OS Release 14.0.R1, and later. Internet Group Management Protocol (IGMP) snooping support for EVPN-MPLS services is supported in SR OS Release 14.0.R4, and later.

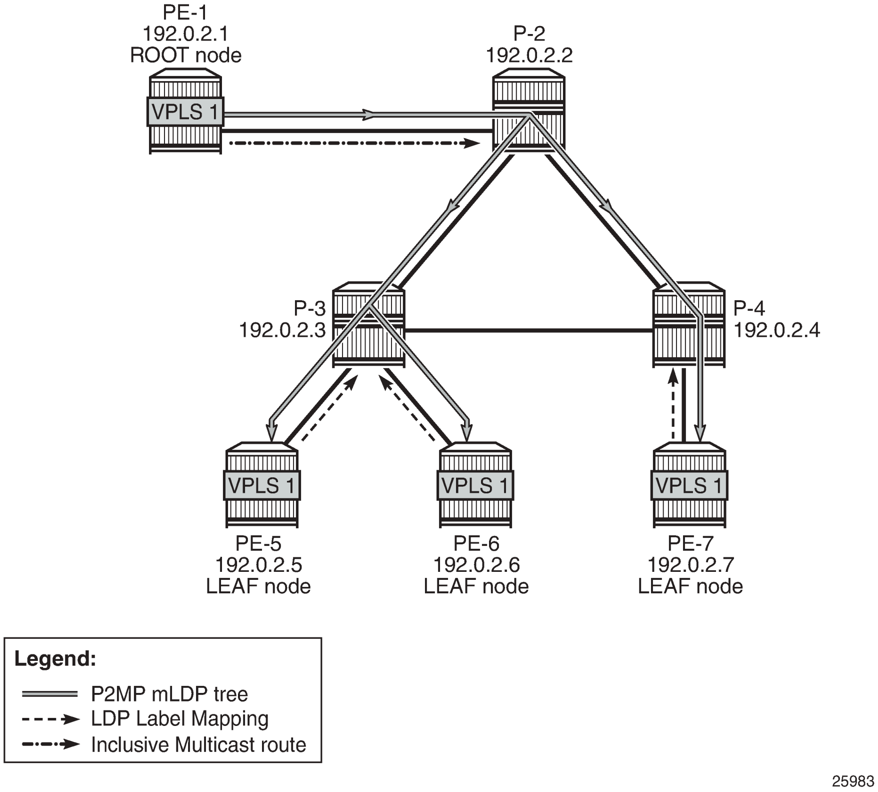
Overview

Service providers are moving their existing VPN services to EVPN. Providers using P2MP LSPs for VPLS services expect the same capabilities in EVPN. Before SR OS Release 14.0.R1, only Ingress Replication (IR) was supported. This works well for broadcast and unknown unicast traffic, but it is inefficient for multicast. Ingress replication does not use a multicast mechanism. Instead, the parent node makes n individual copies and unicasts each copy through an MPLS or IP tunnel to each child node.

BUM traffic is sent from a root node to a number of leaf nodes, but leaf nodes are also allowed to send BUM traffic to root nodes. If most BUM traffic is flowing from a few root nodes to leaf nodes, it would be inefficient to promote all leaf nodes to root-and-leaf nodes because of the amount of P2MP tunnels that would need to be set up. Another solution is to use a combination of P2MP mLDP and ingress replication (IR) tunnels in the service. The root nodes send BUM traffic using P2MP tunnels while the leaf nodes use IR tunnels to send BUM traffic to the root nodes. This avoids the need to set up a P2MP tree from each leaf, while it still allows leaf nodes to send BUM traffic to the root nodes.

Figure 234: P2MP mLDP tree with root node PE-1 and leaf nodes PE-5, PE-6, and PE-7 shows a multicast mLDP tree with root node PE-1 and leaf nodes PE-5, PE-6, and PE-7.

Figure 234: P2MP mLDP tree with root node PE-1 and leaf nodes PE-5, PE-6, and PE-7



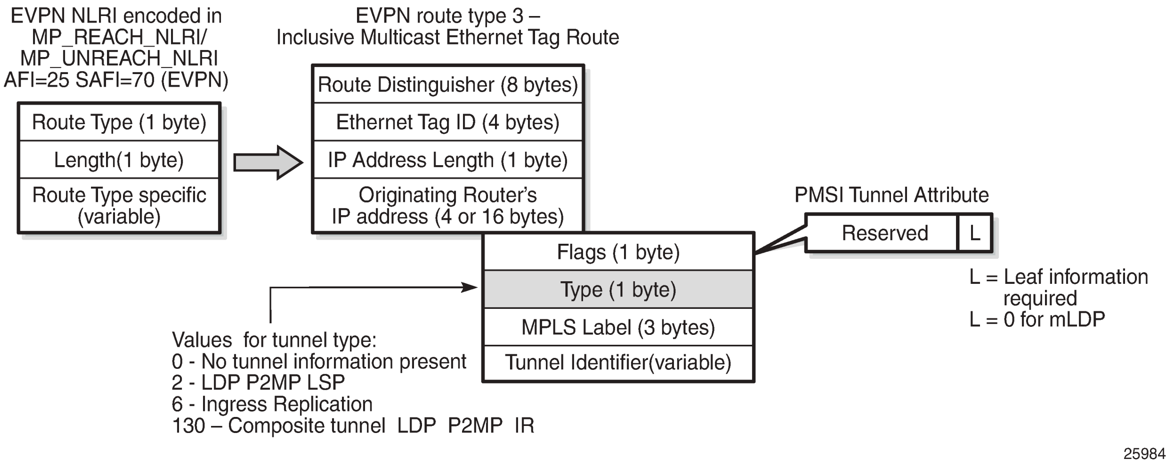
The Inclusive Multicast Ethernet Tag (IMET) route (EVPN route type 3) sent by root node PE-1 contains the required information to set up an mLDP tree, such as the root node IP address and an opaque value. As described in chapter "Multicast Label Distribution Protocol" in the *7450 ESS, 7750 SR, and 7950 XRS MPLS Advanced Configuration Guide for Classic CLI*, the mLDP tree is set up from the leaf nodes toward the root.

The LDP label mapping message contains the root node address, an opaque value, and an MPLS label. The leaf nodes send an LDP label mapping message to their upstream next hop toward the root node of the tree. Each transit node that has received such LDP label mapping message generates a new LDP label mapping message to its upstream next hop toward the root. This is repeated until the root node receives an LDP label mapping message and the multicast tree is completed.

Figure 234: P2MP mLDP tree with root node PE-1 and leaf nodes PE-5, PE-6, and PE-7 shows a P2MP mLDP tree rooted in PE-1, which is optimal for multicast traffic. However, no P2MP mLDP tree needs to be rooted in PE-5, PE-6, and PE-7 for the reverse direction. These three PEs can use IR to send traffic to the root (and to the other leaf nodes if needed).

EVPN route type 3 is used for setting up the flooding tree for a specified VPLS service. EVPN route type 3 includes the Provider Multicast Service Interface (PMSI) Tunnel Attribute (PMSI Tunnel Attribute = PTA), which can have different formats depending on the tunnel type; see Figure 235: BGP-EVPN route type 3 with PTA.

Figure 235: BGP-EVPN route type 3 with PTA


EVPN NLRI encoded in MP_REACH_NLRI/ MP_UNREACH_NLRI AFI=25 SAFI=70 (EVPN)
Route Type (1 byte)
Length(1 byte)
Route Type specific (variable)
EVPN route type 3 – Inclusive Multicast Ethernet Tag Route
Route Distinguisher (8 bytes)
Ethernet Tag ID (4 bytes)
IP Address Length (1 byte)
Originating Router's IP address (4 or 16 bytes)
Flags (1 byte)
Type (1 byte)
MPLS Label (3 bytes)
Tunnel Identifier(variable)
PMSI Tunnel Attribute
Reserved
L
L = Leaf information required
L = 0 for mLDP
Values for tunnel type:
0 - No tunnel information present
2 - LDP P2MP LSP
6 - Ingress Replication
130 – Composite tunnel LDP P2MP IR
25984

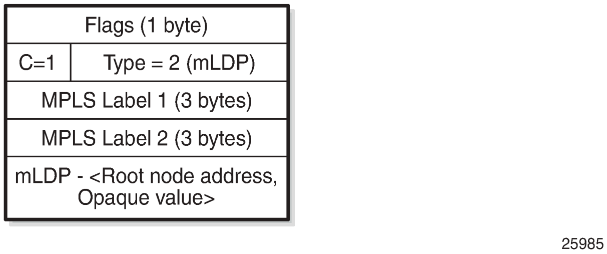
The following route values are used for EVPN-MPLS services:

- The route distinguisher (RD) is taken from the RD of the VPLS service, which can be configured in the BGP context or auto-derived from the BGP-EVPN EVPN Instance (EVI) value. In this case, the RD is auto-derived from the EVI, resulting in a value of 192.0.2.1:1 for VPLS 1 on PE-1.
- The Ethernet tag ID equals 0.
- The IP address length equals 32.
- The originating router's IP address carries the IPv4 system address.
- The PTA can have different formats depending on the tunnel type enabled in the service. The SR OS EVPN-MPLS implementation supports the following tunnel types (SR OS supports different tunnel types for EVPN-VXLAN):
  - Tunnel type 2 - P2MP mLDP
    - The route is referred to as an Inclusive Multicast Ethernet Tag Point-to-Multipoint (IMET-P2MP).
    - Flags: leaf not required.
    - The MPLS label is zero.
    - The tunnel identifier includes the root node address and an opaque number. This is the tunnel identifier that the leaf nodes use to join to the mLDP P2MP tree.
  - Tunnel type 6 - Ingress Replication (IR)
    - The route is referred to as an Inclusive Multicast Ethernet Tag Ingress Replication (IMET-IR).
    - Flags: leaf not required.
    - The MPLS label is a non-zero, downstream allocated label. This MPLS label is allocated to the service and is the same for unicast MAC/IP routes for the same service, unless **ingress-replication-bum-label** is configured in the service.
    - The tunnel identifier is the tunnel endpoint and is equal to the originating IP address.
  - Tunnel type 130 - Composite tunnel: Type: C-bit (composite) + type 2 (mLDP)
    - The route is referred to as an IMET-P2MP-IR.
    - Flags: leaf not required.

- MPLS label 1 equals zero.
- MPLS label 2 is a non-zero, downstream allocated label (as any other IR label). The leaf nodes use the label to set up an EVPN-MPLS binding to the root and add it to the default multicast list.
- The mLDP tunnel identifier is the root node address and an opaque number. This is the tunnel identifier that the leaf nodes use to join the mLDP P2MP tree.

Figure 236: PTA for composite tunnel IMET-P2MP-IR shows the PTA for tunnel type 130.

Figure 236: PTA for composite tunnel IMET-P2MP-IR



The composite bit C is set, indicating that the PTA identifies two tunnels: the transmit tunnel is a P2MP mLDP tunnel and the receive tunnel is an IR tunnel.

IMET-P2MP-IR routes

The composite tunnel type is an optimized solution that combines mLDP and IR within the same EVPN service so that each root node sends BUM traffic using the P2MP tunnel whereas each leaf-only node sends BUM traffic to the root node using IR.

- PEs configured with **root-and-leaf** can send all BUM traffic over P2MP mLDP tunnels while they receive BUM traffic either over P2MP mLDP tunnels (from other root-and-leaf nodes) or over ingress-replication tunnels (from leaf-only nodes).
- PEs configured with **no root-and-leaf** (default setting) can use IR to send BUM traffic to root nodes and other leaf-only nodes, while receiving BUM traffic over either P2MP mLDP tunnels (from root nodes) or ingress-replication tunnels (from leaf-only nodes).

The root PEs signal an IMET-P2MP-IR route, indicating that they intend to transmit BUM traffic using an mLDP P2MP tunnel, while they can receive traffic over an IR EVPN-MPLS binding. Composite tunnels reduce the number of P2MP mLDP tunnels that the PE/P routers in the EVI need to handle, because no full mesh of P2MP tunnels among all the PEs in the EVI is required. This is important (in terms of scaling) in services where there are just a pair of root nodes sending BUM in P2MP tunnels and hundreds of leaf nodes that only need to send BUM traffic to the root nodes using IR tunnels.

Configuration

Initial configuration

The PE and P nodes have the following initial configuration:

- The ports between the routers are configured as network ports and have router interfaces configured.
- IS-IS is enabled on all the router interfaces.
- LDP is enabled on all the router interfaces.
- BGP is enabled on all PEs with route reflector (RR) P-2. The BGP configuration on RR P-2 is as follows:

```
# On P-2:
configure
    router
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family evpn
                cluster 1.1.1.1
                peer-as 64500
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.5
                exit
                neighbor 192.0.2.6
                exit
                neighbor 192.0.2.7
                exit
            exit
            no shutdown
        exit
    exit
```

Configure EVPN P2MP mLDP in VPLS Service

On the root node PE-1, VPLS 1 is configured as follows:

```
# On PE-1:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            bgp
            exit
            bgp-evpn
                ingress-repl-inc-mcast-advertisement    # default setting
                evi 1
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            provider-tunnel
                inclusive
                    owner bgp-evpn-mpls
                    root-and-leaf
                    mldp
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/2/c3/1 create    # sap for ingress traffic from STC
            exit
            no shutdown
        exit
```

The configuration options in the **bgp-evpn** context of the VPLS are as follows:

```
*A:PE-1# configure service vpls 1 bgp-evpn ?
  - bgp-evpn
  - no bgp-evpn

 [no] accept-ivpls-e* - Configure to accept non-zero ethernet-tag MAC routes and process for
 CMAC flushing
 [no] arp-nd-extende* - Enable/disable ARP/ND Ext Community advertisement
 [no] cfm-mac-advert* - Enable/disable the advertisement of MEP, MIP, and VMEP MAC addresses
 over the BGP EVPN
 [no] evi             - EVPN Identifier
 [no] ignore-mtu-mis* - Configure ignore-mtu-mismatch
 [no] incl-mcast-l2-* - Configure BGP EVPN L2 attribute route advertisement
 [no] incl-mcast-ori* - Configure originating IP address
 [no] ingress-repl-i* - Configure BGP EVPN IMET-IR route advertisement
 [no] ip-route-adver* - Configure BGP EVPN IP Route Advertisement
      ip-route-link-* + Configure BGP EVPN IP Route Link Bandwidth
      isid-route-tar* + configure ISID route target information
 [no] mac-advertisem* - Configure BGP EVPN MAC Advertisement
      mac-duplication + Configure BGP EVPN MAC Duplication
 [no] mpls            + Configure BGP EVPN mpls
 [no] segment-routin* + Configure SRv6 instance
 [no] sel-mcast-adve* - Enable/disable selective multicast advertisements
 [no] unknown-mac-ro* - Configure BGP EVPN Unknown MAC Route
 [no] vxlan           + Configure BGP EVPN vxlan
```

By default, the advertisement of the inclusive multicast route with IR is enabled (**ingress-repl-inc-mcast-advertisement**). However, if it is disabled, the router does not send the IMET-IR or IMET-P2MP-IR routes, regardless of the service being enabled for BGP EVPN-MPLS or BGP EVPN-VXLAN.

For information about the other parameters in the **bgp-evpn** context of the VPLS, see chapters EVPN for VXLAN Tunnels (Layer 2) and EVPN for MPLS Tunnels.

The configuration options in the **provider-tunnel inclusive** context are as follows:

```
*A:PE-1# configure service vpls 1 provider-tunnel inclusive ?
  - inclusive

 [no] data-delay-int* - Configure data delay interval
 [no] mldp            - Enable/Disable MLDP
 [no] owner           - Configure provider-tunnel owner
 [no] root-and-leaf   - Configure LSP node type
 [no] rsvp            + Configure RSVP parameters
 [no] shutdown        - Administratively enable/disable the service
```

- The **data-delay-interval** is configured in seconds in the range from 3 to 180 seconds. A node configured with **root-and-leaf** sends all BUM packets (data plane and control plane: ARP, CCMs, and so on) to its provider tunnel after the delay-data-interval has expired. This timer keeps the provider tunnel operationally down until its expiration, and, during that time, the router can use the EVPN-MPLS destinations typically used for IR.

- mLDP is enabled by adding the keyword **mldp** and enabling the provider tunnel (**no shutdown**).
- The owner must be **bgp-evpn-mpls** if MPLS is enabled in the EVPN.

```
*A:PE-1# configure service vpls 1 provider-tunnel inclusive owner ?
  - no owner
  - owner {bgp-ad|bgp-vpls|bgp-evpn-mpls}
```

  Only one of the three possible owner protocols supports the provider tunnel in the service and needs to be set before the provider tunnel can be enabled. By default, no owner is configured. The following error is raised when a user wants to enable the provider tunnel without an owner:

```
*A:PE-1>config>service>vpls>provider-tunnel>inclusive# no shutdown
INFO: SVCMGR #6732 No owner configured for provider-tunnel
```

  After the provider tunnel has an owner and is enabled, the owner can only be changed when the provider tunnel is disabled.

```
*A:PE-1>config>service>vpls>provider-tunnel>inclusive# owner bgp-vpls
INFO: SVCMGR #6721 Provider tunnel is not shutdown
```

  After the owner is set, the corresponding protocol is checked to see if it is enabled in the service configuration.

```
*A:PE-1>config>service>vpls>provider-tunnel>inclusive# shutdown
*A:PE-1>config>service>vpls>provider-tunnel>inclusive# owner bgp-vpls
*A:PE-1>config>service>vpls>provider-tunnel>inclusive# no shutdown
MINOR: SVCMGR #6730 provider-tunnel cannot be enabled - bgp-vpls not enabled
```

- If **ingress-repl-inc-mcast-advertisement** is enabled and the PE is configured with **root-and-leaf**, the router sends an IMET-P2MP-IR route; if the PE is configured with **no root-and-leaf** (default), the router sends an IMET-IR route. However, if **ingress-repl-inc-mcast-advertisement** is disabled and the PE is configured with **root-and-leaf**, the router only sends IMET-P2MP routes. Leaf-only nodes do not send any IMET routes at all in case no IR multicast advertisement is allowed.

  Root-and-leaf nodes only send BUM traffic to the P2MP tunnel as long as it is active. If the P2MP tunnel goes operationally down, it starts sending BUM traffic to IR tunnels (EVPN-MPLS destinations shown in the **show service id 1 evpn-mpls** command).

- If a provider tunnel is configured on a node, the router can join P2MP trees as a leaf, by generating an LDP label mapping message including the corresponding P2MP mLDP FEC. If no provider tunnel is configured, the node does not join P2MP mLDP trees, and can only use IR for BUM.
- If one node is configured as root, all other nodes must be configured with provider tunnels; otherwise, they do not receive BUM traffic sent on P2MP tunnels. The configuration of leaf-only node PE-5 is as follows, the main difference with the configuration for the root being the **no root-and-leaf** (default setting):

```
# On PE-5:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            bgp
            exit
            bgp-evpn
```

```
                evi 1
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            provider-tunnel
                inclusive
                    owner bgp-evpn-mpls
                    mldp
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/2/c1/1:1 create    # sap for egress traffic to VPLS 1
            exit
            no shutdown
        exit
```

As described, the tunnel types for BUM traffic are controlled by **ingress-repl-inc-mcast-advertisement** and the **provider-tunnel** context (**root-and-leaf**). The IMET route sending behavior is summarized in Table 10: IMET routes and Tunnel Types advertised based on the configuration .

Table 10: IMET routes and Tunnel Types advertised based on the configuration

IMET route set	Root + Leaf PE	Leaf-only	No provider-tunnel
IR-mcast advertisement	Composite P2MP + IR	IR	IR
No IR-mcast advertisement	P2MP	-	-

Information about the provider tunnel can be retrieved as follows:

```
*A:PE-1# show service id 1 provider-tunnel

===============================================================================
Service Provider Tunnel Information
===============================================================================
Type               : inclusive         Root and Leaf      : enabled
Admin State        : enabled           Data Delay Intvl   : 15 secs
PMSI Type          : ldp               LSP Template       :
Remain Delay Intvl : 0 secs            LSP Name used      : 8193
PMSI Owner         : bgpEvpnMpls
Oper State         : up                Root Bind Id       : 32767
-------------------------------------------------------------------------------
Type               : selective         Wildcard SPMSI     : disabled
Admin State        : disabled          Data Delay Intvl   : 3 secs
PMSI Type          : none              Max P2MP SPMSI     : 10
PMSI Owner         : none
===============================================================================
```

Note:
The same IMET-P2MP route cannot be imported by two services at the same time. If two VPLS services (where a provider tunnel is enabled) have the same import route-target, only one service joins the mLDP tree (whichever comes first).

EVPN P2MP mLDP operation

After the root node and leaf nodes are configured as shown, the root node sends BGP EVPN composite IMET-P2MP-IR routes, as follows:

```
# On PE-1:
1 2023/07/03 12:23:01.864 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 93
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.1
        Type: EVPN-INCL-MCAST Len: 17 RD: 192.0.2.1:1, tag: 0, orig_addr len: 32, orig_addr:
 192.0.2.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:1
        bgp-tunnel-encap:MPLS
    Flag: 0xc0 Type: 22 Len: 25 PMSI:
        Tunnel-type Composite LDP P2MP IR (130)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf: not required]
        MPLS Label1 Ag  0
        MPLS Label2 IR 8388480
        Root-Node 192.0.2.1, LSP-ID 0x2001
"
```

The PTA for tunnel type 130 (composite tunnel) has two MPLS labels, of which MPLS label 1 equals zero. MPLS label 2 is used by the downstream nodes to set up the EVPN-MPLS destination to the root node and add it to the default multicast list. The actual MPLS label only uses the high-order 20 bits out of the 24 bits advertised in the MPLS label. Therefore, the value 8388480 needs to be divided by 16 to have the MPLS label: 8388480/16 = 524280. This is because the debug message is shown before the router can parse the label field and see whether it corresponds to an MPLS label (20 bits) or a VXLAN VNI (24 bits).

The tunnel identifier field contains the root node address 192.0.2.1 and the opaque value 0x2001, which corresponds to decimal value 8193. With this tunnel identifier, the leaf nodes can join the mLDP multicast tree toward the root node by sending LDP label mapping messages that contain the root node IP address and the opaque value.

Note:
When static P2MP mLDP tunnels and dynamic P2MP mLDP tunnels used by BGP-EVPN coexist on the same router, Nokia recommends that the static tunnels use a tunnel ID less than 8193. If a tunnel ID is statically configured with a value equal to or greater than 8193, BGP-EVPN may attempt to use the same tunnel ID for services with an enabled provider tunnel and fail to set up an mLDP tunnel.

The root node PE-1 receives IMET-IR routes from all leaf nodes, as shown for the BGP update sent by leaf node PE-5 (via RR P-2):

```
# On PE-1:
6 2023/07/03 12:26:12.751 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 91
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.5
        Type: EVPN-INCL-MCAST Len: 17 RD: 192.0.2.5:1, tag: 0, orig_addr len: 32, orig_addr:
 192.0.2.5
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.5
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        1.1.1.1
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:1
        bgp-tunnel-encap:MPLS
    Flag: 0xc0 Type: 22 Len: 9 PMSI:
        Tunnel-type Ingress Replication (6)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf: not required]
        MPLS Label 8388480
        Tunnel-Endpoint 192.0.2.5
"
```

The PTA tunnel type 6 for IR has only one MPLS label, which corresponds to the MPLS label 524280 allocated for the service. The tunnel identifier is the tunnel endpoint 192.0.2.5, which is the system address of the originating leaf node.

On leaf node PE-5, three BGP EVPN inclusive multicast routes have been learned and are used, as follows:

```
*A:PE-5# show router bgp routes evpn incl-mcast
===============================================================================
 BGP Router ID:192.0.2.5        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Flag  Route Dist.         OrigAddr
      Tag                 NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.1:1         192.0.2.1
      0                   192.0.2.1

u*>i  192.0.2.6:1         192.0.2.6
      0                   192.0.2.6

u*>i  192.0.2.7:1         192.0.2.7
      0                   192.0.2.7

-------------------------------------------------------------------------------
```

```
Routes : 3
================================================================================
```

The details of the BGP EVPN inclusive multicast route sent by root node PE-1 to leaf node PE-5 are as follows:

```
*A:PE-5# show router bgp routes evpn incl-mcast rd 192.0.2.1:1 detail
================================================================================
 BGP Router ID:192.0.2.5        AS:64500       Local AS:64500
================================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

================================================================================
BGP EVPN Inclusive-Mcast Routes
================================================================================
Original Attributes

Network        : n/a
Nexthop        : 192.0.2.1
Path Id        : None
From           : 192.0.2.2
Res. Nexthop   : 192.168.35.1
Local Pref.    : 100                    Interface Name : int-PE-5-P-3
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 30
Connector      : None
Community      : target:64500:1 bgp-tunnel-encap:MPLS
Cluster        : 1.1.1.1
Originator Id  : 192.0.2.1              Peer Router Id : 192.0.2.2
Flags          : Used Valid Best IGP
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : INCL-MCAST
Tag            : 0
Originator IP  : 192.0.2.1
Route Dist.    : 192.0.2.1:1
Route Tag      : 0
Neighbor-AS    : n/a
DB Orig Val    : N/A                    Final Orig Val : N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h35m54s
--------------------------------------------------------------------------------
PMSI Tunnel Attributes :
Tunnel-type    : Composite LDP P2MP IR
Flags          : Type: RNVE(0) BM: 0 U: 0 Leaf: not required
MPLS Label1 Ag : LABEL 0
MPLS Label2 IR : LABEL 524280
Root-Node      : 192.0.2.1              LSP-ID         : 8193
--------------------------------------------------------------------------------
---snip---
--------------------------------------------------------------------------------
Routes : 1
================================================================================
```

The MPLS label is 524280, as described. The LSP ID equals 8193, which corresponds to the hexadecimal value 0x2001 in the preceding BGP update message sent by the root node PE-1.

To set up the mLDP tree, leaf node PE-5 has generated an LDP label mapping message to the next hop router toward the root, P-3. The label mapping message includes the root address 192.0.2.1, the opaque value 8193, and MPLS label 524279, as follows:

```
*A:PE-5# show router ldp bindings active p2mp ipv4

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.5)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                     Interface
RootAddr                                    Op
IngLbl                                      EgrLbl
EgrNH                                       EgrIf/LspId
-------------------------------------------------------------------------------
8193                                        73728
192.0.2.1                                   Pop
524279                                        --
  --                                          --

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 1
---snip---
===============================================================================
```

P-3 has received two label mapping messages: one from PE-5 and one from PE-6. P-3 has sent one label mapping message to its upstream next hop P-2 with label 524279, as follows:

```
*A:P-3# show router ldp bindings active p2mp ipv4 opaque-type generic

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.3)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                     Interface
RootAddr                                    Op
IngLbl                                      EgrLbl
EgrNH                                       EgrIf/LspId
-------------------------------------------------------------------------------
8193                                        Unknw
192.0.2.1                                   Swap
524280                                      524279
```

```
192.168.35.2                                    1/1/c3/1

8193                                            Unknw
192.0.2.1                                       Swap
524280                                          524279
192.168.36.2                                    1/1/c4/1

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 2
===============================================================================
```

P-2 has received two label mapping messages: one from P-3 and one from P-4. P-2 has sent a label mapping message toward the root node PE-1 with label 524280, as follows:

```
*A:P-2# show router ldp bindings active p2mp ipv4 opaque-type generic

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.2)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                         Interface
RootAddr                                        Op
IngLbl                                          EgrLbl
EgrNH                                           EgrIf/LspId
-------------------------------------------------------------------------------
8193                                            Unknw
192.0.2.1                                       Swap
524280                                          524280
192.168.23.2                                    1/1/c2/1

8193                                            Unknw
192.0.2.1                                       Swap
524280                                          524280
192.168.24.2                                    1/1/c1/1

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 2
===============================================================================
```

When the LDP label reaches the root node PE-1, the mLDP tree is complete and it can be used for BUM traffic.

The following **tools** command shows the provider tunnels for VPLS 1 on root node and leaf nodes. On root node PE-1, there is one originating inclusive provider tunnel and there are no terminating inclusive provider tunnels, as follows:

```
*A:PE-1# tools dump service id 1 provider-tunnels

===============================================================================
VPLS 1 Inclusive Provider Tunnels Originating
===============================================================================
ipmsi (LDP)                                         P2MP-ID  Root-Addr
```

```
-------------------------------------------------------------------------------
8193                                            8193    192.0.2.1

-------------------------------------------------------------------------------

===============================================================================
VPLS 1 Inclusive Provider Tunnels Terminating
===============================================================================
ipmsi (LDP)                                     P2MP-ID  Root-Addr
-------------------------------------------------------------------------------

No Tunnels Found
-------------------------------------------------------------------------------
---snip---
```

On leaf node PE-5, no originating inclusive provider tunnels are established; only one terminating provider tunnel, as follows:

```
*A:PE-5# tools dump service id 1 provider-tunnels

===============================================================================
VPLS 1 Inclusive Provider Tunnels Originating
===============================================================================
ipmsi (LDP)                                     P2MP-ID  Root-Addr
-------------------------------------------------------------------------------

No Tunnels Found
-------------------------------------------------------------------------------

===============================================================================
VPLS 1 Inclusive Provider Tunnels Terminating
===============================================================================
ipmsi (LDP)                                     P2MP-ID  Root-Addr
-------------------------------------------------------------------------------
                                                8193    192.0.2.1

-------------------------------------------------------------------------------
---snip---
```

The inclusive provider tunnels are identified by the combination of the P2MP ID (opaque value) and the root address. These parameters are in every label mapping message and they are included in the PTA tunnel identifier for tunnel type 130 (IMET-P2MP-IR) and for tunnel type 2 (IMET-P2MP).

In VPLS 1 on root node PE-1, an SDP of type VplsPmsi is auto-created, as follows:

```
*A:PE-1# show service id 1 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
32767:4294967290 VplsPmsi not applicable  Up      Up        None      3
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

The detailed information about this SDP includes the traffic statistics: ingress/egress and forwarding/dropped, as follows:

```
*A:PE-1# show service id 1 sdp detail

===============================================================================
Services: Service Destination Points Details
===============================================================================
-------------------------------------------------------------------------------
 Sdp Id 32767:4294967290  -(not applicable)
-------------------------------------------------------------------------------
Description     : (Not Specified)
SDP Id             : 32767:4294967290         Type              : VplsPmsi
Split Horiz Grp    : (Not Specified)
Etree Root Leaf Tag: Disabled                 Etree Leaf AC     : Disabled
VC Type            : Ether                    VC Tag            : n/a
Admin Path MTU     : 9782                     Oper Path MTU     : 9782
Delivery           : MPLS
Far End            : not applicable           Tunnel Far End    : n/a
---snip---
PMSI Owner         : bgpEvpnMpls

Admin State        : Up                       Oper State        : Up
---snip---
Statistics            :
I. Fwd. Pkts.      : 0                        I. Dro. Pkts.     : 0
I. Fwd. Octs.      : 0                        I. Dro. Octs.     : 0
E. Fwd. Pkts.      : 49766                    E. Fwd. Octets    : 74449936
---snip---
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

IGMP snooping

When IGMP snooping is disabled and a multicast stream enters VPLS 1 on the root node, this stream is sent to all the leaf nodes, even if no receivers join the multicast group on the leaf nodes. In this example, a receiver connected to PE-5 joins a multicast group, but there are no receivers for any multicast group on PE-6 and PE-7. By default, IGMP is disabled and the multicast stream is flooded to all leaf PEs, as can be verified with the following monitor command on PE-6 where no receivers have joined any multicast stream:

```
*A:PE-6# monitor port 1/1/c1/1 1/2/c3/1 1/2/c1/1 1/2/c2/1 repeat 15 interval 4 | match "==|
Port|Input|--|time|Packets" expression
===============================================================================
Monitor statistics for Ports
===============================================================================
                                                   Input                 Output
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
At time t = 12 sec (Mode: Delta)
-------------------------------------------------------------------------------
Port 1/1/c1/1
-------------------------------------------------------------------------------
Packets                                             3293                      3
Port 1/2/c3/1
-------------------------------------------------------------------------------
Packets                                                0                      0
```

```
Port 1/2/c1/1
-------------------------------------------------------------------------------
Packets                                                0                   3290
Port 1/2/c2/1
-------------------------------------------------------------------------------
Packets                                             3290                      0
-------------------------------------------------------------------------------
At time t = 16 sec (Mode: Delta)
-------------------------------------------------------------------------------
Port 1/1/c1/1
-------------------------------------------------------------------------------
Packets                                             3293                      4
Port 1/2/c3/1
-------------------------------------------------------------------------------
Packets                                                0                      0
Port 1/2/c1/1
-------------------------------------------------------------------------------
Packets                                                0                   3289
Port 1/2/c2/1
-------------------------------------------------------------------------------
Packets                                             3289                      0
-------------------------------------------------------------------------------
---snip---
===============================================================================
```

This implies that bandwidth is wasted, which can be prevented by enabling IGMP snooping. IGMP snooping ensures that multicast traffic is only sent to the receivers that joined a multicast group. IGMP snooping can be enabled in VPLS 1 on all PEs, as follows:

```
configure service vpls 1 igmp-snooping no shutdown
```

A receiver connected to PE-5 has sent an IGMP report whereas PE-6 has no receivers that joined a multicast group. The traffic counters are monitored on the outgoing port to the (potential) receivers. On PE-5, traffic is sent to the receiver, as follows:

```
*A:PE-5# monitor port 1/1/c1/1 1/2/c3/1 1/2/c1/1 1/2/c2/1 repeat 15 interval 4 | match "==|
Port|Input|--|time|Packets" expression
===============================================================================
Monitor statistics for Ports
===============================================================================
                                                   Input                 Output
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
At time t = 12 sec (Mode: Delta)
-------------------------------------------------------------------------------
Port 1/1/c1/1
-------------------------------------------------------------------------------
Packets                                             3294                      3
Port 1/2/c3/1
-------------------------------------------------------------------------------
Packets                                                0                      0
Port 1/2/c1/1
-------------------------------------------------------------------------------
Packets                                                0                   3290
Port 1/2/c2/1
-------------------------------------------------------------------------------
Packets                                             3290                      0
---snip---
===============================================================================
```

On PE-6, no traffic is sent to any receiver, as follows:

```
*A:PE-6# monitor port 1/1/c1/1 1/2/c3/1 1/2/c1/1 1/2/c2/1 repeat 15 interval 4 | match "==|
Port|Input|--|time|Packets" expression
===============================================================================
Monitor statistics for Ports
===============================================================================
                                                   Input                 Output
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
At time t = 12 sec (Mode: Delta)
-------------------------------------------------------------------------------
Port 1/1/c1/1
-------------------------------------------------------------------------------
Packets                                             3295                      5
Port 1/2/c3/1
-------------------------------------------------------------------------------
Packets                                                0                      0
Port 1/2/c1/1
-------------------------------------------------------------------------------
Packets                                                0                      0
Port 1/2/c2/1
-------------------------------------------------------------------------------
Packets                                                0                      0
---snip---
===============================================================================
```

IGMP snooping can be enabled in EVPN-MPLS services with IR or provider-tunnel mLDP trees. When IGMP snooping is enabled on the VPLS, all the EVPN-MPLS destinations are added to the MFIB as a single router interface. IGMP queries and reports are properly forwarded to and from EVPN-MPLS destinations.

The following shows the EVPN-MPLS destinations as part of the MFIB when IGMP snooping is enabled:

```
*A:PE-5# show service id 1 mfib

===============================================================================
Multicast FIB, Service 1
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
*               *                     sap:1/2/c1/1:1               Local    Fwd
                                      mpls:192.0.2.1:524280        Local    Fwd
                                      mpls:192.0.2.6:524280        Local    Fwd
                                      mpls:192.0.2.7:524280        Local    Fwd
*               * (mac)               mpls:192.0.2.1:524280        Local    Fwd
                                      mpls:192.0.2.6:524280        Local    Fwd
                                      mpls:192.0.2.7:524280        Local    Fwd
-------------------------------------------------------------------------------
Number of entries: 2
===============================================================================
```

Connected to SAP 1/2/c1/1:1, PE-5 has a receiver that joined the multicast stream. EVPN-MPLS is added as a single logical IGMP snooping interface and treated as an mrouter, also on the other leaf nodes, as follows:

```
*A:PE-5# show service id 1 igmp-snooping base

===============================================================================
IGMP Snooping Base info for service 1
===============================================================================
Admin State : Up
Querier     : 172.16.0.5 on SAP 1/2/c1/1:1
SBD service : N/A
Evpn-proxy  : Disabled
-------------------------------------------------------------------------------
Port                       Oper MRtr Pim  Send Max  Max  Max   MVR       Num
Id                         Stat Port Port Qrys Grps Srcs Grp   From-VPLS Grps
                                                         Srcs
-------------------------------------------------------------------------------
sap:1/2/c1/1:1             Up   Yes  No   No   None None None  Local     0
evpn-mpls
                           Up   Yes  No   N/A  N/A  N/A  N/A   N/A       N/A
===============================================================================
```

On leaf node PE-5, the receiving host connected to SAP 1/2/c1/1:1 has IP address 172.16.0.5, as follows:

```
*A:PE-5# show service id 1 igmp-snooping mrouters

===============================================================================
IGMP Snooping Multicast Routers for service 1
===============================================================================
MRouter          Port Id                    Up Time        Expires   Version
-------------------------------------------------------------------------------
172.16.0.5       sap:1/2/c1/1:1             0d 00:04:19    160s      3
-------------------------------------------------------------------------------
Number of mrouters: 1
===============================================================================
```

On leaf node PE-6, SAP 1/2/c1/1:1 has no receiving host connected, but EVPN-MPLS is always added as an mrouter, as follows:

```
*A:PE-6# show service id 1 igmp-snooping base

===============================================================================
IGMP Snooping Base info for service 1
===============================================================================
Admin State : Up
Querier     : 172.16.0.5 on evpn-mpls
SBD service : N/A
Evpn-proxy  : Disabled
-------------------------------------------------------------------------------
Port                       Oper MRtr Pim  Send Max  Max  Max   MVR       Num
Id                         Stat Port Port Qrys Grps Srcs Grp   From-VPLS Grps
                                                         Srcs
-------------------------------------------------------------------------------
sap:1/2/c1/1:1             Up   No   No   No   None None None  Local     0
evpn-mpls
                           Up   Yes  No   N/A  N/A  N/A  N/A   N/A       N/A
===============================================================================
```

On PE-6, the only mrouter in the list is the receiving host connected to PE-5, with port ID EVPN-MPLS instead of a local SAP, as follows:

```
*A:PE-6# show service id 1 igmp-snooping mrouters

===============================================================================
IGMP Snooping Multicast Routers for service 1
===============================================================================
MRouter          Port Id                    Up Time        Expires   Version
```

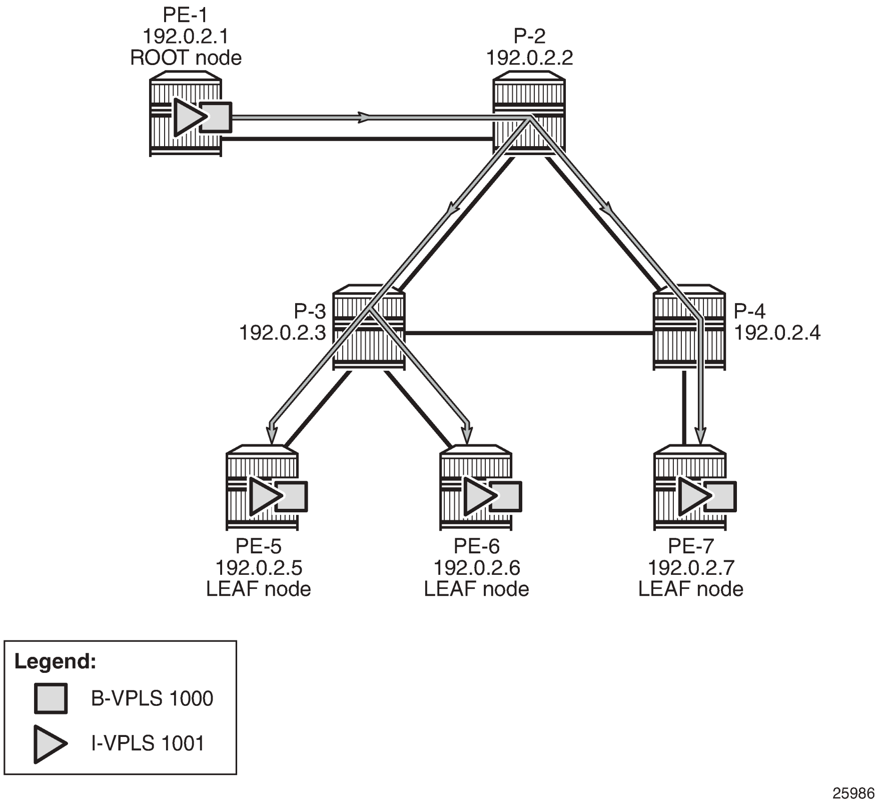
```
-------------------------------------------------------------------------------
172.16.0.5       evpn-mpls                 0d 00:04:21    158s      3
-------------------------------------------------------------------------------
Number of mrouters: 1
===============================================================================
```

PBB-EVPN and P2MP mLDP

Provider Backbone Bridging (PBB) EVPN is described in chapter EVPN for PBB over MPLS (PBB-EVPN).

Figure 237: P2MP mLDP in PBB-EVPN shows the setup for P2MP mLDP in PBB-EVPN.

Figure 237: P2MP mLDP in PBB-EVPN



P2MP mLDP tunnels can also be used in PBB-EVPN services. In Release 14.0, the use of **provider-tunnel inclusive mldp** is only for the default multicast list; no per-ISID IMET-P2MP routes are supported.

The Backbone (B) -VPLS still uses Multicast Forwarding Information Bases (MFIBs) for ISIDs using IR.

If an ISID policy is configured in the B-VPLS, a range of ISIDs configured with **use-def-mcast** use the P2MP tree, and a range of ISIDs configured with **advertise-local** make the router advertise IMET-IR routes for the local ISIDs in the range.

PE-1 is configured with **root-and-leaf**. The configuration for B-VPLS and I-VPLS is as follows:

```
# On PE-1:
configure
    service
```

```
        vpls 1000 name "B-VPLS 1000" customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:00:00:00:00:01
            exit
            bgp
            exit
            bgp-evpn
                evi 1000
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            provider-tunnel
                inclusive
                    owner bgp-evpn-mpls
                    root-and-leaf
                    mldp
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            isid-policy
                entry 10 create
                    use-def-mcast
                    no advertise-local
                    range 1001 to 2000
                exit
            exit
            no shutdown
        exit
        vpls 1001 name "I-VPLS 1001" customer 1 i-vpls create
            pbb
                backbone-vpls 1000
                exit
            exit
            stp
                shutdown
            exit
            sap 1/2/c3/1 create    # sap for ingress traffic from STC
            exit
            no shutdown
        exit
    exit
```

In this example, ISIDs in the range from 1001 to 2000 use the P2MP tree (**use-def-mcast**) and the router does not advertise the IMET-IR routes for the local ISIDs included in that range (**no advertise-local**). Any other local ISID advertises an IMET-IR and uses the MFIB to forward BUM packets to the remote EVPN-MPLS bindings created by IMET-IR routes.

The configuration on the leaf nodes PE-5, PE-6, and PE-7 is similar to the one for the root node, except for the **no root-and-leaf** setting (which is default), as follows:

```
# On PE-5:
configure
    service
        vpls 1000 name "B-VPLS 1000" customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:00:00:00:00:05
            exit
            bgp
            exit
            bgp-evpn
                evi 1000
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            provider-tunnel
                inclusive
                    owner bgp-evpn-mpls
                    mldp
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            isid-policy
                entry 10 create
                    use-def-mcast
                    no advertise-local
                    range 1001 to 2000
                exit
            exit
            no shutdown
        exit
        vpls 1001 name "I-VPLS 1001" customer 1 i-vpls create
            pbb
                backbone-vpls 1000
                exit
            exit
            stp
                shutdown
            exit
            sap 1/2/c1/1:1001 create    # sap for egress traffic to VPLS 1001
            exit
            no shutdown
        exit
    exit
```

A VPLS-PMSI SDP is auto-created in the B-VPLS at the root node, as follows:

```
*A:PE-1# show service id 1000 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
32767:4294967288 VplsPmsi not applicable  Up      Up        None      3
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

The default multicast list for the B-VPLS 1000 can be retrieved on root node and leaf nodes, for instance for leaf node PE-5, as follows:

```
*A:PE-5# tools dump service id 1000 evpn-mpls default-multicast-list
-------------------------------------------------------------------------
TEP Address                              Egr Label
                                         Transport
-------------------------------------------------------------------------
192.0.2.1                                524279
                                         ldp
192.0.2.6                                524279
                                         ldp
192.0.2.7                                524279
                                         ldp
-------------------------------------------------------------------------
```

IGMP snooping can be enabled in the I-VPLS 1001 on all PEs, as follows:

```
configure service vpls 1001 igmp-snooping no shutdown
```

After IGMP snooping is enabled, the multicast stream is not flooded anymore to any receivers until they send an IGMP report for the multicast stream.

On each PE, the logical interface B-EVPN-MPLS is added as a single IGMP snooping interface and treated as an mrouter, as follows:

```
*A:PE-5# show service id 1001 igmp-snooping base

===============================================================================
IGMP Snooping Base info for service 1001
===============================================================================
Admin State : Up
Querier     : 172.16.0.55 on SAP 1/2/c1/1:1001
SBD service : N/A
Evpn-proxy  : Disabled
-------------------------------------------------------------------------------
Port                       Oper MRtr Pim  Send Max  Max  Max   MVR       Num
Id                         Stat Port Port Qrys Grps Srcs Grp   From-VPLS Grps
                                                         Srcs
-------------------------------------------------------------------------------
b-evpn-mpls
                           Up   Yes  No   N/A  N/A  N/A  N/A   N/A       N/A
sap:1/2/c1/1:1001          Up   Yes  No   No   None None None  Local     0
===============================================================================
```

PE-5 has a receiver that sent an IGMP report for a multicast group in I-VPLS 1001 on SAP 1/2/c1/1:1001 and this SAP is an mrouter port. On PE-6, there is no receiver that sent IGMP reports; therefore, the only mrouter port corresponds to the B-EVPN-MPLS logical interface, as follows:

```
*A:PE-6# show service id 1001 igmp-snooping base

===============================================================================
IGMP Snooping Base info for service 1001
===============================================================================
Admin State : Up
Querier     : 172.16.0.55 on evpn-mpls
SBD service : N/A
Evpn-proxy  : Disabled
-------------------------------------------------------------------------------
Port                       Oper MRtr Pim  Send Max  Max  Max   MVR       Num
```

```
Id                         Stat Port Port Qrys Grps Srcs Grp   From-VPLS Grps
                                                         Srcs
-------------------------------------------------------------------------------
b-evpn-mpls
                           Up   Yes  No   N/A  N/A  N/A  N/A   N/A       N/A
sap:1/2/c1/1:1001          Up   No   No   No   None None None  Local     0
===============================================================================
```

PE-5 has a local mrouter 172.16.0.55 on SAP 1/2/c1/1:1001, as follows:

```
*A:PE-5# show service id 1001 igmp-snooping mrouters

===============================================================================
IGMP Snooping Multicast Routers for service 1001
===============================================================================
MRouter          Port Id                    Up Time        Expires   Version
-------------------------------------------------------------------------------
172.16.0.55      sap:1/2/c1/1:1001          0d 00:03:41    199s      3
-------------------------------------------------------------------------------
Number of mrouters: 1
===============================================================================
```

On PE-6, mrouter 172.16.0.55 is not local; therefore, the EVPN-MPLS logical interface is used, as follows:

```
*A:PE-6# show service id 1001 igmp-snooping mrouters

===============================================================================
IGMP Snooping Multicast Routers for service 1001
===============================================================================
MRouter          Port Id                    Up Time        Expires   Version
-------------------------------------------------------------------------------
172.16.0.55      evpn-mpls                  0d 00:03:43    197s      3
-------------------------------------------------------------------------------
Number of mrouters: 1
===============================================================================
```

Conclusion

Service providers are migrating their existing VPN services to EVPN and expect at least the same capabilities in EVPN, including the forwarding of BUM traffic. Ingress replication is a good mechanism for broadcast and unknown unicast traffic in EVPN networks, but not efficient for multicast applications. EVPN P2MP mLDP offers efficiency for multicast, using composite tunnels combining the benefits of P2MP mLDP and IR.

PBB-Epipe

This chapter provides information about Provider Backbone Bridging (PBB) — Ethernet Virtual Leased Line in an MPLS-based network which is applicable to SR OS.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS Release 7.0.R5. The CLI in the current edition corresponds to SR OS Release 20.10.R2. There are no specific prerequisites.

Overview

RFC 7041, *Extensions to VPLS PE model for Provider Backbone Bridging*, describes the PBB-VPLS model supported by SR OS. This model expands the VPLS PE model to support PBB as defined by the IEEE 802.1ah.

The PBB model is organized around a B-component (backbone instance) and an I-component (customer instance). In Nokia’s implementation of the PBB model, the use of an Epipe as I-component is allowed for point-to-point services. Multiple I-VPLS and Epipe services can be all mapped to the same B-VPLS (backbone VPLS instance).

The use of Epipe scales the E-Line services because no MAC switching, learning, or replication is required in order to deliver the point-to-point service. All packets ingressing the customer SAP are PBB-encapsulated and unicasted through the B-VPLS tunnel using the backbone destination MAC of the remote PBB PE. All the packets egressing the B-VPLS destined for the Epipe are PBB de-encapsulated and forwarded to the customer SAP.

Some use cases for PBB-Epipe are:

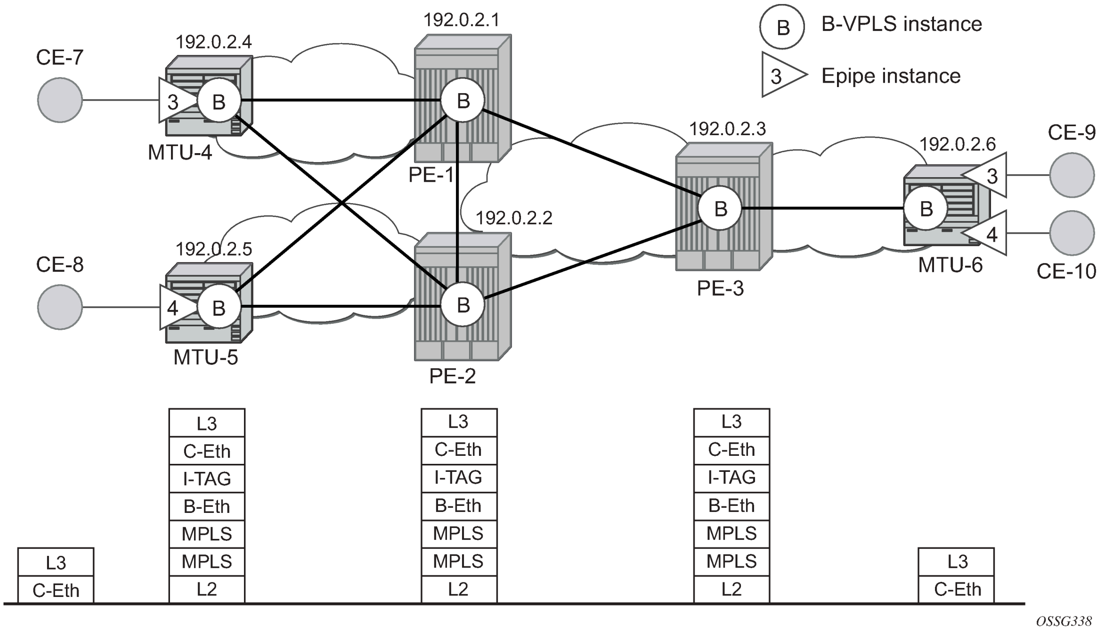
- Get a more efficient and scalable solution for point-to-point services:
  - Up to 8K VPLS services per box are supported (including I-VPLS or B-VPLS) and using I-VPLS for point-to-point services takes VPLS resources as well as unnecessary customer MAC learning. A better solution is to connect a PBB-Epipe to a B-VPLS instance, where there is no customer MAC switching/learning.
- Take advantage of the pseudowire aggregation in the M:1 model:
  - Many Epipe services may use only a single service and set of pseudowires over the backbone.
- Have a uniform provisioning model for both point-to-point (Epipe) and multipoint (VPLS) services.

- Using the PBB-Epipe, the core MPLS/pseudowire infrastructure does not need to be modified: the new Epipe inherits the existing pseudowire and MPLS structure already configured on the B-VPLS and there is no need for configuring new tunnels or pseudowire switching instances at the core.

Knowledge of the PBB-VPLS architecture and functionality on the service router family is assumed throughout this section. For additional information, see the relevant Nokia user documentation.

Figure 238: Example topology shows the example topology that is used throughout the rest of the chapter.

Figure 238: Example topology



The setup consists of a three SR OS routers in the core (PE-1, PE-2, and PE-3) and three Multi-Tenant Unit (MTU) nodes connected to the core. A backbone VPLS instance (B-VPLS 101) will be defined in all the six nodes, whereas two Epipe services will be defined as illustrated in Figure 238: Example topology (Epipe 3 in nodes MTU-4 and MTU-6, Epipe 4 in nodes MTU-5 and MTU-6). Those Epipe services will be multiplexed into the common B-VPLS 101, using the I-Service ID (ISID) field within the I-TAG as the demultiplexer field required at the egress MTU to differentiate each specific customer. I-VPLS and Epipe services can be mapped to the same B-VPLS.

The B-VPLS domain constitutes a H-VPLS network itself, with spoke-SDPs from the MTUs to the core PE layer. Active/standby (A/S) spoke-SDPs can be used from the MTUs to the PEs (like in the MTU-4 and MTU-5 cases) or single non-redundant spoke-SDPs (like MTU-6).

The protocol stack being used along the path between the CEs is represented in Figure 238: Example topology.

Configuration

This section describes all the relevant PBB-Epipe configuration tasks for the setup shown in Figure 238: Example topology. The appropriate B-VPLS and associated IP/MPLS configuration is out of the scope of this document. In this particular example, the following protocols will be configured beforehand in the core:

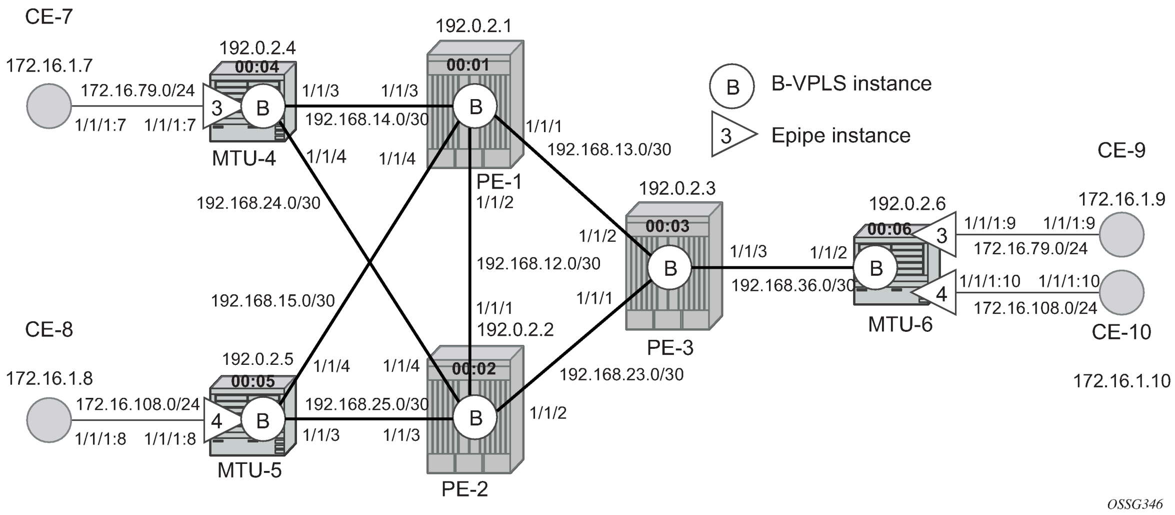
- ISIS-TE as IGP with all the interfaces being level-2. Alternatively, OSPF could have been used.
- RSVP-TE as the MPLS protocol to signal the transport tunnels.
- LSPs between core PEs will be fast re-route protected (facility bypass tunnels) whereas LSP tunnels between MTUs and PEs will not be protected.
- The protection between MTU-4, MTU-5 and PE-1, PE-2 will be based on the A/S pseudowire protection configured in the B-VPLS.
- BGP is configured for auto-discovery—BGP-AD (Layer 2 VPN family), because FEC 129 will be used to establish the pseudowires between PEs in the core (FEC 128 between MTU and PE nodes).

Once the IP/MPLS infrastructure is up and running, the service configuration tasks described in the following sections can be implemented.

PBB Epipe service configuration

Figure 239: Setup detailed view shows an example where the Epipes 3 and 4 are using the B-VPLS 101 in the core. The same B-VPLS which is multiplexing the Epipe services into a common service provider infrastructure can also be used to connect the I-VPLS instances existing in the network for multipoint services.

Figure 239: Setup detailed view



B-VPLS and PBB configuration

First, configure the B-VPLS instance that will carry the PBB traffic. There is no specific requirement on the B-VPLS to support Epipes. The following shows the B-VPLS configuration on MTU-4 and PE-1.

```
# on MTU-4:
configure
    service
        vpls 101 name "B-VPLS 101" customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:04:04:04:04:04
            exit
            endpoint "core" create
                no suppress-standby-signaling
            exit
            spoke-sdp 41:101 endpoint "core" create
                precedence primary
            exit
            spoke-sdp 42:101 endpoint "core" create
            exit
            no shutdown
        exit
```

```
# on PE-1:
configure
    service
        pw-template 1 use-provisioned-sdp create
            split-horizon-group "CORE"
            exit
        exit
        vpls 101 name "B-VPLS 101" customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:01:01:01:01:01
            exit
            bgp
                route-target export target:65000:101 import target:65000:101
                pw-template-binding 1
                exit
            exit
            bgp-ad
                vpls-id 65000:101
                no shutdown
            exit
            spoke-sdp 14:101 create
            exit
            spoke-sdp 15:101 create
            exit
            no shutdown
        exit
```

The relevant B-VPLS commands are in **bold**.

The keyword **b-vpls** is given at creation time and therefore it cannot be added to an existing regular VPLS instance. Besides the **b-vpls** keyword, the B-VPLS is a regular VPLS instance in terms of configuration, with the following exceptions:

- The B-VPLS service MTU must be at least 18 bytes greater than the Epipe MTU of the multiplexed instances. In this example, the I-VPLS instances will have the default service MTU (1514 bytes), therefore, any MTU equal or greater than 1532 bytes must be configured. In this particular example, an MTU of 2000 bytes is configured in the B-VPLS instance throughout the network.
- The source B-MAC is the MAC that will be used as a source when the PBB traffic is originated from that node. It is possible to configure a source B-MAC per B-VPLS instance (if there are more than one B-VPLS) or a common source B-MAC that will be shared by all the B-VPLS instances in the node. A common B-MAC is configured as follows:

  ```
  # on MTU-4:
  configure
      service
          pbb
  ```

```
            source-bmac 00:04:04:04:04:04
```

```
# on MTU-5:
configure
    service
        pbb
            source-bmac 00:05:05:05:05:05
```

```
# on MTU-6:
configure
    service
        pbb
            source-bmac 00:06:06:06:06:06
```

The following considerations will be taken into account when configuring the B-VPLS:

- B-VPLS SAPs:
  - Ethernet null, dot1q, and qinq encapsulations are supported.
  - Default SAP types are blocked in the CLI for the B-VPLS SAP.
- B-VPLS SDPs:
  - For MPLS, both mesh and spoke-SDPs with split-horizon groups are supported.
  - Similar to regular pseudowire, the outgoing PBB frame on an SDP (for example, Bpseudowire) contains a BVID q-tag only if the pseudowire type is Ethernet VLAN (vc-type=vlan). If the pseudowire type is Ethernet (vc-type=ether), the BVID q-tag is stripped before the frame goes out.
  - BGP-AD is supported in the B-VPLS, therefore, spoke-SDPs in the B-VPLS can be signaled using FEC 128 or FEC 129. In this example, BGP-AD and FEC 129 are used. A split-horizon group has been configured to emulate the behavior of mesh SDPs in the core.
- While Multiple MAC Registration Protocol (MMRP) is useful to optimize the flooding in the B-VPLS domain and build a flooding tree on a per I-VPLS basis, it does not have any effect for Epipes because the destination B-MAC used for Epipes is always the destination B-MAC configured in the Epipe and never the group B-MAC corresponding to the ISID.
- If a local Epipe instance is associated with the B-VPLS, local frames originated or terminated on local Epipe(s) are PBB encapsulated or de-encapsulated using the PBB Etype provisioned under the related port or SDP component.

By default, the PBB Etype is 0x88e7 (which is the standard one defined in the 802.1ah, indicating that there is an I-TAG in the payload) but this PBB Etype can be changed if required due to interoperability reasons. This is the way to change it at port and/or SDP level:

```
A:MTU-4# configure port 1/1/3 ethernet pbb-etype
  - pbb-etype <0x0600..0xffff>
  - no pbb-etype

 <0x0600..0xffff>     : [1536..65535] - accepts in decimal or hex
```

```
A:MTU-4# configure service sdp 41 pbb-etype
  - no pbb-etype [<0x0600..0xffff>]
  - pbb-etype <0x0600..0xffff>

 <0x0600..0xffff>     : [1536..65535] - accepts in decimal or hex
```

The following commands are useful to check the actual PBB Etype.

```
A:MTU-4# show service sdp 41 detail | match PBB
Bw BookingFactor     : 100                   PBB Etype          : 0x88e7
```

```
A:MTU-4# show port 1/1/3 | match PBB
PBB Ethertype      : 0x88e7
```

Before configuring the Epipe itself, the operator can optionally configure MAC names under the PBB context. MAC names will simplify the Epipe provisioning later on and in case of any change on the remote node MAC address, only one configuration modification is required as opposed as one change per affected Epipe (potentially thousands of Epipes which are terminated onto the same remote node). The MAC names are configured in the service PBB CLI context:

```
*A:MTU-4# configure service pbb mac-name
  - mac-name <name> <ieee-address>
  - no mac-name <name>

 <name>               : 32 char max
 <ieee-address>       : xx:xx:xx:xx:xx:xx or xx-xx-xx-xx-xx-xx
```

```
# on all nodes:
configure
    service
        pbb
            mac-name "MTU-4" 00:04:04:04:04:04
            mac-name "MTU-5" 00:05:05:05:05:05
            mac-name "MTU-6" 00:06:06:06:06:06
```

It is not required to configure a node with its own MAC address, so on MTU-4, the line defining the mac-name MTU-4 can be omitted.

Epipe configuration

Once the common B-VPLS is configured, the next step is the provisioning of the customer Epipe instances. For PBB-Epipes, the I-component or Epipe is composed of an I-SAP and a PBB tunnel endpoint which points to the backbone destination MAC address (B-DA).

The following outputs show the relevant CLI configuration for the two Epipe instances represented in Figure 239: Setup detailed view. The Epipe instances are configured on the MTU devices, whereas the core PEs are kept as customer-unaware nodes.

Epipes 3 and 4 are configured on MTU-6 as follows:

```
# on MTU-6:
configure
    service
        epipe 3 name "Epipe 3" customer 1 create
            description "pbb epipe number 3"
            pbb
                tunnel 101 backbone-dest-mac "MTU-4" isid 3
            exit
            sap 1/1/1:9 create
            exit
            no shutdown
        exit
        epipe 4 name "Epipe 4" customer 1 create
            description "pbb epipe number 4"
            pbb
                tunnel 101 backbone-dest-mac "MTU-5" isid 4
            exit
            sap 1/1/1:10 create
            exit
            no shutdown
        exit
```

```
# on MTU-4:
configure
    service
        epipe 3 name "Epipe 3" customer 1 create
            description "pbb epipe number 3"
            pbb
                tunnel 101 backbone-dest-mac "MTU-6" isid 3
            exit
            sap 1/1/1:7 create
            exit
            no shutdown
        exit
```

```
# on MTU-5:
configure
    service
        epipe 4 name "Epipe 4" customer 1 create
            description "pbb epipe number 4"
            pbb
                tunnel 101 backbone-dest-mac "MTU-6" isid 4
            exit
            sap 1/1/1:8 create
            exit
            no shutdown
        exit
```

All Ethernet SAPs supported by a regular Epipe are also supported in the PBB Epipe. spoke-SDPs are not supported in PBB-Epipes, for example, no spoke-SDP is allowed when PBB tunnels are configured on the Epipe.

The PBB tunnel links the SAP configured to the B-VPLS 101 existing in the core. The following parameters are accepted in the PBB tunnel configuration:

```
A:MTU-5# configure service epipe 4 pbb tunnel
  - no tunnel
  - tunnel <service-id> backbone-dest-mac <mac-name> isid <ISID>
  - tunnel <service-id> backbone-dest-mac <ieee-address> isid <ISID>

 <service-id>         : [1..2148007978]|<svc-name:64 char max>
 <mac-name>           : 32 char max
 <ieee-address>       : xx:xx:xx:xx:xx:xx or xx-xx-xx-xx-xx-xx
 <ISID>               : [0..16777215]
```

Where:

- The service-id matches the B-VPLS ID.
- The **backbone-dest-mac** can be given by a MAC name (as in this configuration example) or the MAC address itself. It is recommended to use MAC names, as explained in the previous section.
- The ISID must be specified.

Flood avoidance in PBB-Epipes

As already discussed in the previous section, when provisioning a PBB Epipe, the remote **backbone-dest-mac** must be explicitly configured on the PBB tunnel so that the ingress PBB node can build the 802.1ah encapsulation.

If the configured remote backbone destination MAC address is not known in the local FDB, the Epipe customer frames will be 802.1ah encapsulated and flooded into the B-VPLS until the MAC is learned. As previously stated, MMRP does not help to minimize the flooding because the PBB Epipes always use the configured **backbone-destination-mac** for flooding traffic as opposed to the group B-MAC derived from the ISID.

Flooding could be indefinably prolonged in the following cases:

- Configuration mistake of the **backbone-destination-mac**. The service will not work, but the operator will not detect the mistake, because the customer traffic is not dropped at the source node. Every single frame is turned into an unknown unicast PBB frame and therefore flooded into the B-VPLS domain.
- Change the **backbone-smac** in the remote PE B-VPLS instance.
- There is only unidirectional traffic in the Epipe service. In this case, the backbone-dest-mac will never be learned in the local FIB and the frames will always be flooded into the B-VPLS domain.
- The remote node owning the **backbone-destination-mac** simply goes down.

In any of those cases, the operator can easily check whether the PBB Epipe is flooding into the B-VPLS domain, just by looking at the flood flag in the following command output:

```
*A:MTU-4# show service id 3 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 3                   Vpn Id            : 0
Service Type      : Epipe
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/1:7                              q-tag        1518    1518    Up   Up

-------------------------------------------------------------------------------
PBB Tunnel Point
-------------------------------------------------------------------------------
B-vpls     Backbone-dest-MAC Isid     AdmMTU OperState Flood Oper-dest-MAC
-------------------------------------------------------------------------------
101        MTU-6             3        2000   Up        Yes   00:06:06:06:06:06
-------------------------------------------------------------------------------
Last Status Change: 01/05/2021 16:03:03
Last Mgmt Change  : 01/05/2021 16:03:03
===============================================================================
```

In this particular example, the PBB Epipe 3 is flooding into the B-VPLS 101, as the flood flag indicates. The operator can also confirm that the operational destination B-MAC for the PBB tunnel, MTU-6, has not been learned in the B-VPLS FDB:

```
*A:MTU-4# show service id 101 fdb pbb
```

```
===============================================================================
Forwarding Database, b-Vpls Service 101
===============================================================================
MAC               Source-Identifier     iVplsMACs  Epipes     Type/Age
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

In small B-VPLS environments (up to 20 B-VPLSs, each with 10 MC-LAGs), it is possible to configure the PBB V-VPLS MAC notification mechanism to send notification messages at regular intervals (using the renotify parameter), rather than being only event-driven. This can avoid flooding into the B-VPLS.

Flooding cases 1 and 2 — Wrong backbone-dest-mac

Flooding cases 1 and 2 should be fixed after detecting the flooding (see previous commands) and checking the FDBs and PBB tunnel configurations.

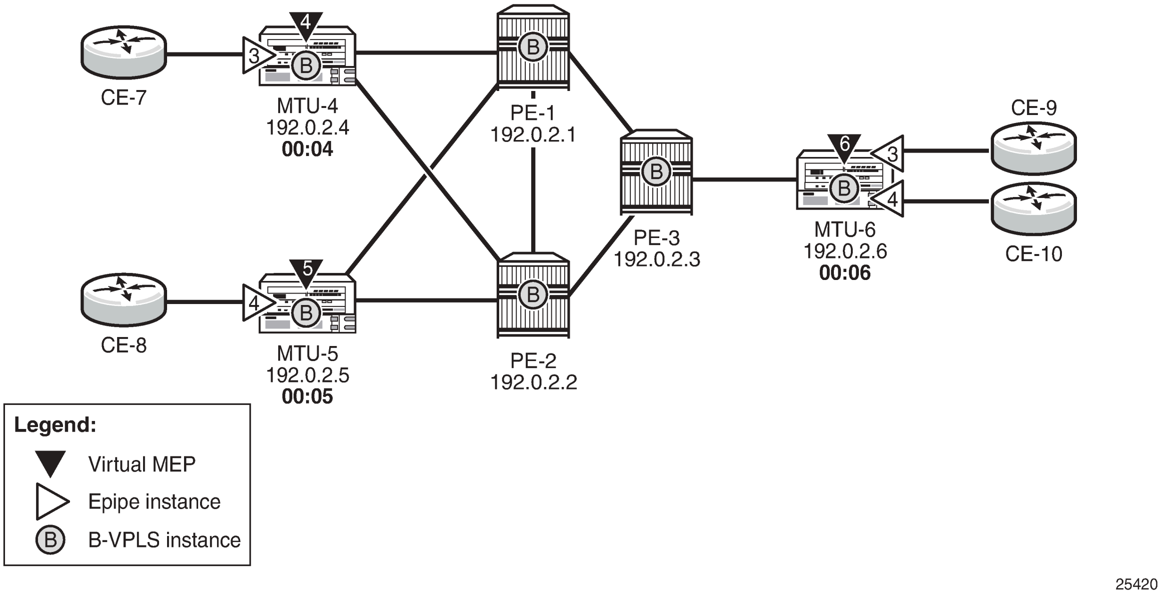
Flooding case 3 — Unidirectional traffic: virtual MEP and CCM configuration

For flooding case 3 (unidirectional traffic), Nokia recommends the use of ETH-CFM (802.1ag/Y.1731 Connectivity Fault Management) virtual Maintenance End Points (MEPs). By defining a virtual MEP per node terminating a PBB-Epipe, configuring the MEP MAC address to be the source-bmac value and activating continuity check messages (CCM), a twofold effect is achieved:

- The **pbb-tunnel backbone-destination-mac** will always be learned at the local FDB, as long as the remote virtual MEP is active and sending CC messages. As a result, there will not be flooding even if we have unidirectional traffic.
- An automatic proactive OAM mechanism exists to detect failures on remote nodes, which ultimately cause unnecessary flooding in the B-VPLS domain.

Figure 240: Virtual MEPs for flooding avoidance shows an example where the virtual MEPs MEP4, MEP5, and MEP6 are configured in B-VPLS 101:

Figure 240: Virtual MEPs for flooding avoidance



The following configuration example uses MTU-4. First, the general ETH-CFM configuration is made:

```
# on MTU-4:
configure
    eth-cfm
        domain 1 format none level 3 admin-name "domain 1"
            association 1 format icc-based name "B-VPLS-000101" admin-name "assoc-1"
                bridge-identifier 101
                exit
                remote-mepid 5
                remote-mepid 6
            exit
        exit
    exit
```

Then the actual virtual MEP configuration is made:

```
# on MTU-4:
configure
    service
        vpls 101
            eth-cfm
                mep 4 domain 1 association 1
                    ccm-enable
                    mac-address 00:04:04:04:04:04
                    no shutdown
                exit
            exit
        exit
```

The MAC address configured for the MEP4 matches the MAC address configured as the **source-bmac** on MTU-4, which is the **backbone-destination-mac** configured on the Epipe 3 PBB tunnel on MTU-6. The source-BMAC address on MTU-4 is 00:04:04:04:04:04, as follows:

```
# on MTU-4:
configure
    service
        pbb
            source-bmac 00:04:04:04:04:04
            mac-name "MTU-4" 00:04:04:04:04:04
            mac-name "MTU-5" 00:05:05:05:05:05
            mac-name "MTU-6" 00:06:06:06:06:06
        exit
```

The backbone destination MAC address configured on MTU-6 uses MAC name “MTU-4”, which corresponds to MAC address 00:04:04:04:04:04, as follows:

```
# on MTU-6:
configure
    service
        pbb
            source-bmac 00:06:06:06:06:06
            mac-name "MTU-4" 00:04:04:04:04:04
            mac-name "MTU-5" 00:05:05:05:05:05
            mac-name "MTU-6" 00:06:06:06:06:06
        exit
        epipe 3 name " Epipe 3" customer 1 create
            description "pbb epipe number 3"
            pbb
                tunnel 101 backbone-dest-mac "MTU-4" isid 3
            exit
            sap 1/1/1:9 create
            exit
            no shutdown
        exit
```

Once MEP4 has been configured, check that MTU-6 is receiving CC messages from MEP4 with the following command:

```
*A:MTU-6# show eth-cfm mep 6 domain 1 association 1 all-remote-mepids

=============================================================================
Eth-CFM Remote-Mep Table
=============================================================================
R-mepId AD Rx CC RxRdi Port-Tlv If-Tlv Peer Mac Addr     CCM status since
-----------------------------------------------------------------------------
4          True  False Absent   Absent 00:04:04:04:04:04 01/05/2021 16:05:12
5          True  False Absent   Absent 00:05:05:05:05:05 01/05/2021 16:05:12
=============================================================================
Entries marked with a 'T' under the 'AD' column have been auto-discovered.
```

As a result of the CC messages coming from MEP4, the MTU-4 MAC is permanently learned in the B-VPLS 101 FDB on node MTU-6 and no flooding takes place. The following output shows that the flooding flag is not set.

```
*A:MTU-6# show service id 3 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 3                   Vpn Id            : 0
Service Type      : Epipe
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                              Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/1:9                             q-tag        1518    1518    Up   Up

-------------------------------------------------------------------------------
PBB Tunnel Point
-------------------------------------------------------------------------------
B-vpls     Backbone-dest-MAC Isid     AdmMTU OperState Flood Oper-dest-MAC
-------------------------------------------------------------------------------
101        MTU-4             3        2000   Up        No    00:04:04:04:04:04
-------------------------------------------------------------------------------
Last Status Change: 01/05/2021 16:03:16
Last Mgmt Change  : 01/05/2021 16:03:16
===============================================================================
```

Flooding case 4 — Remote node failure

If the node owner of the **backbone-dest-mac** fails or gets isolated, the node where the PBB Epipe is initiated will not detect the failure; that is, if MTU-4 fails, the Epipe 3 remote end will also fail but MTU-6 will not detect the failure and as a result of that, MTU-6 will flood the traffic to the network (flooding will occur after MTU-4 MAC is removed from the B-VPLS FDBs, due to either the B-VPLS flushing mechanisms or aging).

In order to avoid/reduce flooding in this case, the following mechanisms are recommended:

- Provision virtual MEPs in the B-VPLS instances terminating PBB Epipes, as already explained. This will guarantee there is no unknown B-MAC unicast being flooded under normal operation.
- CCM timers should be provisioned based on how long the service provider is willing to accept flooding.

```
*A:MTU-6# configure eth-cfm domain 1 association 1 ccm-interval
  - ccm-interval <interval>
  - no ccm-interval

 <interval>           : {10ms|100ms|1|10|60|600} - default 10 seconds
```

- It is possible to provision **discard-unknown** in the B-VPLS, so that flooded traffic due to the destination MAC being unknown in the B-VPLS is discarded immediately. This can be configured on the PEs and the MTUs. On the MTUs, it is important to configure this in conjunction with the CC messages from the virtual MEPs to ensure that the remote B-MACs are learned in both directions. If, for any reason, the remote B-MACs are not in the MTU B-VPLS, no traffic will be forwarded at all on the PBB-Epipe.

```
# on all nodes:
configure
    service
        vpls 101
            discard-unknown
```

As soon as the MTU node recovers, it will start sending CC messages and the backbone MAC address will be learned on the backbone nodes and MTU nodes again.

With the recommended configuration in place, in case MTU-4 fails, the **backbone-dest-mac** configured on the PBB tunnel for Epipe 3 on MTU-6 will be removed from the B-VPLS 101 on all the nodes (either by MAC flush mechanisms on the B-VPLS or by aging). From that point on, traffic originated from CE-9 will be discarded at MTU-6 and won’t be flooded further.

As soon as MTU-4 comes back up, MEP4 will start sending CCM and as such the MTU-4 MAC will be learned throughout the B-VPLS 101 domain and in particular in PE-1, PE-3, and MTU-6 (CCM PDUs use a multicast address). From the moment MTU-4 MAC is known on the backbone nodes and MTU-6, the traffic will not be discarded any more, but forwarded to MTU-4.

PBB-Epipe show commands

The following commands can help to check the PBB Epipe configuration and their related parameters.

For the B-VPLS service:

```
*A:MTU-4# show service id 101 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 101                 Vpn Id            : 0
Service Type      : b-VPLS
MACSec enabled    : no
Name              : B-VPLS 101
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 01/05/2021 16:00:57
Last Mgmt Change  : 01/05/2021 16:08:54
Etree Mode        : Disabled
Admin State       : Up                  Oper State        : Up
MTU               : 2000
SAP Count         : 0                   SDP Bind Count    : 2
Snd Flush on Fail : Disabled            Host Conn Verify  : Disabled
SHCV pol IPv4     : None
Propagate MacFlush: Disabled            Per Svc Hashing   : Disabled
Allow IP Intf Bind: Disabled
Fwd-IPv4-Mcast-To*: Disabled            Fwd-IPv6-Mcast-To*: Disabled
Mcast IPv6 scope  : mac-based
Temp Flood Time   : Disabled            Temp Flood        : Inactive
Temp Flood Chg Cnt: 0
SPI load-balance  : Disabled
TEID load-balance : Disabled
Src Tep IP        : N/A
Vxlan ECMP        : Disabled
MPLS ECMP         : Disabled
VSD Domain        : <none>
Oper Backbone Src : 00:04:04:04:04:04
Use SAP B-MAC     : Disabled
i-Vpls Count      : 0
Epipe Count       : 1
Use ESI B-MAC     : Disabled

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type        AdmMTU  OprMTU  Adm  Opr
```

```
-------------------------------------------------------------------------------
sdp:41:101 S(192.0.2.1)                  Spok         8000    8000    Up   Up
sdp:42:101 S(192.0.2.2)                  Spok         8000    8000    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

For the Epipe service:

```
*A:MTU-4# show service id 3 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 3                   Vpn Id            : 0
Service Type      : Epipe
MACSec enabled    : no
Name              : Epipe 3
Description       : pbb epipe number 3
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 01/05/2021 16:03:03
Last Mgmt Change  : 01/05/2021 16:03:03
Test Service      : No
Admin State       : Up                  Oper State        : Up
MTU               : 1514
Vc Switching      : False
SAP Count         : 1                   SDP Bind Count    : 0
Per Svc Hashing   : Disabled
Vxlan Src Tep Ip  : N/A
Force QTag Fwd    : Disabled
Oper Group        : <none>

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/1:7                              q-tag        1518    1518    Up   Up

-------------------------------------------------------------------------------
PBB Tunnel Point
-------------------------------------------------------------------------------
B-vpls     Backbone-dest-MAC Isid     AdmMTU OperState Flood Oper-dest-MAC
-------------------------------------------------------------------------------
101        MTU-6             3        2000   Up        No    00:06:06:06:06:06
-------------------------------------------------------------------------------
Last Status Change: 01/05/2021 16:03:03
Last Mgmt Change  : 01/05/2021 16:03:03
===============================================================================
```

The following command shows all the Epipe instances multiplexed into a particular B-VPLS and its status.

```
*A:MTU-4# show service id 101 epipe

===============================================================================
Related Epipe services for b-Vpls service 101
===============================================================================
Epipe SvcId         Oper ISID           Admin               Oper
-------------------------------------------------------------------------------
3                   3                   Up                  Up
-------------------------------------------------------------------------------
Number of Entries : 1
-------------------------------------------------------------------------------
```

```
===============================================================================
```

To check the virtual MEP information, the following command shows the local virtual MEPs configured on the node:

```
*A:MTU-4# show eth-cfm cfm-stack-table all-virtuals
===============================================================================
CFM Stack Table Defect Legend:
R = Rdi, M = MacStatus, C = RemoteCCM, E = ErrorCCM, X = XconCCM
A = AisRx, L = CSF LOS Rx, F = CSF AIS/FDI rx, r = CSF RDI rx
G = receiving grace PDU (MCC-ED or VSM) from at least one peer

===============================================================================
CFM Virtual Stack Table
===============================================================================
Service            Lvl Dir Md-index   Ma-index  MepId  Mac-address     Defect G
-------------------------------------------------------------------------------
101                  3 U          1          1    4 00:04:04:04:04:04 ------- -
===============================================================================
```

The following command shows all the information related to the remote MEPs configured in the association, for example, the remote virtual MEPs configured in MTU-5 and MTU-6:

```
*A:MTU-4# show eth-cfm mep 4 domain 1 association 1 all-remote-mepids

=============================================================================
Eth-CFM Remote-Mep Table
=============================================================================
R-mepId AD Rx CC RxRdi Port-Tlv If-Tlv Peer Mac Addr     CCM status since
-----------------------------------------------------------------------------
5          True  False Absent   Absent 00:05:05:05:05:05 01/05/2021 16:04:56
6          True  False Absent   Absent 00:06:06:06:06:06 01/05/2021 16:04:56
=============================================================================
Entries marked with a 'T' under the 'AD' column have been auto-discovered.
```

The following command shows the detail information and status of the local virtual MEP configured in MTU-4:

```
*A:MTU-4# show eth-cfm mep 4 domain 1 association 1
===============================================================================
Eth-Cfm MEP Configuration Information
===============================================================================
Md-index         : 1                         Direction         : Up
Ma-index         : 1                         Admin             : Enabled
MepId            : 4                         CCM-Enable        : Enabled
SvcId            : 101
Description      : (Not Specified)
FngAlarmTime     : 0                         FngResetTime      : 0
FngState         : fngReset                  ControlMep        : False
LowestDefectPri  : macRemErrXcon             HighestDefect     : none
Defect Flags     : None
Mac Address      : 00:04:04:04:04:04         Collect LMM Stats : disabled
LMM FC Stats     : None
LMM FC In Prof   : None
TxAis            : noTransmit                TxGrace           : noTransmit
Facility Fault   : disabled
CcmLtmPriority   : 7                         CcmPaddingSize    : 0 octets
CcmTx            : 47                        CcmSequenceErr    : 0
CcmTxIfStatus    : Absent                    CcmTxPortStatus   : Absent
CcmTxRdi         : False                     CcmTxCcmStatus    : transmit
CcmIgnoreTLVs    : (Not Specified)
```

```
Fault Propagation: disabled                FacilityFault     : n/a
MA-CcmInterval   : 10                      MA-CcmHoldTime    : 0ms
MA-Primary-Vid   : Disabled
Eth-1Dm Threshold: 3(sec)                  MD-Level          : 3
Eth-1Dm Last Dest: 00:00:00:00:00:00
Eth-Dmm Last Dest: 00:00:00:00:00:00
Eth-Ais          : Disabled
Eth-Ais Tx defCCM: allDef
Eth-Tst          : Disabled
Eth-CSF          : Disabled

Eth-Cfm Grace Tx : Enabled                 Eth-Cfm Grace Rx  : Enabled
Eth-Cfm ED Tx    : Disabled                Eth-Cfm ED Rx     : Enabled
Eth-Cfm ED Rx Max: 0
Eth-Cfm ED Tx Pri: CcmLtmPri (7)

Eth-BNM Receive  : Disabled                Eth-BNM Rx Pacing : 5

Redundancy:
    MC-LAG State : n/a

CcmLastFailure Frame:
    None

XconCcmFailure Frame:
    None
===============================================================================
```

When there is a failure on a remote Epipe node, as described, the source node keeps sending traffic. The 802.1ag/Y.1731 virtual MEP configured can help to detect and troubleshoot the problem. For instance, when a failure happens in MTU-6 (node goes down or the B-VPLS instance is disabled), the virtual MEP show commands will show the following information:

```
# on MTU-6:
configure
    service
        vpls 101
            shutdown
```

```
*A:MTU-4# show eth-cfm mep 4 domain 1 association 1
===============================================================================
Eth-Cfm MEP Configuration Information
===============================================================================
Md-index         : 1                       Direction         : Up
Ma-index         : 1                       Admin             : Enabled
MepId            : 4                       CCM-Enable        : Enabled
SvcId            : 101
Description      : (Not Specified)
FngAlarmTime     : 0                       FngResetTime      : 0
FngState         : fngDefectReported       ControlMep        : False
LowestDefectPri  : macRemErrXcon           HighestDefect     : defRemoteCCM
Defect Flags     : bDefRDICCM bDefRemoteCCM
Mac Address      : 00:04:04:04:04:04       Collect LMM Stats : disabled
LMM FC Stats     : None
LMM FC In Prof   : None
TxAis            : noTransmit              TxGrace           : noTransmit
Facility Fault   : disabled
CcmLtmPriority   : 7                       CcmPaddingSize    : 0 octets
CcmTx            : 70                      CcmSequenceErr    : 0
CcmTxIfStatus    : Absent                  CcmTxPortStatus   : Absent
CcmTxRdi         : True                    CcmTxCcmStatus    : transmit
CcmIgnoreTLVs    : (Not Specified)
```

```
Fault Propagation: disabled                 FacilityFault     : n/a
MA-CcmInterval   : 10                       MA-CcmHoldTime    : 0ms
MA-Primary-Vid   : Disabled
Eth-1Dm Threshold: 3(sec)                   MD-Level          : 3
Eth-1Dm Last Dest: 00:00:00:00:00:00
Eth-Dmm Last Dest: 00:00:00:00:00:00
Eth-Ais          : Disabled
Eth-Ais Tx defCCM: allDef
Eth-Tst          : Disabled
Eth-CSF          : Disabled

Eth-Cfm Grace Tx : Enabled                  Eth-Cfm Grace Rx  : Enabled
Eth-Cfm ED Tx    : Disabled                 Eth-Cfm ED Rx     : Enabled
Eth-Cfm ED Rx Max: 0
Eth-Cfm ED Tx Pri: CcmLtmPri (7)

Eth-BNM Receive  : Disabled                 Eth-BNM Rx Pacing : 5

Redundancy:
    MC-LAG State : n/a

CcmLastFailure Frame:
    None

XconCcmFailure Frame:
    None
===============================================================================
```

The bDefRemoteCCMdefect flag clearly shows that there is a remote MEP in the association which has stopped sending CCMs. In order to find out which node is affected, see the following output:

```
*A:MTU-4# show eth-cfm mep 4 domain 1 association 1 all-remote-mepids

=============================================================================
Eth-CFM Remote-Mep Table
=============================================================================
R-mepId AD Rx CC RxRdi Port-Tlv If-Tlv Peer Mac Addr     CCM status since
-----------------------------------------------------------------------------
5          True  True  Absent   Absent 00:05:05:05:05:05 01/05/2021 16:04:56
6          False False Absent   Absent 00:00:00:00:00:00 01/05/2021 16:14:28

5          True  True  Absent   Absent 00:05:05:05:05:05 06/14/2019 09:13:39
6          False False Absent   Absent 00:00:00:00:00:00 06/14/2019 09:17:58
=============================================================================
Entries marked with a 'T' under the 'AD' column have been auto-discovered.
```

CCMs are no longer received from virtual MEP 6 (the one defined in MTU-6) since 01/05/2021 16:14:28. This conveys which node has failed and when it failed.

Conclusion

Point-to-Point Ethernet services can use the same operational model followed by PBB VPLS for multipoint services. In other words, Epipes can be linked to the same B-VPLS domain being used by I-VPLS instances and use the existing H-VPLS network infrastructure in the core. The use of PBB Epipes reduces dramatically the number of services and pseudowires in the core and therefore allows the service provider to scale the number of E-Line services in the network.

The example used in this chapter shows the configuration of the PBB Epipes as well as all the related features which are required for this environment. Show commands have also been suggested so that the operator can verify and troubleshoot the service.

PBB-EVPN ISID-based CMAC Flush

This chapter provides information about PBB-EVPN ISID-based CMAC Flush.

Topics in this chapter include:

Applicability

This chapter was initially written for SR OS Release 15.0.R4, but the CLI in the current edition is based on SR OS Release 21.2.R2. PBB-EVPN ISID-based CMAC flush is supported on the following objects in an I-VPLS:

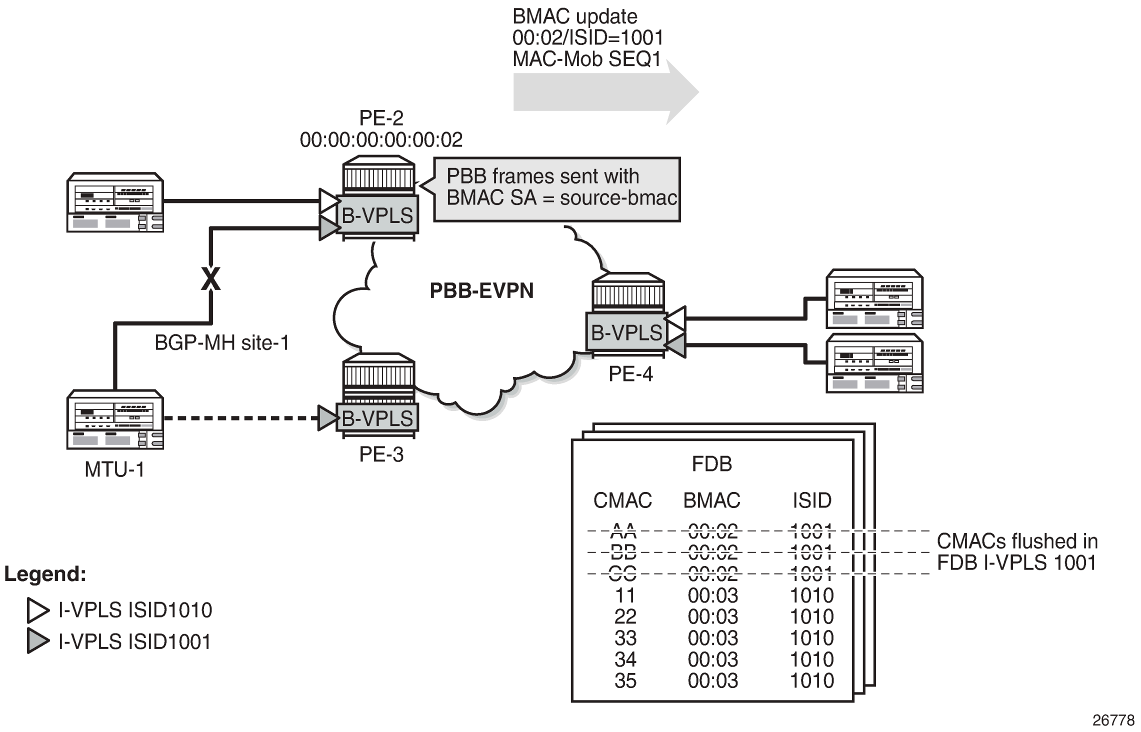
- SAPs in a BGP multi-homing site (no Ethernet Segment (ES))-supported in SR OS Release 14.0.R4, and later
- SAPs in ESs or virtual ESs (vESs)-SR OS Release 15.0.R1, and later
- Spoke-SDPs (that may be part of an ES/vES or not)-SR OS Release 15.0.R4, and later.

Chapter EVPN for PBB over MPLS (PBB-EVPN) is prerequisite reading.

Overview

Figure 241: CMAC flush when SAP in BGP multi-homing site fails shows an example topology with PBB-EVPN where a CMAC flush is triggered after a SAP in a BGP multi-homing site fails.

Figure 241: CMAC flush when SAP in BGP multi-homing site fails



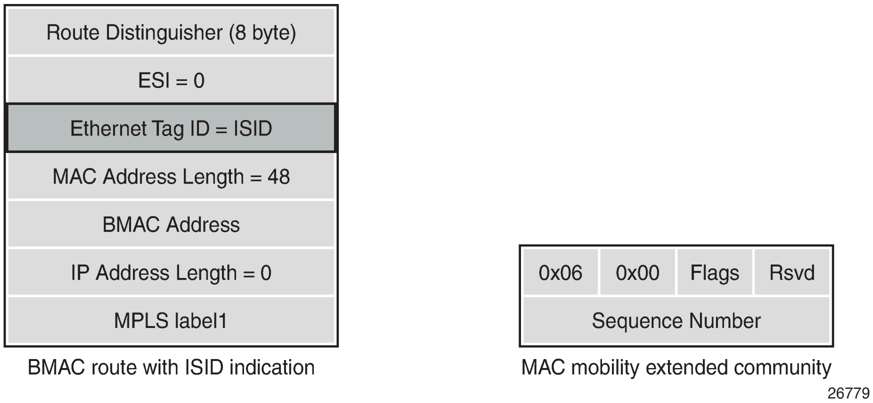
FDB

CMAC	BMAC	ISID
~~AA~~	~~00:02~~	~~1001~~
~~BB~~	~~00:02~~	~~1001~~
~~CC~~	~~00:02~~	~~1001~~
11	00:03	1010
22	00:03	1010
33	00:03	1010
34	00:03	1010
35	00:03	1010

I-VPLS 1001 is configured in PE-2 and PE-3 with **send-bvpls-evpn-flush** and connected to MTU-1. In the example, the SAP goes operationally down in I-VPLS 1001 on PE-2. To speed up convergence without flushing CMAC addresses in other I-VPLS services, PE-2 sends a BGP-EVPN BMAC route for ISID 1001 with increased sequence number to trigger a MAC-flush for I-VPLS 1001 on the remote PEs. All CMAC addresses in the FDB for other I-VPLS services, such as I-VPLS 1010 in this example, will be preserved. When PE-4 needs to send traffic to one of the flushed CMAC addresses in I-VPLS 1001, it will flood the frames until the CMAC address is learned again (via PE-3).

When SAPs or SDP-bindings-associated with ESs, vESs, or BGP-MH sites-in an I-VPLS service fail, a BGP-EVPN BMAC route (route type 2) can trigger an ISID-based CMAC flush on the remote PEs. For the CMAC addresses to be flushed from the FDB of the I-VPLS, the existing EVPN BMAC routes will be used with the Ethernet tag equal to the ISID. Figure 242: EVPN BMAC route with ISID indication shows the EVPN BMAC route with ISID indication (BMAC/ISID). A BMAC/ISID update may trigger a selective MAC-flush for a specific I-VPLS, whereas a BMAC/0 update (BMAC/ISID route where ISID=0) may trigger a MAC-flush for all I-VPLS services. This procedure is based on *draft-snr-bess-pbb-evpn-isid-cmacflush*.

Figure 242: EVPN BMAC route with ISID indication



By default, ISID-based CMAC flush is disabled: no I-VPLS will send a B-VPLS EVPN flush message and no B-VPLS will accept any I-VPLS EVPN flush messages. The router only installs CMAC entries corresponding to a zero Ethernet tag and ignores non-zero Ethernet tag MAC routes. However, when the B-VPLS is configured to accept BMAC/ISID routes, non-zero Ethernet tag BMAC routes can be processed for CMAC flush. The CMAC flush trigger will be an EVPN BMAC/ISID route with a sequence number that is higher than before. The receiving PE will then flush all CMACs associated with this BMAC address in the I-VPLS.

The first time that a BMAC/ISID route is received, it is added to the database as a baseline. It does not cause a CMAC flush. Only subsequent BMAC/ISID updates with increased sequence number or withdrawals will cause CMAC flush.

The following command shows that B-VPLS 1000 does not accept any I-VPLS EVPN flush messages. This is the default behavior.

```
*A:PE-2# show service id 1000 bgp-evpn | match "Accept IVPLS Flush"
Accept IVPLS Flush : Disabled
```

At the receiving node, B-VPLS 1000 will accept BMAC/ISID routes when the following command is configured:

```
# on PE-2:
configure
    service
        vpls "B-VPLS 1000"
            bgp-evpn
                accept-ivpls-evpn-flush
```

By default, I-VPLS 1001 will not send any B-VPLS EVPN flush messages, as follows:

```
*A:PE-2# show service id 1001 base | match SendBvplsEvpnFlush
SendBvplsEvpnFlush : Disabled
```

The following configuration allows I-VPLS 1001 to send B-VPLS EVPN flush messages when a SAP or SDP-binding fails:

```
# on PE-2:
configure
    service
```

```
        vpls "I-VPLS 1001"
            pbb
                send-bvpls-evpn-flush
```

When enabled, the I-VPLS will send a BMAC/ISID route and subsequent updates with a higher sequence number whenever a SAP fails in the I-VPLS on the node. The default setting for a SAP allows a B-VPLS EVPN flush message to be sent (when enabled in the I-VPLS itself):

```
*A:PE-2# show service id 1001 sap 1/2/1:1001 detail | match SendBvplsEvpnFlush
SendBvplsEvpnFlush : Enabled
```

When no alternative route via another node is available for specific SAPs (single-homed SAPs), no CMAC flush should be triggered. When no B-VPLS EVPN flush messages need to be sent from PE-4 when SAP 1/2/1:1001 goes down, the configuration is as follows:

```
# on PE-4:
configure
    service
        vpls "I-VPLS 1001"
            sap 1/2/1:1001
                disable-send-bvpls-evpn-flush
```

The router only installs the BMACs received in MAC routes that have Ethernet tag zero. When CMAC flush is enabled, MAC routes with Ethernet tag equal to the ISID (always non-zero) are for CMAC flush, but not for installing the conveyed BMACs.

BMAC/ISID routes have the following characteristics:

- BMAC/ISID routes are sent with the static bit flag set as for any other BMAC route. The static bit is ignored at reception because this route is never used to install a BMAC in the FDB.
- BMAC/ISID routes received with non-zero ESI and non-zero Ethernet tag are treated as withdraw by the router at application level. Route Reflectors (RRs) treat such BMAC/ISID routes as valid routes that can be forwarded.
- BMAC/ISID routes are shown as valid in the **show router bgp routes evpn mac** commands, as in the following output, even though they are not used to populate the FDB. This shows that BGP is sending the routes to the application layer for CMAC flush processing. The BMAC/0 route should be sent before the BMAC/ISID routes for the same BMAC. Also, when the B-VPLS goes operationally down, the BMAC/0 should be withdrawn before the BMAC/ISID routes.

```
*A:PE-2# show router bgp routes evpn mac rd 192.0.2.3:1000
===============================================================================
 BGP Router ID:192.0.2.2        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.3:1000      00:00:00:00:00:03 ESI-0
```

```
      0                   Static            LABEL 524282
                          n/a
                          192.0.2.3

u*>i  192.0.2.3:1000      00:00:00:00:00:03 ESI-0
      1001                Static            LABEL 524282
                          n/a
                          192.0.2.3

-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

When **send-bvpls-evpn-flush** is enabled in an I-VPLS that is associated with a B-VPLS, BGP-EVPN BMAC/ISID updates will be sent when certain events take place in the I-VPLS or B-VPLS. Table 11: CMAC flush transmission behavior shows the CMAC flush transmission behavior at the egress PE.

Table 11: CMAC flush transmission behavior

Local Event	Send-bvpls-evpn-flush	SAP disable-bvpls-evpn-flush	Action
Reconfigure I-VPLS: enable or disable send-bvpls-evpn-flush	Enable or disable	N/A	Send update/withdraw source BMAC/ISID with Seq=0
Associate/disassociate I-VPLS to/from B-VPLS	Enabled	N/A	Send update/withdraw source BMAC/ISID with Seq=0
I-VPLS oper-up/oper-down	Enabled	N/A	Send update/withdraw source BMAC/ISID with Seq=0
B-VPLS oper-up/oper-down	Enabled	N/A	Send update/withdraw source BMAC/ISID with Seq=0 Note: All BMACs are also advertised/withdrawn.
B-VPLS bgp-evpn mpls no shut/shut	Enabled	N/A	Send update/withdraw source BMAC/ISID with Seq=0
B-VPLS operational source BMAC change	Enabled	N/A	Send update/withdraw source BMAC/ISID with Seq=0
SAP oper-up	Enabled	N/A	No operation
SAP oper-down	Enabled	No disable	Send update source BMAC/ISID Seq=Seq+1
	Enabled	Disable	No operation

Table 12: CMAC flush reception behavior shows the reception behavior at the ingress PE. For the CMAC flush triggered by a BMAC/ISID update with increased sequence number, the B-VPLS in the receiving PE must be configured with **accept-ivpls-evpn-flush**. BMAC/0 refers to a BMAC route where the Ethernet Tag is 0.

Table 12: CMAC flush reception behavior

Received Route	Action
BMAC/0 withdraw	Flush all CMACs for that BMAC
BMAC/ISID withdraw	Flush all CMACs for that BMAC and ISID
BMAC/0 update + Seq change	Flush all CMACs for that BMAC
BMAC/ISID update + Seq change	Flush all CMACs for that BMAC and ISID
BMAC/0 update + PE (NHop) change	No CMAC-flush
BMAC/ISID update + PE (NHop) change	Flush all CMACs for that BMAC and ISID

BMAC/ISID updates will trigger CMAC flush procedures regardless of the Termination Endpoint (TEP) or Route Distinguisher (RD) with which the update is received. CMAC flush will be processed even if the BMAC-ISID comes from a TEP or RD different from the BMAC/0 route. Even when the sequence number is the same as in the previous BMAC/ISID update, CMAC flush will happen when the TEP is different. When the same BMAC/ISID is received from two PEs, both are accepted and any change in sequence number causes a MAC flush. However, when the same BMAC/ISID route is received from two PEs with the same RD, BGP will select only one, so the router only sees one.

CMAC flush for ES/vES

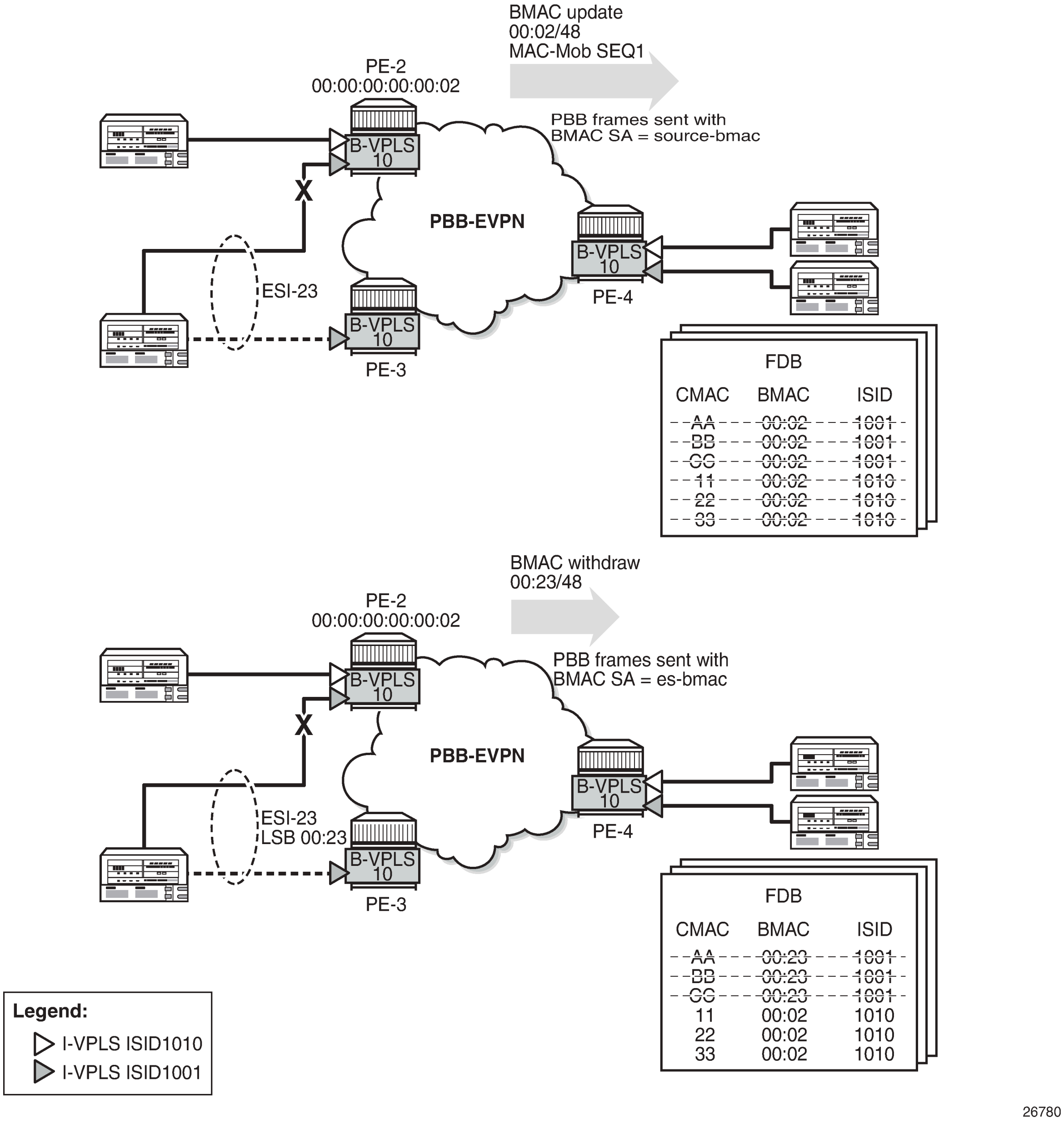
RFC 7623 (PBB-EVPN) defines the following CMAC Flush notification mechanisms for single-active multi-homing. These notifications do not include the local ISIDs:

- When ES-BMACs are used and the ES goes operationally down, the ES-BMAC will be withdrawn.
- When source-BMACs are used and the ES goes operationally down, a BGP-EVPN BMAC/0 is sent with a higher sequence number.

Figure 243: ISID-independent CMAC flush when ES fails shows the following two scenarios for ISID-independent CMAC flush that are supported in SR OS Release 13.0.R4, and later:

- PBB frames are sent with the source-BMAC. When the ES goes operationally down, a BMAC update is sent with an increased sequence number, triggering a CMAC flush for all CMAC addresses associated with the BMAC address in I-VPLS, regardless of the ISID.
- PBB frames are sent with the ES-BMAC address. When the ES goes operationally down, a BMAC withdraw message is sent, triggering the remote PEs to flush all CMAC addresses associated to the ES-BMAC address, regardless of the ISID.

Figure 243: ISID-independent CMAC flush when ES fails



In addition to the preceding ISID-independent CMAC flush mechanisms, ISID-based CMAC flush is also supported in I-VPLS services with SAP or spoke-SDPs that are part of an ES or vES. ISID-based CMAC flush is enabled in the I-VPLS with the **send-bvpls-evpn-flush** command. An I-VPLS that is configured with **send-bvpls-evpn-flush** requires one of the following conditions to be met:

- The SAP or spoke-SDP has **disable-send-bvpls-evpn-flush** configured.
- The SAP or spoke-SDP has **no disable-send-bvpls-evpn-flush** configured (default) and one of the following conditions is met:
  - The SAP or spoke-SDP is not on an ES.

  - The SAP or spoke-SDP is on an ES or vES with **no src-bmac-lsb** configured.
  - The B-VPLS has **no use-es-bmac** configured.

For ES SAPs with **no disable-send-bvpls-evpn-flush** in I-VPLS services that have **send-bvpls-evpn-flush** configured, the ISID-based CMAC flush replaces the RFC 7623-based CMAC flush mechanism.

For each ES/vES and B-VPLS, the system will check whether all I-VPLS services in the ES/B-VPLS have ISID-based MAC-flush enabled.

- If all I-VPLSs have **send-bvpls-evpn-flush** enabled:
  - No BMAC/0 updates with increased sequence number will be triggered when the ES/vES goes operationally down.
  - Only BMAC/ISID updates with increased sequence number will be sent when the I-VPLS attachment circuit goes operationally down.
- If at least one I-VPLS has **no send-bvpls-evpn-flush** enabled:
  - BMAC/0 updates with increased sequence number will be triggered when the ES/vES goes operationally down.
  - Also, BMAC/ISID updates with increased sequence number will be generated for those I-VPLS services that have **send-bvpls-evpn-flush** enabled.

The number of CMAC addresses that may be flushed at the remote nodes can be reduced by enabling ISID-based MAC-flush for all the I-VPLS services in the ES/vES.

When attempting to set **use-es-bmac** in B-VPLS 1000 on PE-4 when the SAP/SDP-binding has default settings (and **send-bvpls-evpn-flush** is enabled in the I-VPLS), the following error is raised:

```
*A:PE-4>config>service>vpls>pbb# use-es-bmac
MINOR: SVCMGR #1433 Cannot set use-es-bmac - spoke 46:1001 on ethernet-segment ESI-45 has "no
 disable-send-bvpls-evpn-flush"
```

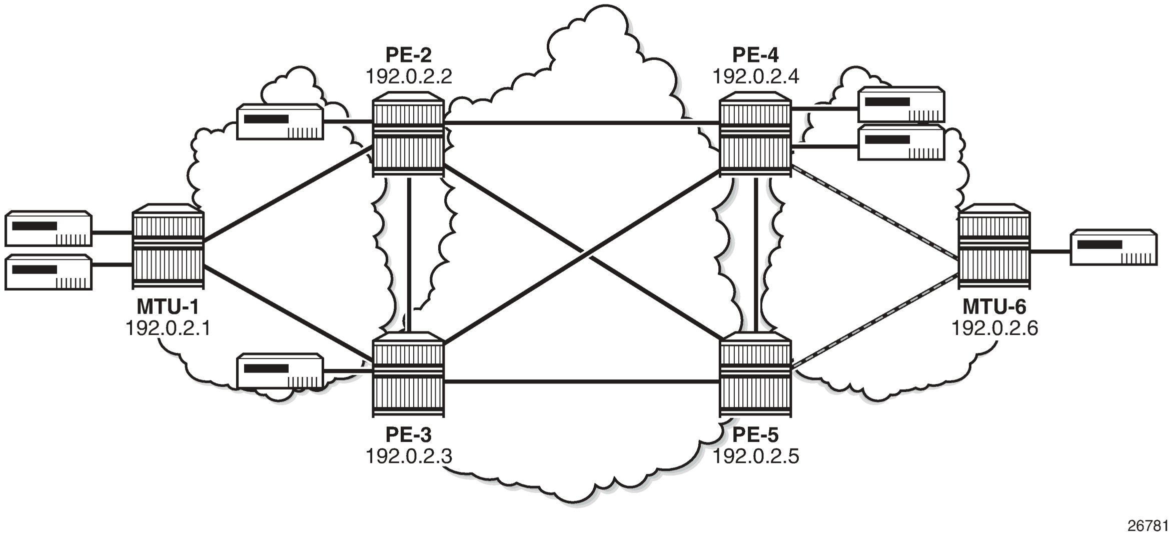
When the ES is disabled, the B-VPLS can be configured with **use-es-bmac**. When attempting to enable the ES afterward, the following error is raised.

```
*A:PE-4# configure service system bgp-evpn ethernet-segment "ESI-45" shutdown
*A:PE-4# configure service vpls "B-VPLS 1000" pbb use-es-bmac
*A:PE-4# configure service system bgp-evpn ethernet-segment "ESI-45" no shutdown
MINOR: SVCMGR #8057 Ethernet segment cannot change admin state -
spoke 46:1001 has "no disable-send-bvpls-evpn-flush"
```

Configuration

Figure 244: Example topology shows the example topology.

Figure 244: Example topology



The initial configuration includes the following:

- Cards, MDAs
- Ports: the ports between the MTUs and the PEs are hybrid or access ports with dot1q encapsulation; the ports between the PEs are network ports with null encapsulation
- Router interfaces
- IS-IS on all router interfaces (alternatively, OSPF could be used)
- LDP on all router interfaces

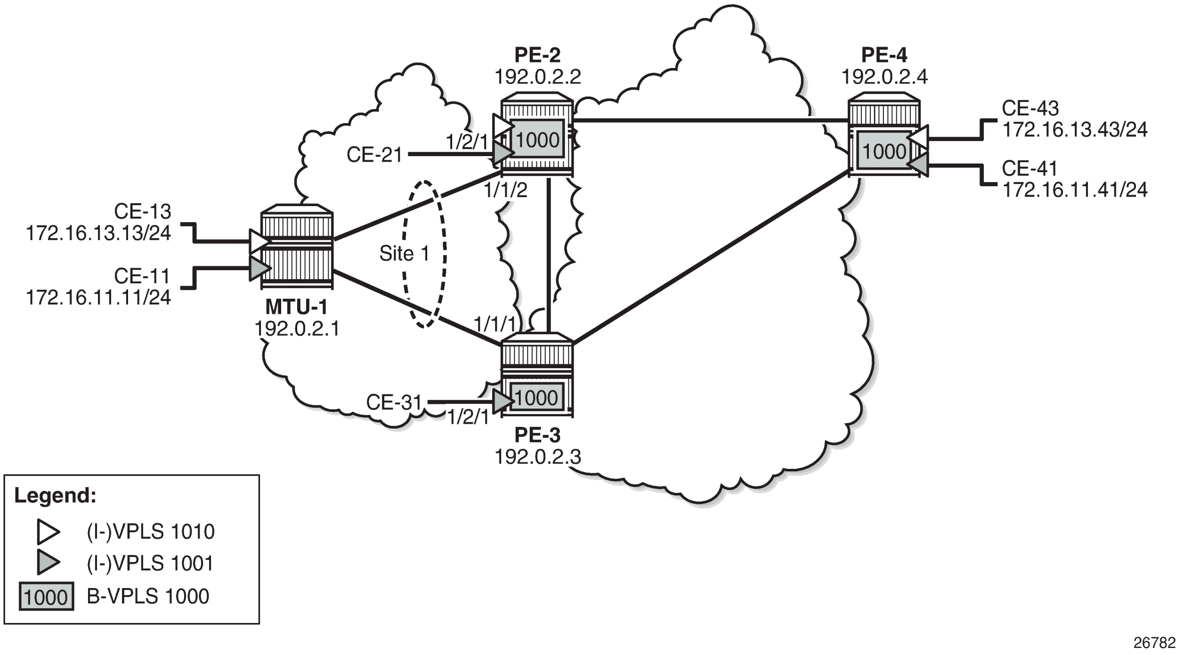
The following use cases are described in this section:

- ISID-based CMAC flush for BGP non-EVPN multi-homing (no ES)
- ISID-based CMAC flush for BGP-EVPN in a single-active ES

ISID-based CMAC flush for BGP multi-homing

Figure 245: Example topology with BGP multi-homing shows the example topology with BGP multi-homing site 1 between PE-2 and PE-3. B-VPLS 1000 is configured on all the core nodes (PEs) and I-VPLS 1001 and I-VPLS 1010 are associated with this B-VPLS in the PEs. On MTU-1, regular VPLSs are configured. For more information about BGP non-EVPN multi-homing, see chapter BGP Multi-Homing for VPLS Networks.

Figure 245: Example topology with BGP multi-homing



BGP is configured for address family EVPN on all PEs with PE-2 as RR. For BGP multi-homing, address family L2-VPN is enabled between PE-2 and PE-3. The BGP configuration on PE-2 is as follows:

```
# on PE-2:
configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update l2-vpn evpn
            group "internal"
                cluster 1.1.1.1
                peer-as 64500
                neighbor 192.0.2.3
                    family l2-vpn evpn
                exit
                neighbor 192.0.2.4
                    family evpn
                exit
            exit
        exit
```

The BGP configuration on PE-4 is as follows:

```
# on PE-4:
configure
    router Base
        autonomous-system 64500
        bgp
```

```
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family evpn
                peer-as 64500
                neighbor 192.0.2.2
                exit
            exit
        exit
```

The configuration of B-VPLS 1000 and I-VPLS 1001 on PE-2 is as follows. ISID-based CMAC flush is disabled by default. BGP multi-homing site "site 1" is configured on PE-2 with SAP 1/1/2:1001 associated with it, whereas SAP 1/2/1:1001 is not associated to the MH site. CE-21 is attached to I-VPLS 1001 with SAP 1/2/1:1001.

```
# on PE-2:
configure
    service
        system
            bgp-auto-rd-range 192.0.2.2 comm-val 1 to 999
        exit
        vpls 1000 name "B-VPLS 1000" customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:00:00:00:00:02
            exit
            bgp
            exit
            bgp-evpn
                evi 1000
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
        vpls 1001 name "I-VPLS 1001" customer 1 i-vpls create
            pbb
                backbone-vpls 1000
                exit
            exit
            bgp
                route-distinguisher auto-rd
                route-target export target:64500:1001 import target:64500:1001
            exit
            stp
                shutdown
            exit
            site "MH-site-1" create
                site-id 1
                1/1/2:1001
                no shutdown
            exit
            sap 1/1/2:1001 create
                no shutdown
            exit
            sap 1/2/1:1001 create
                no shutdown
            exit
            no shutdown
        exit
        vpls 1010 name "I-VPLS 1010" customer 1 i-vpls create
            pbb
                backbone-vpls 1000
                exit
            exit
            bgp
                route-distinguisher auto-rd
                route-target export target:64500:1010 import target:64500:1010
            exit
            stp
                shutdown
            exit
            sap 1/1/2:1010 create
                no shutdown
            exit
            no shutdown
        exit
```

I-VPLS 1010 is configured without multi-homing. The configuration of VPLS 1001 on PE-3 is similar, but without I-VPLS 1010.

ISID-based CMAC flush is not enabled yet. The PEs exchange BGP-EVPN MAC routes with Ethernet tag zero. PE-3 has received BMAC/0 routes from PE-2 and PE-4, as follows:

```
*A:PE-3# show router bgp routes evpn mac
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.2:1000      00:00:00:00:00:02 ESI-0
      0                   Static            LABEL 524282
                          n/a
                          192.0.2.2

u*>i  192.0.2.4:1000      00:00:00:00:00:04 ESI-0
      0                   Static            LABEL 524282
                          n/a
                          192.0.2.4

-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

PE-2 and PE-4 have also received BMAC/0 routes from the other PEs.

ISID-based CMAC flush is enabled in I-VPLS 1001 on PE-2 and PE-3. PE-4 has no multi-homing in I-VPLS 1001, so it should not send any CMAC flush. I-VPLS 1010 has no multi-homing in any PE, so ISID-based MAC-flush should not be enabled in I-VPLS 1010.

```
# on PE-2, PE-3:
configure
    service
        vpls "I-VPLS 1001"
            pbb
                send-bvpls-evpn-flush
```

PE-2 and PE-3 will send BMAC/1001 updates with sequence number 0 to the other two PEs. As an example, the following EVPN-MAC route for BMAC 00:00:00:00:00:03 with tag 1001 is sent by PE-3:

```
22 2021/04/15 08:07:57.818 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 89
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.3
        Type: EVPN-MAC Len: 33 RD: 192.0.2.3:1000 ESI: ESI-0, tag: 1001, mac len: 48
                       mac: 00:00:00:00:00:03, IP len: 0, IP: NULL, label1: 8388512
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:1000
        bgp-tunnel-encap:MPLS
        mac-mobility:Seq:0/Static
"
```

PE-4 has received the following BMAC routes from PE-2 and PE-3, with Ethernet tag zero and Ethernet tag 1001. BMAC routes are always static (received with the sticky bit set).

```
*A:PE-4# show router bgp routes evpn mac
===============================================================================
 BGP Router ID:192.0.2.4        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.2:1000      00:00:00:00:00:02 ESI-0
      0                   Static            LABEL 524282
                          n/a
                          192.0.2.2

u*>i  192.0.2.2:1000      00:00:00:00:00:02 ESI-0
```

```
      1001                Static            LABEL 524282
                          n/a
                          192.0.2.2

u*>i  192.0.2.3:1000      00:00:00:00:00:03 ESI-0
      0                   Static            LABEL 524282
                          n/a
                          192.0.2.3

u*>i  192.0.2.3:1000      00:00:00:00:00:03 ESI-0
      1001                Static            LABEL 524282
                          n/a
                          192.0.2.3

-------------------------------------------------------------------------------
Routes : 4
===============================================================================
```

When a failure occurs on PE-2, PE-3, and PE-4 should accept the BMAC/ISID with increased sequence number; for a failure on PE-3, PE-2, and PE-4 should accept the BMAC/ISID update. Therefore, the B-VPLS on all PEs should accept the CMAC flush message for ISID 1001, and this is configured as follows:

```
# on PE-2, PE-3, PE-4, PE-5:
configure
    service
        vpls "B-VPLS 1000"
            bgp-evpn
                accept-ivpls-evpn-flush
```

The FDB for VPLS 1001 on PE-4 includes MAC address 00:00:11:11:11:11 with source-identifier 192.0.2.2:524282, so PE-4 will forward traffic toward that MAC address to PE-2.

```
*A:PE-4# show service id 1001 fdb detail

===============================================================================
Forwarding Database, Service 1001
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1001       00:00:11:11:11:11 b-mpls:                 L/420    04/15/21 08:03:47
                             192.0.2.2:524282
           ldp:65537
1001       00:00:41:41:41:41 sap:1/2/1:1001          L/0      04/15/21 08:11:36
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

A failure is simulated on SAP 1/1/2:1001 in multi-homing site 1 on PE-2 as follows:

```
# on PE-2:
configure
    service
        vpls "I-VPLS 1001"
            sap 1/1/2:1001
                shutdown
```

SAP 1/1/2:1001 has the default **no disable-send-bvpls-evpn-flush** and I-VPLS 1001 is configured with **send-bvpls-evpn-flush**, so PE-2 will send BMAC/ISID updates for BMAC 00:00:00:00:00:02, ISID 1001, and sequence number 1 to its BGP peers. The following BGP update is sent by PE-2 to PE-4:

```
# on PE-2:
64 2021/04/15 08:12:55.058 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 89
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.2
        Type: EVPN-MAC Len: 33 RD: 192.0.2.2:1000 ESI: ESI-0, tag: 1001, mac len: 48
                       mac: 00:00:00:00:00:02, IP len: 0, IP: NULL, label1: 8388512
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:1000
        bgp-tunnel-encap:MPLS
        mac-mobility:Seq:1/Static
"
```

This BMAC/ISID with sequence number 1 triggers a CMAC flush in the FDB for VPLS 1001, so the entry for 00:00:11:11:11:11 will be flushed, along with all other MAC addresses associated with BMAC 00:00:00:00:00:02. The FDB on PE-4 does not contain any entries with source-identifier BMAC 00:00:00:00:00:02, as follows:

```
*A:PE-4# show service id 1001 fdb detail

===============================================================================
Forwarding Database, Service 1001
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1001       00:00:41:41:41:41 sap:1/2/1:1001          L/150    04/15/21 08:11:36
-------------------------------------------------------------------------------
No. of MAC Entries: 1
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

When the MAC address 00:00:11:11:11:11 is learned via PE-3, the FDB is as follows:

```
*A:PE-4# show service id 1001 fdb detail

===============================================================================
Forwarding Database, Service 1001
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1001       00:00:11:11:11:11 b-mpls:                 L/0      04/15/21 08:15:16
                             192.0.2.3:524282
           ldp:65538
1001       00:00:41:41:41:41 sap:1/2/1:1001          L/0      04/15/21 08:11:36
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
```

```
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

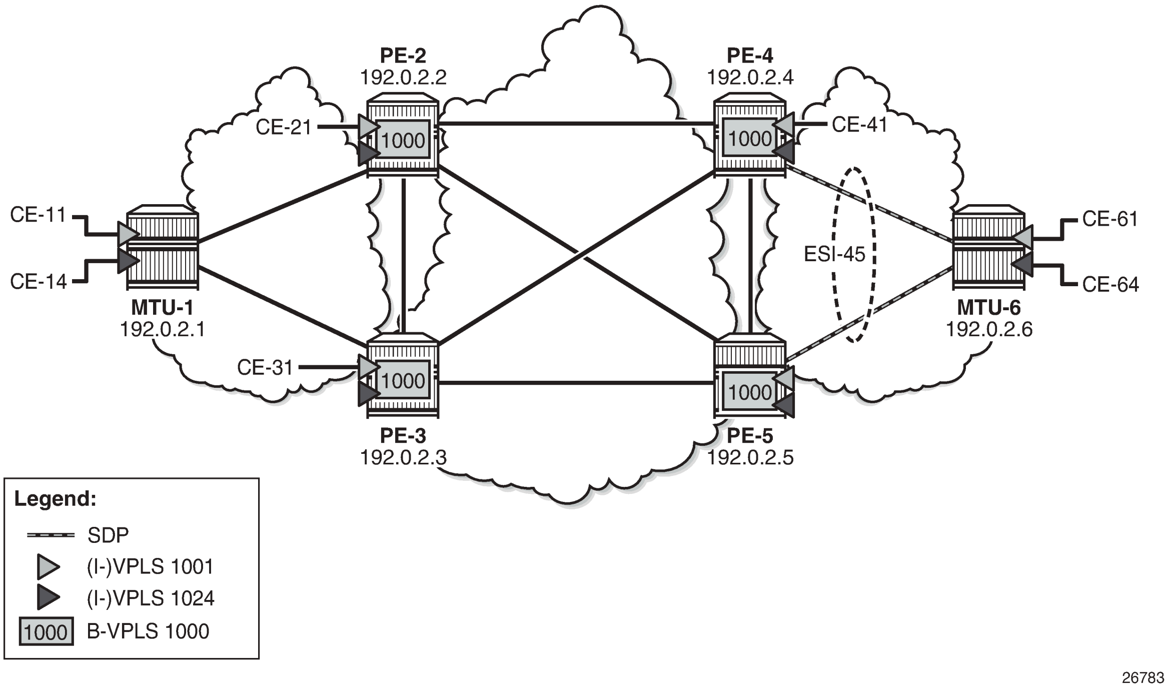
The CMAC flush is only applied for VPLS 1001, so the FDB for VPLS 1010 on PE-4 will keep entries learned from PE-2, as follows:

```
*A:PE-4# show service id 1010 fdb detail

===============================================================================
Forwarding Database, Service 1010
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1010       00:00:13:13:13:13 b-mpls:                 L/0      04/15/21 08:03:48
                             192.0.2.2:524282
           ldp:65537
1010       00:00:43:43:43:43 sap:1/2/1:1010          L/0      04/15/21 08:11:36
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

ISID-based CMAC flush in single-active ES

CMAC flush only makes sense for single-active multi-homing. Also, CMAC flush only works for single-active multi-homing; not for all-active multi-homing, because ES-BMAC is required in all-active multi-homing. Figure 246: Example topology with single-active ES shows the example topology with a single-active ES "ESI-45" configured in PE-4 and PE-5.

Figure 246: Example topology with single-active ES



The multi-homing configuration has been removed from PE-2 and PE-3, so no CMAC flush should be sent by PE-2 or PE-3. VPLS 1001 is configured as follows on PE-2 and PE-3:

```
# on PE-2, PE-3:
configure
    service
        vpls 1001 name "I-VPLS 1001" customer 1 i-vpls create
            pbb
                backbone-vpls 1000
                exit
            exit
            bgp
                route-distinguisher auto-rd
                route-target export target:64500:1001 import target:64500:1001
            exit
            stp
                shutdown
            exit
            sap 1/2/1:1001 create
                no shutdown
            exit
            sap lag-1:1001 create
                no shutdown
            exit
            no shutdown
        exit
```

SDPs are configured between PE-4 and MTU-6, and between PE-5 and MTU-6. These SDPs are associated with the single-active ES "ESI-45".

The configuration of B-VPLS 1000 on PE-4 is as follows. The B-VPLS configuration on the other PEs is similar, but with a different source BMAC.

```
# on PE-4:
configure
    service
        vpls 1000 name "B-VPLS 1000" customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:00:00:00:00:04
            exit
            bgp
            exit
            bgp-evpn
                accept-ivpls-evpn-flush
                evi 1000
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
```

The service configuration on PE-4 includes an SDP toward PE-6 and a single-active multi-homing ES, as follows:

```
# on PE-4:
configure
    service
        sdp 46 mpls create
            far-end 192.0.2.6
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        system
            bgp-evpn
                ethernet-segment "ESI-45" create
                    esi 01:00:00:00:00:45:00:00:00:01
                    source-bmac-lsb 45-04 es-bmac-table-size 8
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 46
                    no shutdown
                exit
            exit
        exit
```

The configuration on PE-5 is similar. The configuration of B-VPLS 1000 is similar to the one for PE-2, with only a different BMAC. The configuration of I-VPLS 1001 on PE-4 is as follows:

```
# on PE-4:
configure
    service
        vpls 1001 name "I-VPLS 1001" customer 1 i-vpls create
            pbb
                backbone-vpls 1000
                exit
                send-bvpls-evpn-flush
            exit
            bgp
                route-distinguisher auto-rd
                route-target export target:64500:1001 import target:64500:1001
            exit
            stp
                shutdown
            exit
            sap 1/2/1:1001 create
                no shutdown
            exit
            spoke-sdp 46:1001 create
                no shutdown
            exit
            no shutdown
        exit
```

ISID-based MAC-flush is enabled in B-VPLS 1000 and I-VPLS 1001 on all PEs.

I-VPLS 1024 is also associated with B-VPLS 1000 and contains one object (SAP or spoke-SDP) in each PE. The configuration of I-VPLS 1024 is identical on PE-2 and PE-3, as follows:

```
# on PE-2, PE-3:
configure
    service
        vpls 1024 name "I-VPLS 1024" customer 1 i-vpls create
            pbb
                backbone-vpls 1000
                exit
            exit
            stp
                shutdown
            exit
            sap lag-1:1024 create
                no shutdown
            exit
            no shutdown
        exit
```

The configuration of I-VPLS 1024 on PE-4 has **send-bvpls-evpn-flush** enabled and contains a spoke-SDP instead of a SAP, as follows. The configuration on PE-5 is similar, but with a different SDP.

```
# on PE-4:
configure
    service
        vpls 1024 name "I-VPLS 1024" customer 1 i-vpls create
            pbb
                backbone-vpls 1000
                exit
                send-bvpls-evpn-flush
            exit
            stp
                shutdown
            exit
            spoke-sdp 46:1024 create
                no shutdown
            exit
            no shutdown
        exit
```

ISID-based MAC-flush is enabled on PE-4 and PE-5 for both I-VPLS 1001 and I-VPLS 1024, and BMAC/ISID updates are sent for ISID 1001 and ISID 1024, as follows:

```
*A:PE-3# show router bgp routes evpn mac rd 192.0.2.4:1000
===============================================================================
 BGP Router ID:192.0.2.3        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.4:1000      00:00:00:00:00:04 ESI-0
      0                   Static            LABEL 524282
                          n/a
                          192.0.2.4

u*>i  192.0.2.4:1000      00:00:00:00:00:04 ESI-0
      1001                Static            LABEL 524282
                          n/a
                          192.0.2.4

u*>i  192.0.2.4:1000      00:00:00:00:00:04 ESI-0
      1024                Static            LABEL 524282
                          n/a
                          192.0.2.4

-------------------------------------------------------------------------------
Routes : 3
===============================================================================
```

PE-5 is the DF for VPLS 1001 in the single-active ES "ESI-45", but not for VPLS 1024, as follows:

```
*A:PE-5# show service id 1001 ethernet-segment
No sap entries

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
56:1001               ESI-45                           DF
===============================================================================
No vxlan instance entries
```

```
*A:PE-5# show service id 1024 ethernet-segment
No sap entries

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                  Eth-Seg                         Status
-------------------------------------------------------------------------------
56:1024              ESI-45                          NDF
===============================================================================
No vxlan instance entries
```

The following FDB for VPLS 1001 on PE-5 shows that traffic toward CMAC 00:00:11:11:11:11 (CE-11) in VPLS 1001 will be forwarded to PE-3:

```
*A:PE-5# show service id 1001 fdb detail

===============================================================================
Forwarding Database, Service 1001
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1001       00:00:11:11:11:11 b-mpls:                 L/0      04/15/21 08:19:47
                             192.0.2.3:524282
           ldp:65539
1001       00:00:41:41:41:41 b-mpls:                 L/0      04/15/21 08:19:47
                             192.0.2.4:524282
           ldp:65537
1001       00:00:61:61:61:61 sdp:56:1001             L/0      04/15/21 08:19:42
-------------------------------------------------------------------------------
No. of MAC Entries: 3
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The following FDB for VPLS 1024 on PE-4 shows that traffic toward CMAC 00:00:14:14:14:14 (CE-14) will be forwarded to PE-2:

```
*A:PE-4# show service id 1024 fdb detail

===============================================================================
Forwarding Database, Service 1024
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1024       00:00:14:14:14:14 b-mpls:                 L/0      04/15/21 08:19:48
                             192.0.2.2:524282
           ldp:65537
1024       00:00:64:64:64:64 sdp:46:1024             L/0      04/15/21 08:19:48
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The following FDB for VPLS 1001 on PE-3 shows that traffic toward CMAC 00:00:61:61:61:61 (CE-61) will be forwarded to PE-5:

```
*A:PE-3# show service id 1001 fdb detail

===============================================================================
Forwarding Database, Service 1001
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1001       00:00:11:11:11:11 sap:lag-1:1001          L/0      04/15/21 08:19:47
1001       00:00:41:41:41:41 b-mpls:                 L/0      04/15/21 08:19:47
                             192.0.2.4:524282
           ldp:65538
1001       00:00:61:61:61:61 b-mpls:                 L/0      04/15/21 08:19:42
                             192.0.2.5:524282
           ldp:65539
-------------------------------------------------------------------------------
No. of MAC Entries: 3
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The following FDB for VPLS 1024 on PE-2 shows that traffic toward CMAC 00:00:64:64:64:64 (CE-64) will be forwarded to PE-4:

```
*A:PE-2# show service id 1024 fdb detail

===============================================================================
Forwarding Database, Service 1024
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1024       00:00:14:14:14:14 sap:lag-1:1024          L/0      04/15/21 08:19:48
1024       00:00:64:64:64:64 b-mpls:                 L/0      04/15/21 08:19:48
                             192.0.2.4:524282
           ldp:65538
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

PE-5 is the DF for VPLS 1001 in "ESI-45". A failure is simulated by disabling the SDP toward PE-5 on MTU-6, as follows:

```
# on MTU-6:
configure
    service
        sdp 65
            shutdown
```

PE-5 sends the following BMAC/ISID with increased sequence number for ISID 1001 to the RR PE-2:

```
50 2021/04/15 08:24:35.567 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 89
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.5
        Type: EVPN-MAC Len: 33 RD: 192.0.2.5:1000 ESI: ESI-0, tag: 1001, mac len: 48
                       mac: 00:00:00:00:00:05, IP len: 0, IP: NULL, label1: 8388496
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:1000
        bgp-tunnel-encap:MPLS
        mac-mobility:Seq:1/Static
"
```

When PE-3 receives this BMAC/ISID, all MAC routes with next-hop PE-5 are flushed and the FDB will contain the following MAC entries:

```
*A:PE-3# show service id 1001 fdb detail

===============================================================================
Forwarding Database, Service 1001
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1001       00:00:11:11:11:11 sap:lag-1:1001          L/0      04/15/21 08:19:47
1001       00:00:41:41:41:41 b-mpls:                 L/0      04/15/21 08:19:47
                             192.0.2.4:524282
           ldp:65538
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

If MAC address 00:00:61:61:61:61 is learned again, the next hop will be PE-4 instead of PE-5.

The configuration is restored as follows:

```
# on MTU-6:
configure
    service
        sdp 65
            no shutdown
```

No CMAC/ISID update will be sent when the last SAP/SDP-binding in a service goes operationally down. VPLS 1024 only has one SAP/SDP-binding in DF PE-4: spoke-SDP 46:1024. A failure of the spoke-SDP is simulated as follows:

```
# on MTU-6:
configure
    service
        sdp 64
            shutdown
```

When the last SAP/SDP-binding is down, the service will be operationally down, as follows:

```
*A:PE-4# show service id 1024 base | match "Oper State"
Admin State       : Up                  Oper State        : Down
```

PE-4 sends the following withdrawal message instead of a CMAC/ISID:

```
56 2021/04/15 08:26:10.691 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 61
    Flag: 0x90 Type: 15 Len: 57 Multiprotocol Unreachable NLRI:
        Address Family EVPN
        Type: EVPN-INCL-MCAST Len: 17 RD: 192.0.2.4:1000, tag: 1024,
                              orig_addr len: 32, orig_addr: 192.0.2.4
        Type: EVPN-MAC Len: 33 RD: 192.0.2.4:1000 ESI: ESI-0, tag: 1024, mac len: 48
                       mac: 00:00:00:00:00:04, IP len: 0, IP: NULL, label1: 0
"
```

The configuration is restored as follows:

```
# on MTU-6:
configure
    service
        sdp 64
            no shutdown
```

ISID-based and regular CMAC flush in ES

When ISID-based CMAC flush is not enabled in all I-VPLS services using the ES, a failure in the ES will trigger BMAC/0 updates and BMAC/ISID updates with increased sequence number. An additional I-VPLS is configured on the nodes with **no send-bvpls-evpn-flush** (default). The configuration of I-VPLS 1021 on PE-5 is as follows:

```
# on PE-5:
configure
    service
        vpls 1021 name "I-VPLS 1021" customer 1 i-vpls create
            pbb
                backbone-vpls 1000
                exit
            exit
            stp
                shutdown
            exit
            sap 1/2/1:1021 create
                no shutdown
            exit
            spoke-sdp 56:1021 create
                no shutdown
            exit
            no shutdown
        exit
```

The configuration on PE-4 is similar; PE-2 and PE-3 have SAP lag-1:1021 instead of the spoke-SDP.

On MTU-6, SDP 65 is disabled, which will cause an ES failure on PE-5:

```
# on MTU-6:
configure
    service
        sdp 65
```

```
            shutdown
```

The following BMAC updates are sent by PE-5:

- BMAC/0 with increased sequence number, which will trigger a CMAC flush for all entries received from PE-5 for all I-VPLS services (ISID-independent)
- BMAC/ISID with increased sequence number, which will trigger a CMAC flush for all entries received from PE-5 for VPLS 1001

```
73 2021/04/15 08:32:57.204 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 89
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.5
        Type: EVPN-MAC Len: 33 RD: 192.0.2.5:1000 ESI: ESI-0, tag: 0, mac len: 48
                       mac: 00:00:00:00:00:05, IP len: 0, IP: NULL, label1: 8388496
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:1000
        bgp-tunnel-encap:MPLS
        mac-mobility:Seq:1/Static
"

74 2021/04/15 08:32:57.204 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 89
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.5
        Type: EVPN-MAC Len: 33 RD: 192.0.2.5:1000 ESI: ESI-0, tag: 1001, mac len: 48
                       mac: 00:00:00:00:00:05, IP len: 0, IP: NULL, label1: 8388496
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:1000
        bgp-tunnel-encap:MPLS
        mac-mobility:Seq:3/Static
"
```

Conclusion

ISID-based MAC-flush speeds up convergence after a SAP or spoke-SDP failure, triggering a selective CMAC flush on the receiving nodes, which flushes all CMAC entries associated with that ISID and BMAC. The feature can be enabled per I-VPLS and disabled for those SAPs or spoke-SDPs for which no alternative route is available, or for those SAPs that are contained in an all-active Ethernet Segment. The BMAC/ISID update always contains the source-BMAC, not the ES-BMAC. CMAC flush based on ES-BMAC is not performed per ISID.

PBB-EVPN ISID-based Route Targets

This chapter provides information about PBB-EVPN ISID-based Route Targets.

Topics in this chapter include:

Applicability

This chapter was initially written based on SR OS Release 15.0.R4, but the CLI in the current edition corresponds to SR OS Release 21.5.R1. PBB-EVPN ISID-based route targets are supported in SR OS Release 15.0.R1, and later.

Overview

The following BGP-EVPN routes are used in PBB-EVPN according to RFC 7623:

- B-MAC routes—based on BGP-EVPN route type 2—are sent with the B-VPLS Route Target (RT), so they are sent to all the PEs where the B-VPLS is defined.
- Ethernet Segment (ES) routes—route type 4—are used for multi-homing. ES routes are sent with an RT auto-derived from the ES Identifier (ESI). If the RT-constraint is enabled, the routes are sent to only those PEs that are part of the ES.
- Inclusive Multicast Ethernet Tag (IMET) routes—route type 3—are used for the setup of per-ISID flooding domains and can be sent with a B-VPLS RT or with an ISID-based RT.
  - IMET routes are, by default, sent with a B-VPLS RT (referred to as IMET/0 routes), so they are imported by all the PEs where the B-VPLS is defined, as per RFC 7623, and supported in SR OS Release 13.0.R4, and later.
  - IMET routes with an ISID-based RT (referred to as IMET/ISID routes) are imported by only the PEs where the ISID is defined. RFC 7623 recommends these routes for deployments where the ISIDs are sparsely distributed in the network. This is supported in SR OS Release 15.0.R1, and later. The service ISID is encoded in the Ethernet tag field.

Figure 247: PBB-EVPN B-VPLS-based RT shows how the B-MAC and IMET routes with a B-VPLS RT sent by PE-1 are advertised to all other PEs (via the Route Reflector (RR)), regardless of the ISID.

Figure 247: PBB-EVPN B-VPLS-based RT

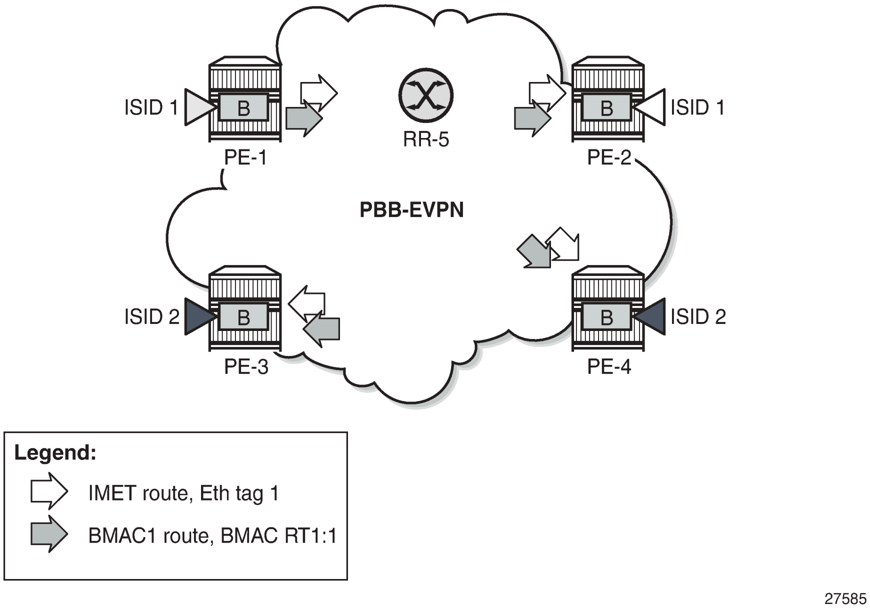
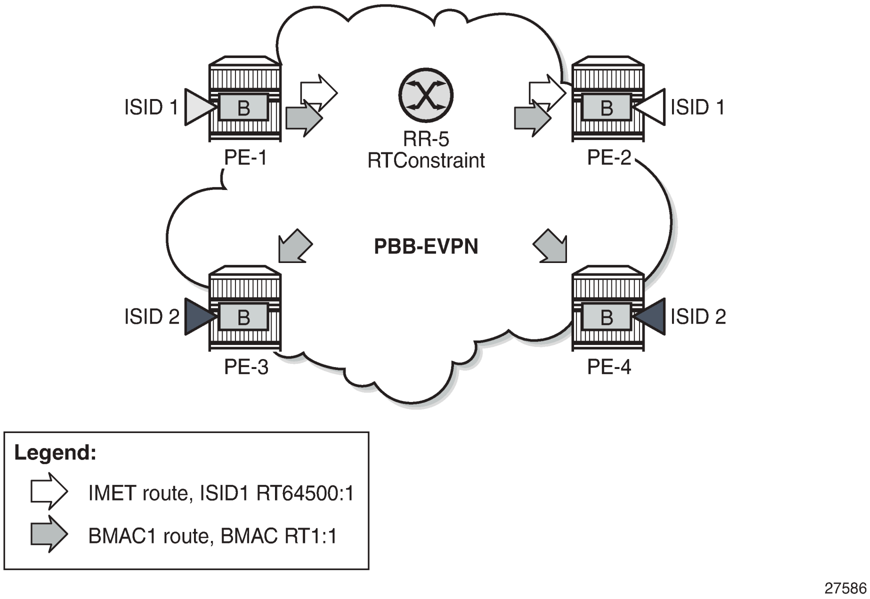


Figure 248: PBB-EVPN ISID-based RT shows how the B-MAC routes are sent to all PEs within the B-VPLS, whereas the IMET routes sent by PE-1 are selectively reflected by the RR (due to RT-constraints) and only sent to PE-2, which is the only PE with the same ISID.

Figure 248: PBB-EVPN ISID-based RT



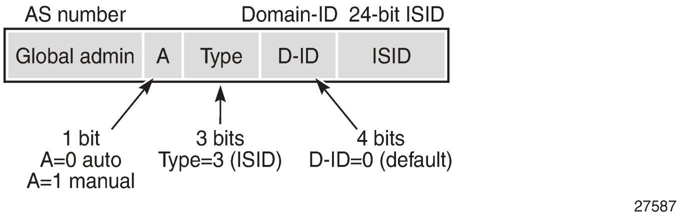
IMET routes with ISID-based RTs (IMET/ISID) can significantly reduce the number of IMET/ISID routes distributed by the RRs. The RT for the IMET/ISID route can be auto-derived from the corresponding Ethernet tag (ISID).

In addition to RFC 7623, the ISID-derived RTs can be used for BMAC/ISID routes if ISID-based CMAC flush is enabled, as per *draft-snr-bess-pbb-evpn-isid-cmacflush*. The service ISID is encoded in the Ethernet tag field.

PBB-EVPN ISID-based RT format

Figure 249: PBB-EVPN ISID-based RT format shows the ISID-based RT format:

Figure 249: PBB-EVPN ISID-based RT format



For an auto-derived ISID-based RT, the values are as follows:

- The Autonomous System (AS) number is obtained from the **config router autonomous-system** command:
  - Value = 2-byte AS number
  - For AS numbers with more than 2 bytes, the low-order 16-bit value is used.
- A = 0 for auto-derivation
- Type = 011 = 3 for ISID-based RT
- Domain ID = 0000 (default)
- ISID value

The auto-derived RT will be AS:00110000+ISID = AS:0x30+ISID Hex.

The type and sub-type of the BGP extended community is 0x00 and 0x02.

Enabling ISID-based RT

The following command is used to enable ISID-based RT for specific ISID ranges for IMET/ISID and BMAC/ISID routes.

```
*A:PE-1>config>service>vpls>bgp-evpn$ isid-route-target ?
  - isid-route-target

 [no] isid-range      - Configure ISID range information
```

The RT to be used for the I-VPLS can be auto-derived or explicitly configured.

```
*A:PE-1>configure>service>vpls>bgp-evpn>isid-route-target# isid-range ?
  - isid-range <from> [to <to>] auto-rt
  - isid-range <from> [to <to>] route-target <rt>
  - no isid-range <from>
```

```
 <from>               : [1..16777215]
 <to>                 : [1..16777215]
 <rt>                 : target:{<ip-addr:comm-val>|<2byte-asnumber:ext-comm-val>|
                        <4byte-asnumber:comm-val>}
                        ip-addr        - a.b.c.d
                        comm-val       - [0..65535]
                        2byte-asnumber - [0..65535]
                        ext-comm-val   - [0..4294967295]
                        4byte-asnumber - [0..4294967295]
```

The following configures an ISID range from 20 to 29 with auto-derived RT, whereas ISID 30 has a manually configured RT of 64500:30.

```
# on PE-1:
configure
    service
        vpls "B-VPLS 100"
            bgp-evpn
                isid-route-target
                    isid-range 20 to 29 auto-rt
                    isid-range 30 route-target target:64500:30
                exit
```

If **isid-route-target** is enabled, the IMET/ISID and BMAC/ISID route processing is modified in the export and import directions:

- "Exported IMET/ISID and BMAC/ISID routes:
  - IMET/ISID routes are sent with an ISID-based RT for the local I-VPLS ISIDs and static ISIDs, unless the ISID is contained in an ISID policy for which **no advertise-local** is configured.
  - When **isid-route-target** and **send-bvpls-evpn-flush** are both enabled for an I-VPLS, the BMAC/ISID route will also be sent with the ISID-based RT instead of the B-VPLS-based RT.
  - The **isid-route-target** command has impact only on IMET/ISID and BMAC/ISID, not on IMET/0, BMAC/0, or ES routes.
  - When a new ISID-based RT is added for an I-VPLS, a BGP update is sent for the existing IMET/ISID and BMAC/ISID routes. The new RT will be added when the routes are advertised.
- Imported IMET/ISID and BMAC/ISID routes:
  - When **isid-route-target** is enabled for an I-VPLS, BGP will start importing IMET/ISID routes and—if **bgp-evpn accept-ivpls-evpn-flush** is enabled—BMAC/ISID routes with ISID-based RTs.
  - ISID-based RTs are added for import operations when the I-VPLS is associated with the B-VPLS (regardless of the operational state of the I-VPLS) and/or when the static ISID has been added.
  - Ensure that the ISID-based RTs are configured consistently in the network. The system does not keep a mapping of RTs and ISIDs for imported routes.
  - The system will not check the format of the received auto-derived RTs. Routes will be imported when the RT is on the list of RTs for the B-VPLS.
- When **isid-route-target** is configured for an I-VPLS, VSI import/export policies are blocked in the B-VPLS, whereas BGP import/export policies are allowed and matching on the export ISID-based RT is supported.

Some other considerations:

- ISID ranges cannot overlap within a B-VPLS, but they can overlap across different B-VPLSs.
- The explicitly configured RT is meant to be used in two cases:

- ISID aggregation - when multiple ISIDs are using the same ISID RT
- Interoperability - in case the peer sends an RT in a different format

ISID-based RTs and RT-constraint

The use of the RT-constraint feature (BGP family route-target) maximizes the benefits of using different RTs per ISID; therefore, service providers are expected to enable both ISID-based RTs and RT-constraint. RT-constraint is enabled by adding the BGP address family route-target in the general BGP settings, per group, or per neighbor, as follows:

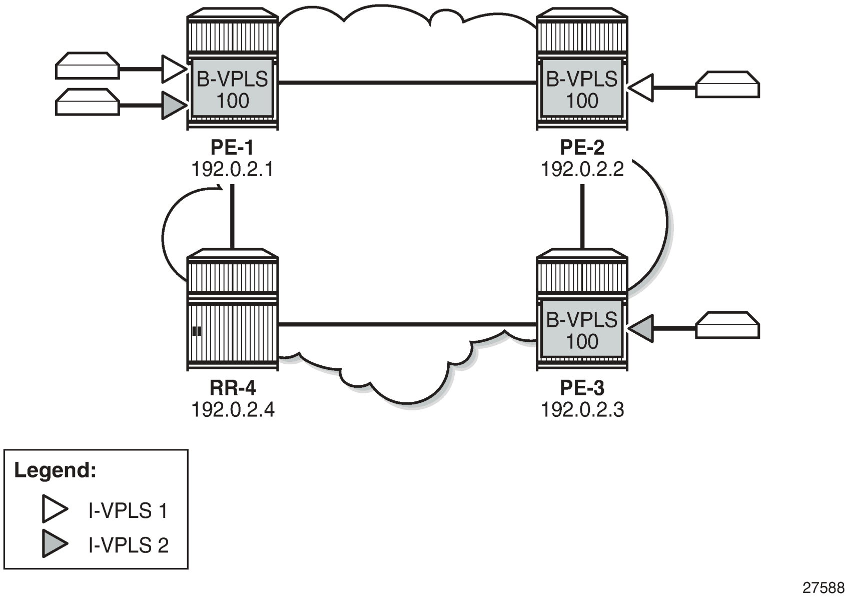
```
configure router bgp family route-target evpn
configure router bgp group "internal" family route-target evpn
configure router bgp group "internal" neighbor 192.0.2.4 family route-target evpn
```

The system will advertise the RT-constraint route when the I-VPLS is associated with the B-VPLS, regardless of the operational state of the I-VPLS. However, the IMET/ISID and the BMAC/ISID routes are sent based on the I-VPLS operational state.

Configuration

Figure 250: Example topology shows the example topology with three PEs and an RR.

Figure 250: Example topology



Initial configuration

The initial configuration on the nodes includes the following:

- Cards, MDAs, ports
- Router interfaces
- IS-IS enabled on all router interfaces (alternatively, OSPF could be used)
- SR-ISIS enabled on the PEs (but disabled on the RR)

BGP is configured on all PEs for address family EVPN, as follows.

```
# on PE-1, PE-2, PE-3:
configure
    router Base
        autonomous-system 64500
        bgp
            family evpn
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                peer-as 64500
                neighbor 192.0.2.4
                exit
            exit
        exit
```

On RR-4, BGP is configured as follows:

```
# on RR-4:
configure
    router Base
        autonomous-system 64500
        bgp
            family evpn
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                cluster 1.1.1.1
                peer-as 64500
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.2
                exit
                neighbor 192.0.2.3
                exit
            exit
        exit
```

For the RT-constraint feature, the route-target address family can be configured in combination with the EVPN address family; see section ISID-based RTs and RT-constraint.

The initial service configuration on PE-1 without ISID-based RTs is as follows:

```
# on PE-1:
configure
    service
        system
            bgp-auto-rd-range 192.0.2.1 comm-val 10 to 99
        exit
        vpls 100 name "B-VPLS 100" customer 1 b-vpls create
            service-mtu 2000
            pbb
```

```
                source-bmac 00:00:00:00:00:01
            exit
            bgp
            exit
            bgp-evpn
                evi 100
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            no shutdown
        exit
        vpls 1 name "I-VPLS 1" customer 1 i-vpls create
            pbb
                backbone-vpls 100
                exit
            exit
            bgp
                route-distinguisher auto-rd
                route-target export target:64500:1 import target:64500:1
            exit
            stp
                shutdown
            exit
            sap 1/2/1:1 create
                no shutdown
            exit
            no shutdown
        exit
        vpls 2 name "I-VPLS 2" customer 1 i-vpls create
            pbb
                backbone-vpls 100
                exit
            exit
            bgp
                route-distinguisher auto-rd
                route-target export target:64500:2 import target:64500:2
            exit
            stp
                shutdown
            exit
            sap 1/2/1:2 create
                no shutdown
            exit
            no shutdown
        exit
```

The service configuration on PE-2 is similar, but only I-VPLS 1 is configured. On PE-3, only I-VPLS 2 is configured.

PE-1 sends the following default BGP-EVPN IMET/0 update to the RR:

```
2 2021/05/28 08:55:18.406 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 77
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.1
        Type: EVPN-INCL-MCAST Len: 17 RD: 192.0.2.1:100, tag: 0, orig_addr len: 32,
                              orig_addr: 192.0.2.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:100
        bgp-tunnel-encap:MPLS
    Flag: 0xc0 Type: 22 Len: 9 PMSI:
        Tunnel-type Ingress Replication (6)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf: not required]
        MPLS Label 8388560
        Tunnel-Endpoint 192.0.2.1
"
```

The following BGP-EVPN IMET routes are received on PE-1. Toward each other PE, there is a route with Ethernet tag 0; toward PE-2, there is a route with Ethernet tag 1 for ISID 1; toward PE-3, there is a route with Ethernet tag 2.

```
*A:PE-1# show router bgp routes evpn incl-mcast
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Flag  Route Dist.         OrigAddr
      Tag                 NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.2:100       192.0.2.2
      0                   192.0.2.2

u*>i  192.0.2.2:100       192.0.2.2
      1                   192.0.2.2

u*>i  192.0.2.3:100       192.0.2.3
      0                   192.0.2.3

u*>i  192.0.2.3:100       192.0.2.3
      2                   192.0.2.3

-------------------------------------------------------------------------------
Routes : 4
===============================================================================
```

All these routes have a B-VPLS-based RT equal to 64500:100, as follows:

```
*A:PE-1# show router bgp routes evpn incl-mcast detail | match Community
Community      : target:64500:100 bgp-tunnel-encap:MPLS
Community      : target:64500:100 bgp-tunnel-encap:MPLS
Community      : target:64500:100 bgp-tunnel-encap:MPLS
Community      : target:64500:100 bgp-tunnel-encap:MPLS
Community      : target:64500:100 bgp-tunnel-encap:MPLS
Community      : target:64500:100 bgp-tunnel-encap:MPLS
Community      : target:64500:100 bgp-tunnel-encap:MPLS
Community      : target:64500:100 bgp-tunnel-encap:MPLS
```

In the preceding output, each of the four inclusive multicast routes occurs twice: the first time with the original attributes, the second time with the modified attributes, but in this example, the attribute did not change.

For the EVPN MAC routes, the output is similar. ISID-based CMAC flush is not enabled yet, so there are only BMAC/0 routes, no BMAC/ISID routes, as follows:

```
*A:PE-1# show router bgp routes evpn mac
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.2:100       00:00:00:00:00:02 ESI-0
      0                   Static            LABEL 524285
                          n/a
                          192.0.2.2

u*>i  192.0.2.3:100       00:00:00:00:00:03 ESI-0
      0                   Static            LABEL 524285
                          n/a
                          192.0.2.3

-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

Both EVPN MAC routes have the same B-VPLS-based RT with value 64500:100, as follows:

```
*A:PE-1# show router bgp routes evpn mac detail | match Community
Community      : target:64500:100 bgp-tunnel-encap:MPLS
Community      : target:64500:100 bgp-tunnel-encap:MPLS
Community      : target:64500:100 bgp-tunnel-encap:MPLS
Community      : target:64500:100 bgp-tunnel-encap:MPLS
```

ISID-based RTs

On the PEs, B-VPLS 100 is configured with ISID-based RTs, but initially without ISID-based CMAC flush, as follows:

```
# on PE-1, PE-2:
configure
    service
        vpls "B-VPLS 100"
            bgp-evpn
                isid-route-target
                    isid-range 1 to 2 auto-rt
                    isid-range 10 to 11 route-target target:64500:10
                exit
            exit
```

B-VPLS 100 has two ISID-ranges configured:

- For ISIDs 1 and 2, the RT is auto-derived. The hexadecimal value for ISID 1 is 0x30000001, which corresponds to decimal value 805306369. The hexadecimal value for ISID 2 is 0x30000002 (decimal value 805306370). For ISID 1, the RT is 64500: 805306369; for ISID 2, the RT is 64500: 805306370.
- For ISIDs 10 and 11, the RT is manually configured as 64500:10.

The configuration is identical on PE-2. On PE-3, only ISID range 2 is configured, as follows:

```
# on PE-3:
configure
    service
        vpls "B-VPLS 100"
            bgp-evpn
                isid-route-target
                    isid-range 2 auto-rt
                exit
            exit
```

On PE-1, the same four BGP-EVPN IMET routes are shown, as follows:

```
*A:PE-1# show router bgp routes evpn incl-mcast
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Inclusive-Mcast Routes
===============================================================================
Flag  Route Dist.         OrigAddr
      Tag                 NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.2:100       192.0.2.2
      0                   192.0.2.2

u*>i  192.0.2.2:100       192.0.2.2
      1                   192.0.2.2

u*>i  192.0.2.3:100       192.0.2.3
      0                   192.0.2.3

u*>i  192.0.2.3:100       192.0.2.3
      2                   192.0.2.3

-------------------------------------------------------------------------------
Routes : 4
```

The IMET route with Ethernet tag 1 now has RT 64500:805306369 (ISID 1) and the IMET route with Ethernet tag 2 has RT 64500:805306370 (ISID 2), as follows:

```
*A:PE-1# show router bgp routes evpn incl-mcast detail | match Community
Community      : target:64500:100 bgp-tunnel-encap:MPLS
Community      : target:64500:100 bgp-tunnel-encap:MPLS
Community      : target:64500:805306369 bgp-tunnel-encap:MPLS
Community      : target:64500:805306369 bgp-tunnel-encap:MPLS
Community      : target:64500:100 bgp-tunnel-encap:MPLS
Community      : target:64500:100 bgp-tunnel-encap:MPLS
Community      : target:64500:805306370 bgp-tunnel-encap:MPLS
Community      : target:64500:805306370 bgp-tunnel-encap:MPLS
```

Again, each route has two identical entries in the preceding command: one with the original attributes and another with the modified attributes.

The following BGP-EVPN IMET/ISID route is sent by PE-1 for ISID 1. The Ethernet tag is 1 and the RT is 64500:805306369.

```
# on PE-1:
11 2021/05/28 08:59:47.220 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 77
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.1
        Type: EVPN-INCL-MCAST Len: 17 RD: 192.0.2.1:100, tag: 1, orig_addr len: 32,
                              orig_addr: 192.0.2.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:805306369
        bgp-tunnel-encap:MPLS
    Flag: 0xc0 Type: 22 Len: 9 PMSI:
        Tunnel-type Ingress Replication (6)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf: not required]
        MPLS Label 8388560
        Tunnel-Endpoint 192.0.2.1
"
```

The following BGP-EVPN IMET/ISID route is sent by PE-1 for ISID 2. The Ethernet tag is 2 and the RT is 64500:805306370.

```
# on PE-1:
12 2021/05/28 08:59:47.220 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 77
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.1
        Type: EVPN-INCL-MCAST Len: 17 RD: 192.0.2.1:100, tag: 2, orig_addr len: 32,
                              orig_addr: 192.0.2.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:805306370
        bgp-tunnel-encap:MPLS
    Flag: 0xc0 Type: 22 Len: 9 PMSI:
        Tunnel-type Ingress Replication (6)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf: not required]
        MPLS Label 8388560
        Tunnel-Endpoint 192.0.2.1
```

```
"
```

When a SAP (or SDP binding) is added with static ISID 11, RT 64500:10 will be added. The service configuration on PE-1 is modified as follows:

```
# on PE-1:
configure
    service
        vpls "B-VPLS 100"
            bgp-evpn
                isid-route-target
                    isid-range 1 to 2 auto-rt
                    isid-range 10 to 11 route-target target:64500:10
                exit
            exit
            isid-policy
                entry 10 create
                    range 11
                exit
            exit
            sap 1/1/1:100 create
                static-isid
                    range 1 create isid 11
                    exit
                exit
            exit
```

The configuration is similar on PE-2. Only on PE-1 and PE-2, SAPs are configured, with static ISID 11. The following IMET/ISID route with RT 64500:10 is sent by PE-1:

```
# on PE-1:
13 2021/05/28 08:59:47.251 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 77
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.1
        Type: EVPN-INCL-MCAST Len: 17 RD: 192.0.2.1:100, tag: 11, orig_addr len: 32,
                              orig_addr: 192.0.2.1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:10
        bgp-tunnel-encap:MPLS
    Flag: 0xc0 Type: 22 Len: 9 PMSI:
        Tunnel-type Ingress Replication (6)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf: not required]
        MPLS Label 8388560
        Tunnel-Endpoint 192.0.2.1
"
```

This RT 64500:10 is not auto-derived, but configured manually for ISID range 10 to 11.

ISID-based CMAC flush

ISID-based CMAC flush is described in chapter PBB-EVPN ISID-based CMAC Flush and requires the following configuration on PE-1:

```
# on PE-1:
configure
    service
        vpls "I-VPLS 1"
            pbb
                send-bvpls-evpn-flush
            exit
        exit
        vpls "I-VPLS 2"
            pbb
                send-bvpls-evpn-flush
            exit
        exit
        vpls "B-VPLS 100"
            bgp-evpn
                accept-ivpls-evpn-flush
            exit
        exit
```

The configuration on PE-2 and PE-3 is similar, but only needs to be applied for I-VPLS 1 on PE-2 (I-VPLS 2 is not configured on PE-2) and for I-VPLS 2 on PE-3. The configuration for B-VPLS 100 is the same on all PEs.

When ISID-based CMAC flush is enabled on the PEs, additional BGP-EVPN MAC routes are sent by PE-1 for ISIDs 1 and 2:

```
27 2021/05/28 09:02:38.769 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 89
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.1
        Type: EVPN-MAC Len: 33 RD: 192.0.2.1:100 ESI: ESI-0, tag: 2, mac len: 48
                       mac: 00:00:00:00:00:01, IP len: 0, IP: NULL, label1: 8388560
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:805306370
        bgp-tunnel-encap:MPLS
        mac-mobility:Seq:0/Static
"

25 2021/05/28 09:02:38.769 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 89
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.1
        Type: EVPN-MAC Len: 33 RD: 192.0.2.1:100 ESI: ESI-0, tag: 1, mac len: 48
                       mac: 00:00:00:00:00:01, IP len: 0, IP: NULL, label1: 8388560
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 24 Extended Community:
        target:64500:805306369
        bgp-tunnel-encap:MPLS
        mac-mobility:Seq:0/Static
"
```

The BGP-EVPN MAC routes for ISIDs 1 and 2 use the same auto-derived RT values as the IMET/ISID routes. The following four BGP-EVPN MAC routes are received in PE-1:

```
*A:PE-1# show router bgp routes evpn mac
===============================================================================
 BGP Router ID:192.0.2.1        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.2:100       00:00:00:00:00:02 ESI-0
      0                   Static            LABEL 524285
                          n/a
                          192.0.2.2

u*>i  192.0.2.2:100       00:00:00:00:00:02 ESI-0
      1                   Static            LABEL 524285
                          n/a
                          192.0.2.2

u*>i  192.0.2.3:100       00:00:00:00:00:03 ESI-0
      0                   Static            LABEL 524285
                          n/a
                          192.0.2.3

u*>i  192.0.2.3:100       00:00:00:00:00:03 ESI-0
      2                   Static            LABEL 524285
                          n/a
                          192.0.2.3

-------------------------------------------------------------------------------
Routes : 4
===============================================================================
```

The BMAC/0 routes have an RT based on the B-VPLS, whereas the BMAC/ISID routes have an RT derived from the ISID, as follows:

```
*A:PE-1# show router bgp routes evpn mac detail | match Community
Community      : target:64500:100 bgp-tunnel-encap:MPLS
Community      : target:64500:100 bgp-tunnel-encap:MPLS
Community      : target:64500:805306369 bgp-tunnel-encap:MPLS
Community      : target:64500:805306369 bgp-tunnel-encap:MPLS
Community      : target:64500:100 bgp-tunnel-encap:MPLS
Community      : target:64500:100 bgp-tunnel-encap:MPLS
Community      : target:64500:805306370 bgp-tunnel-encap:MPLS
Community      : target:64500:805306370 bgp-tunnel-encap:MPLS
```

ISID-based RTs and RT-constraint

To show that RT BGP updates are sent when the I-VPLS is associated with the B-VPLS, the I-VPLSs are initially disassociated from B-VPLS 100 on PE-1, as follows:

```
# on PE-1:
configure
    service
        vpls "I-VPLS 1"
            pbb
                no backbone-vpls
            exit
        exit
        vpls "I-VPLS 2"
            pbb
                no backbone-vpls
            exit
        exit
```

The BGP configuration is modified on all nodes to include address families route-target and EVPN, as follows:

```
# on PE-1, PE-2, PE-3, RR-4:
configure
    router Base
        bgp
            family route-target evpn
```

The following RT-constraint route is sent by PE-1 after I-VPLS 1 is associated with B-VPLS 100. The RT is auto-derived from the ISID 1:

```
# on PE-1:
configure
    service
        vpls "I-VPLS 1"
            pbb
                backbone-vpls 100
            exit
        exit
        vpls "I-VPLS 2"
            pbb
                backbone-vpls 100
            exit
        exit
```

```
# on PE-1:
73 2021/05/28 09:09:34.587 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 47
    Flag: 0x90 Type: 14 Len: 22 Multiprotocol Reachable NLRI:
        Address Family RTC_V4
        NextHop len 4 NextHop 192.0.2.1
        [RT-Const-V4] origin-as 64500, Target target:64500:805306369
    Flag: 0x40 Type: 1 Len: 1 Origin: 2
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
"
```

When the I-VPLS goes operationally down, the IMET/ISID and BMAC/ISID routes are withdrawn, but not the RT-constraint route.

```
# on PE-1:
configure
    service
        vpls "I-VPLS 1"
            shutdown
```

```
# on PE-1:
83 2021/05/28 09:10:33.458 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 61
    Flag: 0x90 Type: 15 Len: 57 Multiprotocol Unreachable NLRI:
        Address Family EVPN
        Type: EVPN-INCL-MCAST Len: 17 RD: 192.0.2.1:100, tag: 1, orig_addr len: 32,
                       orig_addr: 192.0.2.1
        Type: EVPN-MAC Len: 33 RD: 192.0.2.1:100 ESI: ESI-0, tag: 1, mac len: 48
                       mac: 00:00:00:00:00:01, IP len: 0, IP: NULL, label1: 0
"
```

The RT-constraint route is withdrawn when the I-VPLS is disassociated from B-VPLS 100, as follows:

```
# on PE-1:
configure
    service
        vpls "I-VPLS 1"
            pbb
                no backbone-vpls
```

```
# on PE-1:
84 2021/05/28 09:11:28.205 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 20
    Flag: 0x90 Type: 15 Len: 16 Multiprotocol Unreachable NLRI:
        Address Family RTC_V4
        [RT-Const-V4] origin-as 64500, Target target:64500:805306369
"
```

Conclusion

PBB-EVPN ISID-based RTs, in combination with RT-constraint, reduce the number of advertised IMET routes to only those nodes where the ISID is configured. The ISID-based RT can be auto-derived from the ISID or configured manually. When ISID-based CMAC flush is also enabled, the BMAC/ISID routes will contain the same auto-derived RT.

PBB-VPLS

This chapter provides information about Provider Backbone Bridging (PBB) in a Multi-Protocol Label Switching (MPLS) based network.

Topics in this chapter include:

Applicability

This chapter is applicable to SR OS and was initially written for SR OS Release 7.0.R6. The CLI in the current edition is based on SR OS Release 20.10.R2.

Note:
Although it can be used in an MPLS-based PBB network as explained in this document, the MAC notification feature for dual-homed access is normally used in native PBB networks.

Overview

RFC 7041, *Extensions to the Virtual Private LAN Service (VPLS) Provider Edge (PE) Model for Provider Backbone Bridging*, describes the PBB-VPLS model supported by SR OS. This model expands the VPLS PE model to support PBB as defined by the IEEE 802.1ah.

PBB-VPLS combines the best of the PBB and VPLS technologies to deliver the most scalable multi-point Layer 2 VPN in the market. PBB-VPLS inherits all the benefits derived from MPLS (for example, sub-50ms Fast Reroute (FRR) protection, Traffic Engineering (TE), no need for Multiple Spanning Tree Protocol (MSTP) in the backbone) while greatly increasing the scalability of the network by providing MAC hiding, service multiplexing, and pseudowire aggregation.

The SR OS PBB-VPLS implementation also includes support for:

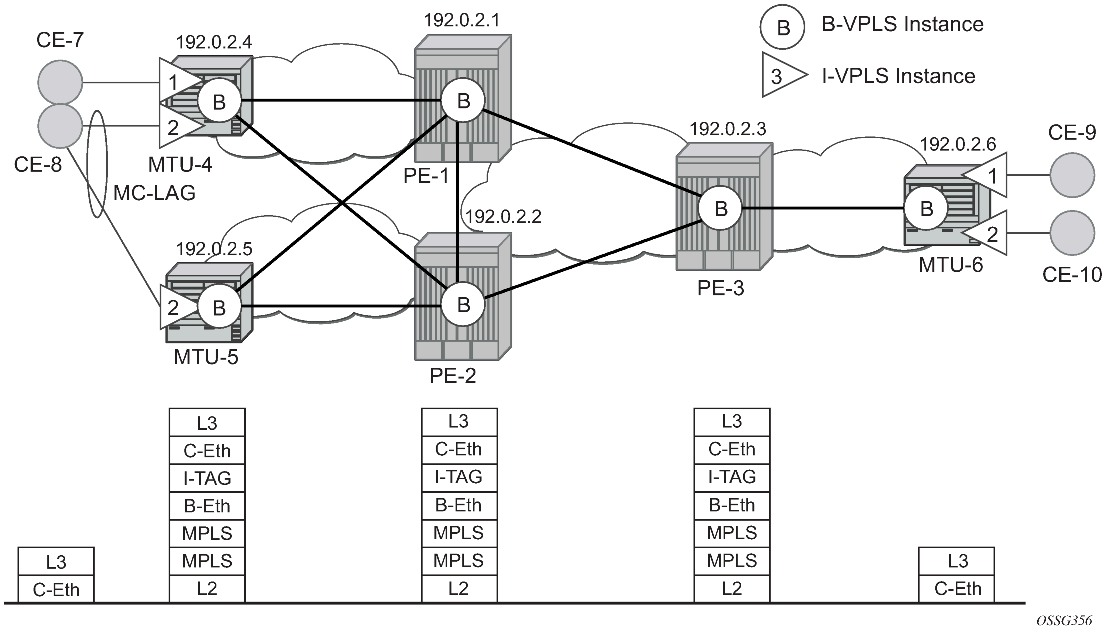
- Multiple MAC Registration Protocol (MMRP), application within IEEE 802.1ak for flood containment in the backbone instances, as specified in Section 6 of RFC 7041.
- Extensions to LDP signaling for PBB-VPLS, according to *draft-balus-l2vpn-pbb-ldp-ext-00*. These extensions will avoid network black-hole issues, as described in the Section 3 of the mentioned draft.

This chapter describes how to configure and troubleshoot a PBB-VPLS network.

Knowledge of the VPLS and H-VPLS (RFC 4762, *Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling*) architecture and functionality is assumed throughout this chapter. The most relevant concepts will be briefly explained in this chapter. For further information, see the relevant Nokia documentation.

Figure 251: Example topology including B-VPLS, I-VPLSs, and protocol stacks shows the example topology that will be used throughout the rest of the chapter along with the protocol stack being used along the path between the CEs.

Figure 251: Example topology including B-VPLS, I-VPLSs, and protocol stacks



The topology consists of three core nodes (PE-1, PE-2, and PE-3) and three Multi-Tenant Unit (MTU) nodes connected to the core. A backbone VPLS instance (B-VPLS 100) will be defined in all the six nodes, whereas a few customer I-VPLS instances will be defined on the three MTU nodes.

Those I-VPLS instances will be multiplexed into the common B-VPLS, using the ISID field within the I-TAG as the demultiplexer field at the egress MTU to differentiate each specific customer.

The B-VPLS domain constitutes an H-VPLS network itself, with spoke-SDPs from the MTUs to the core PE layer. Active/standby spoke-SDPs can be used from the MTUs to the PEs (for example, in the MTU-4 and MTU-5 cases) or single non-redundant spoke-SDPs (for example, MTU-6). CE-8 is dual-connected to the service provider network through MC-LAG.

Configuration

This section describes all the relevant PBB-VPLS configuration tasks for the setup shown in Figure 251: Example topology including B-VPLS, I-VPLSs, and protocol stacks. The appropriate associated IP/MPLS configuration is out of the scope of this example. In this particular example, the following protocols will be configured beforehand:

- ISIS-TE as IGP with all the interfaces being Level-2 (OSPF-TE could have been used instead).
- RSVP-TE as the MPLS protocol to signal the transport tunnels (LDP could have been used instead).
- LSPs between core PEs will be fast reroute protected (facility bypass tunnels) whereas LSP tunnels between MTUs and PEs will not be protected.

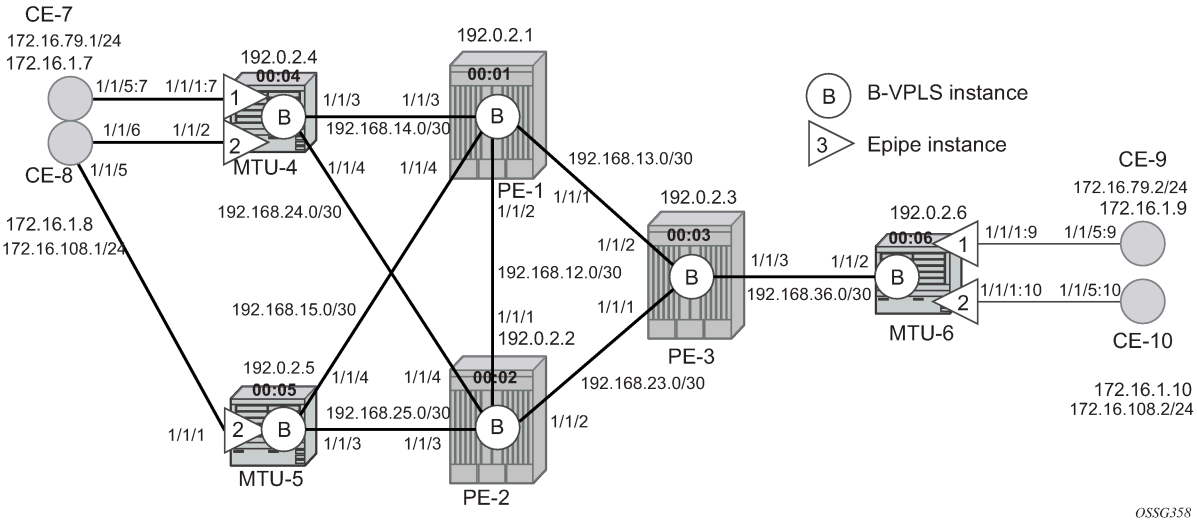
- The protection between MTU-4, MTU-5 and PE-1, PE-2 will be based on the active/standby pseudowire protection configured in the B-VPLS.
- BGP is configured for auto-discovery (Layer 2-VPN family), because FEC 129 will be used for the pseudowires between PEs in the core.

Once the IP/MPLS infrastructure is up and running, the service configuration tasks described in the following sections can be implemented.

PBB-VPLS M:1 service configuration

This section explains the process to configure PBB-VPLS services in a M:1 fashion, M being the number of customer I-VPLS services multiplexed into the same B-VPLS instance (instance 100). An alternative configuration is 1:1, where each customer I-VPLS has its own B-VPLS. MTU-4 and PE-1 will be picked to show the relevant CLI configuration commands. The bold digits separated by colons *00:xx* are abbreviations for the backbone MAC addresses. Figure 252: Example topology with port numbers and IP addresses shows the example topology.

Figure 252: Example topology with port numbers and IP addresses



B-VPLS configuration

The first step is to configure the B-VPLS instance that will carry the PBB traffic. The following shows the B-VPLS configuration on MTU-4 and PE-1. The configuration on MTU-5 and MTU-6 resembles the configuration on MTU-4; the configuration on PE-2 and PE-3 resembles the configuration on PE-1.

The configuration for B-VPLS 100 on MTU-4 is as follows:

```
# on MTU-4:
configure
    service
        vpls 100 name "B-VPLS 100" customer 1 b-vpls create
            endpoint "core" create
                no suppress-standby-signaling
            exit
            service-mtu 2000
            pbb
                source-bmac 00:04:04:04:04:04
            exit
            spoke-sdp 41:100 endpoint "core" create
                precedence primary
            exit
            spoke-sdp 42:100 endpoint "core" create
            exit
            no shutdown
        exit
```

On PE-1, B-VPLS 100 is configured as follows:

```
# on PE-1:
configure
    service
        pw-template 1 use-provisioned-sdp create
            split-horizon-group "CORE"
            exit
        exit
        vpls 100 name "B-VPLS 100" customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:01:01:01:01:01
            exit
            bgp
                route-target export target:65000:100 import target:65000:100
                pw-template-binding 1
                exit
            exit
            bgp-ad
                vpls-id 65000:100
                no shutdown
            exit
            spoke-sdp 14:100 create
            exit
            spoke-sdp 15:100 create
            exit
            no shutdown
        exit
```

The keyword **b-vpls** is given at creation time and therefore it cannot be added to a regular existing VPLS instance. Besides the **b-vpls** keyword, the B-VPLS is a regular VPLS instance in terms of configuration, with the following exceptions:

- The B-VPLS service MTU must be at least 18 bytes greater than the I-VPLS MTU of the multiplexed instances. In this example, the I-VPLS instances will have the default service MTU (1500 bytes); therefore, any MTU equal to or greater than 1518 bytes must be configured. In this particular example, a MTU of 2000 bytes is configured in the B-VPLS instance throughout the network.
- The source B-MAC is the MAC address that will be sourced when the PBB traffic is originated from that node. A source B-MAC per B-VPLS instance can be configured (if there are more than one B-VPLS) or a common source B-MAC that will be shared by all the B-VPLS instances in the node. If no specific source B-MAC is provisioned, the system MAC address is used as the source B-MAC. When using the access multi-homing feature for native PBB, the source B-MAC must be a configured one and never the chassis MAC address. The way to configure a common B-MAC for all the B-VPLS instances on MTU-4 is as follows:

  ```
  # on MTU-4:
  ```

```
configure
    service
        pbb
            source-bmac 00:04:04:04:04:04
```

The following considerations will be taken into account when configuring the B-VPLS:

- B-VPLS SAPs:
  - Ethernet null, dot1q, and qinq encapsulations are supported
  - Default SAP (:*) types are blocked in the CLI for the B-VPLS SAP
- B-VPLS SDPs:
  - For MPLS, both mesh and spoke-SDPs with split-horizon groups are supported.
  - Similar to regular pseudowires, the outgoing PBB frame on an SDP (for example, B-pseudowire) contains a BVID qtag only if the pseudowire type is Ethernet VLAN. If the pseudowire type is *Ethernet*, the BVID q-tag is stripped before the frame goes out.
  - BGP-AD is supported in the B-VPLS; therefore, spoke-SDPs in the B-VPLS can be signaled using FEC 128 or FEC 129. In this example, BGP-AD and FEC 129 are used. A split-horizon group (SHG) has been configured to emulate the behavior of mesh-SDPs in the core.
- If a local I-VPLS instance is associated with the B-VPLS, local frames originated/terminated on local I-VPLS(s) are PBB encapsulated/de-encapsulated using the PBB Ethertype provisioned under the related port or SDP component.

By default, the PBB Ethertype is 0x88e7 (which is the standard one defined in 802.1ah for the I-TAG) but this PBB Ethertype can be changed if required due to interoperability reasons. This is the way to change it at port and/or SDP level:

```
*A:MTU-4# configure port 1/1/3 ethernet pbb-etype ?
  - pbb-etype <0x0600..0xffff>
  - no pbb-etype

 <0x0600..0xffff>     : [1536..65535] - accepts in decimal or hex
```

```
*A:MTU-4# configure service sdp 41 pbb-etype ?
  - no pbb-etype [<0x0600..0xffff>]
  - pbb-etype <0x0600..0xffff>

 <0x0600..0xffff>     : [1536..65535] - accepts in decimal or hex
```

The following commands are useful to check the actual PBB Ethertype:

```
*A:MTU-4# show service sdp 41 detail | match PBB
Bw BookingFactor     : 100                   PBB Etype          : 0x88e7
```

```
*A:MTU-4# show port 1/1/3 | match PBB
PBB Ethertype      : 0x88e7
```

I-VPLS configuration

Once the common B-VPLS is configured, the next step is to provision the customer I-VPLS instances. The following shows the relevant configuration on MTU-4 for the two I-VPLS instances represented in Figure

252: Example topology with port numbers and IP addresses. The I-VPLS instances are configured on the MTU devices, whereas the core PEs are customer-unaware nodes.

```
# on MTU-4:
configure
    service
        vpls 1 name "I-VPLS 1" customer 1 i-vpls create
            pbb
                backbone-vpls 100
                exit
            exit
            sap 1/1/1:7 create
            exit
            no shutdown
        exit
        vpls 2 name "I-VPLS 2" customer 1 i-vpls create
            pbb
                backbone-vpls 100 isid 2
                exit
            exit
            sap lag-1 create
            exit
            no shutdown
        exit
```

The keyword **i-vpls** is given at creation time and therefore it cannot be added to a regular existing VPLS instance. After creating the I-VPLS instance, it has to be linked to its corresponding transport B-VPLS instance. That link is given by the **backbone-vpls** <*b-vpls*> **isid** <*isid*> command. If no ISID (20 bit customer identification in the ITAG) is specified, the system will take the VPLS instance identifier as the ISID value.

The following considerations will be taken into account when configuring the I-VPLS:

- I-VPLS SAPs:
  - SAPs can be defined on ports with any Ethernet encapsulation type (null, dot1q, and qinq)
  - The I-VPLS SAPs can coexist on the same port with SAPs for other business services, for example, VLL and VPLS SAPs.
- I-VPLS SDPs:
  - GRE and MPLS SDPs are supported.
  - No mesh-SDPs are supported, only spoke-SDP. Mesh-SDPs can be emulated by using SHGs.

Existing SAP processing rules still apply for the I-VPLS case; the SAP encapsulation definition on Ethernet ingress ports defines which VLAN tags are used to determine the service that the packet belongs to:

- Null encapsulation defined on ingress — Any VLAN tags are ignored and the packet goes to a default service for the SAP;
- Dot1q encapsulation defined on ingress — only first VLAN tag is considered;
- QinQ encapsulation defined on ingress — both VLAN tags are considered; wildcard for the inner VLAN tag is supported.
- For dot1q/qinq encapsulations, traffic encapsulated with VLAN tags for which there is no definition is discarded.
- Any VLAN tag used for service selection on the I-SAP is stripped before the PBB encapsulation is added. Appropriate VLAN tags are added at the remote PBB PE when sending the packet out on the egress SAP.

MMRP for flooding optimization

When the M:1 model is used (as in this example), any I-VPLS broadcast, unknown unicast, or multicast (BUM) frame is flooded throughout the B-VPLS domain regardless of the nodes where the originating I-VPLS is defined. In other words, in our example in Figure 251: Example topology including B-VPLS, I-VPLSs, and protocol stacks, any BUM frame coming from CE-7 would be flooded in the B domain and would reach PE-2 and MTU-5, even though that traffic only needs to go to PE-3 and MTU-6. In order to build customer-based flooding trees and optimize the flooding, Multiple MAC Registration Protocol (MMRP) must be configured on the B-VPLS.

MMRP can be enabled with its default settings just by executing a **mrp no shutdown** command on all nodes:

```
# on all nodes:
configure
    service
        vpls "B-VPLS 100"
            mrp
                no shutdown
```

There are certain B-VPLS MRP settings that can be modified. These are the default values:

```
*A:MTU-4>config>service>vpls>mrp# info detail
----------------------------------------------
                mmrp
                    no end-station-only
                    attribute-table-size 2048
                    attribute-table-low-wmark 90
                    attribute-table-high-wmark 95
                    no flood-time
                    no shutdown
                exit
                no shutdown
----------------------------------------------
```

These attributes can be changed in order to control the number of MMRP attributes per B-VPLS and optimize the convergence time in case of failures in the B-VPLS:

- Controlling the number of attributes per B-VPLS

  The MMRP exchanges create one entry per attribute (group B-MAC) in the B-VPLS where MMRP protocol is running. PBB uses a group B-MAC address—built using a specific OUI (00:1e:83) with the multicast bit set, and the ISID value for the last 24 bits—as a destination MAC address for flooding any BUM frame into the B-domain.

  When the first registration is received for an attribute, an MFIB entry is created for it. The **attribute-table-size** allows the user to control the number of MMRP attributes (group B-MACs) created on a per B-VPLS basis, between 1 and 2048. Based on the configured size, high and low watermarks can be set (in percentage) so that alarms can be triggered upon exceeding the watermarks. This ensures that no B-VPLS will take up all the resources from the total pool. The maximum number of attributes per B-VPLS is 2048 and 4000 can be configured globally on the system.

- Optimizing the convergence time

  Assuming that MMRP is used in a certain B-VPLS, under failure conditions, the time it takes for the B-VPLS forwarding to resume may depend on the data plane and control plane convergence plus the time it takes for MMRP exchanges to stabilize the flooding trees on a per ISID basis. In order to minimize the convergence time, the PBB SR OS implementation offers the selection of a mode where B-VPLS forwarding reverts for a short time to flooding so that MMRP has enough time to converge. This mode can be selected through configuration using the **flood-time** <*value*> command where value represents the amount of time in seconds (between 3 and 600) that flooding will be enabled. If this behavior is selected, the forwarding plane starts with B-VPLS flooding for a configurable time period, then it reverts back to the MFIB entries installed by MMRP. The following B-VPLS events initiate the switch from per I-VPLS (MMRP) MFIB entries to B-VPLS flooding:

– Reception or local triggering of a Spanning Tree Topology Change Notification (TCN)
– B-SAP failure
– Failure of a B-SDP binding
– Pseudowire activation in a primary/standby H-VPLS resiliency solution
– SF/CPM switchover due to STP reconvergence

The IEEE 802.1ak standard, which defines MRP, requires the implementation of different state machines with associated timers that can be tuned. A full MRP participant maintains the following state machines:

- Registrar state machine
- Applicant state machine
- LeaveAll state machine
- PeriodicTransmission state machine

The two first state machines are maintained for each attribute in which the participant is interested, whereas the two latter are global to all the attributes.

The job of the registrar function is to record declarations of the attribute made by other participants on the LAN. A registrar does not send any protocol messages, because the applicant looks after the interests of all would-be participants.

The job of the applicant is twofold: first, to ensure that this participant’s declaration is correctly registered by other participants’ registrars, and next, to prompt other participants to register again after one withdraws a declaration.

The associated timers can be tuned on a per SAP/SDP basis:

```
*A:MTU-4>config>service>vpls>spoke-sdp# mrp ?
  - mrp

 [no] join-time       - Configure timer value in 10th of seconds for sending
                        join-messages
 [no] leave-all-time  - Configure timer value in 10th of seconds for refreshing
                        all attributes
 [no] leave-time      - Configure timer value in 10th of seconds to hold
                        attribute in leave-state
 [no] mrp-policy      - Configure mrp-policy
 [no] periodic-time   - Configure timer value in 10th of seconds for
                        re-transmission of attribute declarations
 [no] periodic-timer  - Control re-transmission of attribute declarations
```

```
*A:MTU-4>config>service>vpls>spoke-sdp>mrp# info detail
----------------------------------------------
                    join-time 2
                    leave-time 30
                    leave-all-time 100
                    periodic-time 10
                    no periodic-timer
                    no mrp-policy
----------------------------------------------
```

A brief description of the MRP SAP/SDP attributes follows:

- **join-time** — This command controls the interval between transmit opportunities that are applied to the applicant state machine. An instance of this join period timer is required on a per-port, per-MRP participant basis. For additional information, see IEEE 802.1ak-2007 section 10.7.4.1.
- **leave-time** — This command controls the period of time that the registrar state machine will wait in the leave state before transitioning to the MT state when it is removed. An instance of the timer is required for each state machine that is in the leave state. The leave period timer is set to the value leave-time when it is started. A registration is normally in "in" state where there is an MFIB entry and traffic being forwarded. When a "leave all" is performed (periodically around every 10-15 seconds per SAP/SDP binding – see leave-all-time below), a node sends a message to its peer indicating a leave all is occurring and puts all of its registrations in leave state. The peer refreshes its registrations based on the leave all PDU it receives and sends a PDU back to the originating node with the state of all its declarations. See IEEE 802.1ak-2007 section 10.7.4.2.
- **leave-all-time** — This command controls the frequency with which the leaveall state machine generates leaveall PDUs. The timer is required on a per-port, per-MRP participant basis. The leaveall period timer is set to a random value, T, in the range leave-all-time<T<1.5*leave-all-time when it is started. See IEEE 802.1ak-2007, section 10.7.4.3.
- **periodic-time** — This command controls the frequency the periodic transmission state machine generates periodic events if the periodic transmission timer is enabled. The timer is required on a per-port basis. The periodic transmission timer is set to one second when it is started.
- **periodic-timer** — This command enables or disables the periodic transmission timer.

The following command shows the MRP configuration and statistics on a per SAP/SDP basis within the B-VPLS:

```
*A:MTU-4# show service id 100 all | match MRP post-lines 10
Sdp Id 41:100 MRP Information
-------------------------------------------------------------------------------
Join Time          : 0.2 secs                 Leave Time        : 3.0 secs
Leave All Time     : 10.0 secs                Periodic Time     : 1.0 secs
Periodic Enabled   : false
Mrp Policy         : N/A
Rx Pdus            : 234                      Tx Pdus           : 252
Dropped Pdus       : 0
Rx New Event       : 0                        Rx Join-In Event  : 246
Rx In Event        : 0                        Rx Join Empty Evt : 217
Rx Empty Event     : 0                        Rx Leave Event    : 0
SDP MMRP Information
-------------------------------------------------------------------------------
MAC Address       Registered        Declared
-------------------------------------------------------------------------------
01:1e:83:00:00:01 Yes               Yes

01:1e:83:00:00:02 Yes               Yes

-------------------------------------------------------------------------------
Number of MACs=2 Registered=2 Declared=2
-------------------------------------------------------------------------------
Sdp Id 42:100 MRP Information
-------------------------------------------------------------------------------
Join Time          : 0.2 secs                 Leave Time        : 3.0 secs
Leave All Time     : 10.0 secs                Periodic Time     : 1.0 secs
Periodic Enabled   : false
```

```
Mrp Policy         : N/A
Rx Pdus            : 0                         Tx Pdus           : 0
Dropped Pdus       : 0
Rx New Event       : 0                         Rx Join-In Event  : 0
Rx In Event        : 0                         Rx Join Empty Evt : 0
Rx Empty Event     : 0                         Rx Leave Event    : 0
SDP MMRP Information
-------------------------------------------------------------------------------
MAC Address       Registered        Declared
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Number of MACs=0 Registered=0 Declared=0
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
* indicates that the corresponding row element may have been truncated.
Service MRP Information
===============================================================================
Admin State          : enabled
-------------------------------------------------------------------------------
MMRP
-------------------------------------------------------------------------------
Admin Status         : enabled              Oper Status      : up
Register Attr Cnt    : 2                    Declared Attr Cnt: 2
End-station-only     : disabled
Max Attributes       : 2048                 Attribute Count  : 2
Hi Watermark         : 95%                  Low Watermark    : 90%
Failed Registers     : 0                    Flood Time       : Off
-------------------------------------------------------------------------------
MVRP
-------------------------------------------------------------------------------
MRP SAP Table
=============================================================================
SAP                             Join      Leave     Leave All Periodic
                                Time(sec) Time(sec) Time(sec) Time(sec)
-----------------------------------------------------------------------------
=============================================================================

=============================================================================
MRP SDP-BIND Table
=============================================================================
SDP-BIND                        Join      Leave     Leave All Periodic
                                Time(sec) Time(sec) Time(sec) Time(sec)
-----------------------------------------------------------------------------
41:100                          0.2       3.0       10.0      1.0
42:100                          0.2       3.0       10.0      1.0
=============================================================================
===============================================================================

-------------------------------------------------------------------------------
```

The following command is useful to check the MRP configuration and status.

```
*A:MTU-4# show service id 100 mrp

===============================================================================
Service MRP Information
===============================================================================
Admin State          : enabled
-------------------------------------------------------------------------------
MMRP
-------------------------------------------------------------------------------
```

```
Admin Status         : enabled              Oper Status      : up
Register Attr Cnt    : 2                    Declared Attr Cnt: 2
End-station-only     : disabled
Max Attributes       : 2048                 Attribute Count  : 2
Hi Watermark         : 95%                  Low Watermark    : 90%
Failed Registers     : 0                    Flood Time       : Off
-------------------------------------------------------------------------------
MVRP
-------------------------------------------------------------------------------
Admin Status         : disabled             Oper Status      : down
Max Attr             : 4095                 Failed Register  : 0
Register Attr Count  : 0                    Declared Attr    : 0
Hi Watermark         : 95%                  Low Watermark    : 90%
Hold Time            : disabled             Attr Count       : 0
-------------------------------------------------------------------------------

=============================================================================
MRP SAP Table
=============================================================================
SAP                             Join      Leave     Leave All Periodic
                                Time(sec) Time(sec) Time(sec) Time(sec)
-----------------------------------------------------------------------------
=============================================================================

=============================================================================
MRP SDP-BIND Table
=============================================================================
SDP-BIND                        Join      Leave     Leave All Periodic
                                Time(sec) Time(sec) Time(sec) Time(sec)
-----------------------------------------------------------------------------
41:100                          0.2       3.0       10.0      1.0
42:100                          0.2       3.0       10.0      1.0
=============================================================================
===============================================================================
```

In the example throughout the chapter, as soon as MMRP is enabled, an optimized flooding tree will be built for ISID 1, because the I-VPLS 1 is only defined in MTU-4 and MTU-6, but not in MTU-5. A good way to track the flooding tree for a particular ISID is the following command:

```
*A:MTU-4# show service id 100 mmrp mac
-------------------------------------------------------------------------------
SAP/SDP                                 MAC Address       Registered  Declared
-------------------------------------------------------------------------------
sdp:41:100                              01:1e:83:00:00:01 Yes         Yes
sdp:41:100                              01:1e:83:00:00:02 Yes         Yes
-------------------------------------------------------------------------------
Number of Entries=2 SAPs=0 SDPs=2
-------------------------------------------------------------------------------
```

```
*A:MTU-5# show service id 100 mmrp mac
-------------------------------------------------------------------------------
SAP/SDP                                 MAC Address       Registered  Declared
-------------------------------------------------------------------------------
sdp:52:100                              01:1e:83:00:00:01 Yes         No
sdp:52:100                              01:1e:83:00:00:02 Yes         No
-------------------------------------------------------------------------------
Number of Entries=2 SAPs=0 SDPs=2
-------------------------------------------------------------------------------
```

The group B-MAC ending in *01* corresponds to the I-VPLS 1 whereas the one ending in *02* to the I-VPLS 2. MMRP PDUs for the two attributes are sent throughout the loop-tree topology (not over STP blocked ports or standby spoke-SDPs and observing the split-horizon rules). The two attributes are registered on every B-VPLS virtual port; however, the tree is only built on those ports where the attribute is also declared, and not only registered. For instance, the spoke-SDP 52:100 in MTU-5 will not be part of the ISID 1 or ISID 2 flooding trees. Neither attribute is declared because I-VPLS 1 does not exist on MTU-5 and I-VPLS 2 is operationally down on MTU-5 (MC-LAG SAP is in standby state, so the I-VPLS is down).

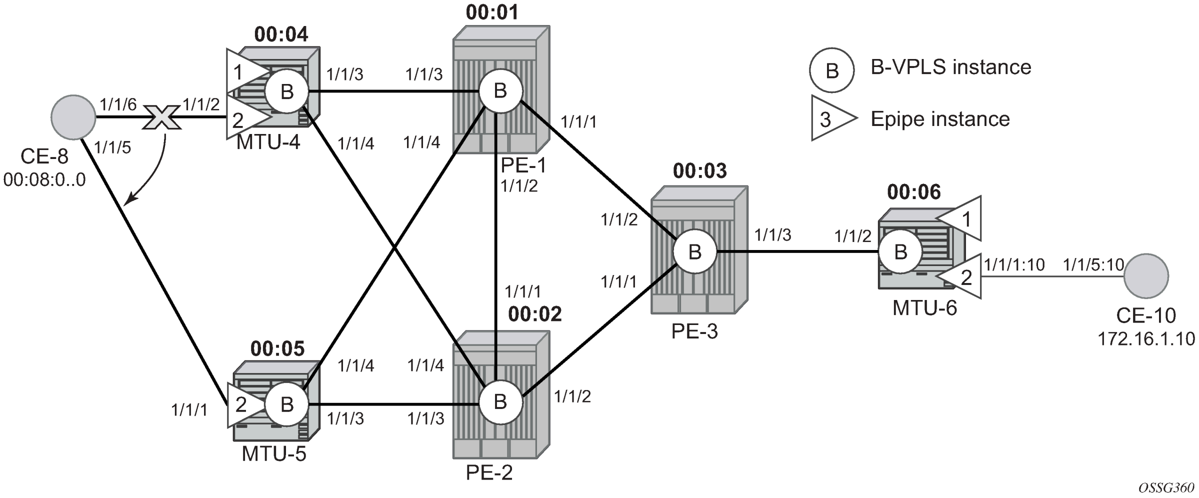
As soon as a group B-MAC attribute is registered on a particular port, an MFIB entry is added for that B-MAC on that port, regardless of the declaration state for that attribute on the port. For instance, neither B-MAC is declared on MTU-5, however, the two MFIB entries are created as soon as the attributes are registered:

```
*A:MTU-5# show service id 100 mfib

===============================================================================
Multicast FIB, Service 100
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
*               01:1e:83:00:00:01     b-sdp:52:100                 Local    Fwd
*               01:1e:83:00:00:02     b-sdp:52:100                 Local    Fwd
-------------------------------------------------------------------------------
Number of entries: 2
===============================================================================
```

MAC flush: avoiding black-holes

Both the I-VPLS and B-VPLS components inherit the MAC flush capabilities of a regular VPLS clearing the related C-MAC and respectively B-MAC FIBs. All types of MAC flush—**flush-all-but-mine** and **flush-all-from-me**—are supported together with the related CLI. In addition to these features, some extensions have been added so that MAC flush can be triggered on the B-VPLS based on some events happening on the I-VPLS. Figure 253: Black-hole shows a potential scenario where black-holes can occur if the proper configuration is not added.

Figure 253: Black-hole



Under normal conditions, the I-VPLS 2 FIB on MTU-6 shows that CE-8 MAC address is learned through B-MAC 00:04 of MTU-4:

```
*A:MTU-6# show service id 2 fdb pbb

===============================================================================
Forwarding Database, i-Vpls Service 2
===============================================================================
MAC               Source-Identifier     B-Svc     b-Vpls MAC        Type/Age
 Transport:Tnl-Id
-------------------------------------------------------------------------------
00:08:00:00:00:00 b-sdp:63:100          100       00:04:04:04:04:04 L/180
00:10:00:00:00:00 sap:1/1/1:10          100       N/A               L/0
===============================================================================
```

When a failure happens in the CE-8 MC-LAG active link, the link to MTU-5 takes over. However, the FIB on MTU-6 still points at the BMAC of MTU-4 and that will still be the B-MAC used in the PBB encapsulation. Therefore, a black-hole occurs until either bidirectional traffic is sent or the FIB aging timer expires.

The configuration in the I-VPLS can be modified to trigger a MAC flush in the B-VPLS with the following command:

```
*A:MTU-4# configure service vpls "I-VPLS 2" pbb send-bvpls-flush ?
  - send-bvpls-flush {[all-but-mine] [all-from-me]}
  - no send-bvpls-flush

 <all-but-mine>       : keyword
 <all-from-me>        : keyword
```

The following command is executed on all MTUs to solve the black-hole:

```
# on all MTUs:
configure
    service
        vpls "I-VPLS 2"
            pbb
                send-bvpls-flush all-from-me
```

By enabling **send-bvpls-flush all-from-me** on I-VPLS 2, a failure on the MC-LAG active link on I-VPLS 2 will trigger an LDP MAC **flush-all-from-me** into the B-VPLS that will flush the FIB in MTU-6 for I-VPLS 2, avoiding the black-hole. A MC-LAG failure is emulated by disabling the LAG on MTU-4, as follows:

```
# on MTU-4:
configure
    lag 1
        shutdown
```

MTU-4 sends the following LDP MAC flush for all MAC addresses learned from MTU-4:

```
1 2021/01/12 17:02:25.211 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Address Withdraw packet (msgId 263) to 192.0.2.1:0
Protocol version = 1
MAC Flush (All MACs learned from me)
Service FEC PWE3: ENET(5)/100 Group ID = 0 cBit = 0
Number of PBB-BMACs = 1
BMAC 1 = 00:04:04:04:04:04
Number of PBB-ISIDs = 1
ISID 1 = 2
```

```
Number of Path Vectors : 1
Path Vector(  1) = 192.0.2.4
"
```

On MTU-6:

```
1 2021/01/12 17:02:25.227 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Recv Address Withdraw packet (msgId 206) from 192.0.2.3:0
Protocol version = 1
MAC Flush (All MACs learned from me)
Service FEC PWE3: ENET(5)/100 Group ID = 0 cBit = 0
Number of PBB-BMACs = 1
BMAC 1 = 00:04:04:04:04:04
Number of PBB-ISIDs = 1
ISID 1 = 2
Number of Path Vectors : 3
Path Vector(  1) = 192.0.2.4
Path Vector(  2) = 192.0.2.1
Path Vector(  3) = 192.0.2.3
"
```

Immediately after receiving the MAC flush, the CE-8 MAC is flushed. The CE-8 MAC is learned again, but this time linked to the B-MAC 00:05, which is the B-MAC of MTU-5:

```
*A:MTU-6# show service id 2 fdb pbb

===============================================================================
Forwarding Database, i-Vpls Service 2
===============================================================================
MAC               Source-Identifier     B-Svc     b-Vpls MAC        Type/Age
 Transport:Tnl-Id
-------------------------------------------------------------------------------
00:08:00:00:00:00 b-sdp:63:100          100       00:05:05:05:05:05 L/0
00:10:00:00:00:00 sap:1/1/1:10          100       N/A               L/120
===============================================================================
```

The following I-VPLS events are propagated into the B-VPLS depending on the **flush-all-but-mine** or **flush-all-from-me** keywords used in the configuration:

If the **flush-all-but-mine** keyword is configured (positive flush), the following events in the I-VPLS trigger a MAC flush into the B-VPLS:

1. TCN event in one or more of the related I-VPLS/M-VPLS.
2. Pseudowire/SDP binding activation with active/standby pseudowire (standby to active or down to up).
3. Reception of an LDP MAC withdraw flush-all-but-mine in the related I-VPLS.

If the **flush-all-from-me** keyword is configured (negative flush) the following events in the I-VPLS trigger a MAC flush into the B-VPLS:

1. MC-LAG active link failure (in our example).
2. Failure of a local SAP – requires **send-flush-on-failure** to be enabled in I-VPLS.
3. Failure of a local pseudowire/SDP binding – requires **send-flush-on-failure** to be enabled in I-VPLS.
4. Reception of an LDP MAC withdraws flush-all-from-me in the related I-VPLS.

In addition to this and regardless of what type, MAC flush has been optimized to avoid flushing in the core PEs, flushing only the C-MACs mapped to a certain B-MAC (belonging to a specific ISID FIB) and the ability to indicate to core PEs which messages should always be forwarded endpoint-to-endpoint toward all

PBB PEs regardless of the propagate-mac-flush setting in B-VPLS. All of this is implemented without the need of any additional CLI commands and it is part of *draft-balus-l2vpn-pbb-ldp-ext-00*.

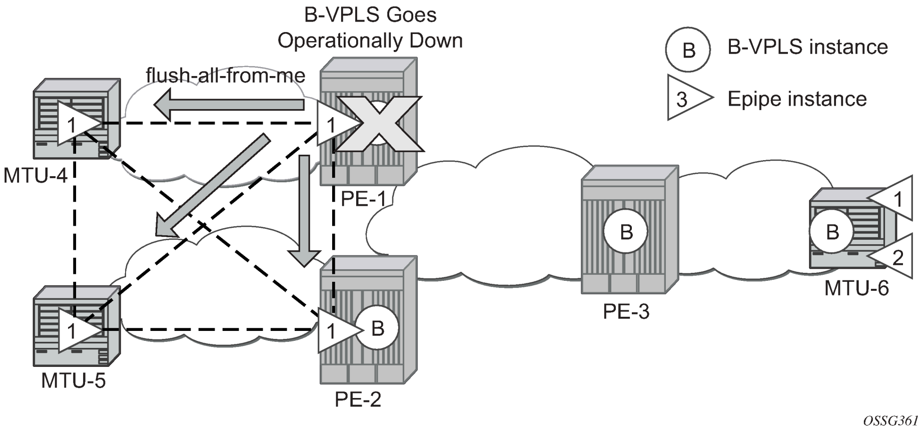
Another extension supported to avoid black-holes within this mix of I- and B-VPLS environments is the **block-on-mesh-failure** feature in PBB. When the VPLS mesh exists only in I-VPLS or in B-VPLS, and the **block-on-mesh-failure** feature is enabled, the regular VPLS behavior will apply (when all the mesh-SDPs go down an LDP notification with pseudowire status bits = 0x01—Pseudo Wire Not Forwarding—is sent over the spoke-SDPs). When the active/standby pseudowire resiliency is implemented in I-VPLS such that the PBB PE performs the role of a PE-rs, the B-VPLS core replaces the pseudowire (SDP binding) mesh. The block-on-mesh notification (LDP notification indicating pseudowire not forwarding) will be sent to the MTUs only when the related B-VPLS is operationally down. The B-VPLS core is operationally down only when all of its SAPs and SDPs are down.

The final feature that can be enabled in an I-VPLS with CLI is the **send-flush-on-bvpls-failure** feature.

```
*A:MTU-4# configure service vpls "I-VPLS 2" pbb send-flush-on-bvpls-failure ?
  - no send-flush-on-bvpls-failure
  - send-flush-on-bvpls-failure
```

This feature is required to avoid black-holes when there is a full mesh of pseudowires in the I-VPLS domain and the B-VPLS instance can go operationally down. Figure 254: Send flush on B-VPLS failure example shows a typical scenario where this feature is needed (normally when PBB-VPLS and multi-chassis end point are combined together).

Figure 254: Send flush on B-VPLS failure example



Access dual-homing and MAC notification

Although this section is focused on PBB in a MPLS based network, Nokia PBB implementation also allows the operator to use a native Ethernet infrastructure in the PBB core. Native Ethernet tunneling can be emulated using Ethernet SAPs to interconnect the related B-VPLS instances. In those cases, there is no LDP signaling available; therefore, there is no MAC flush sent when the active link in a multi-homed access device fails.

The SR OS supports a mechanism to avoid potential black-holes in native Ethernet PBB networks. In addition to the source B-MAC associated with each B-VPLS, an additional B-MAC is associated with each MC-LAG supporting Multi-homed I-VPLS SAPs. The nodes that are in a multi-homed MC-LAG configuration share a common B-MAC on the related MC-LAG interfaces. When the MAC notification is enabled (**mac-notification no shutdown**), an Ethernet CFM notification message is sent from the node holding the active link. That message will be flooded in the B-VPLS domain using the MC-LAG SAP B-MAC as the source MAC address. The remote nodes will learn the customer MAC addresses behind the MC-LAG and will link them to this new SAP B-MAC. MC-LAG will keep track of the active link for each particular LAG associated to a SAP B-MAC. Should MC-LAG detect any new active link in a node, a new CFM notification message will be flooded from the new active node.

The following caveats and considerations must be taken into account:

- Only MC-LAG is supported as dual-home mechanism.
- This mechanism is supported for native PBB and/or MPLS-based PBB-VPLS. Although it is mostly beneficial when native PBB is used in the core, it can also help to optimize the re-learning process in a MPLS-based core in case of MC-LAG failures, in addition to the existing LDP MAC flush procedures.

The example of this configuration shows the setup being used in this configuration example. MAC-notification will be configured in MTU-4 and MTU-5 for the dual-homed CE-8.

The first step is to configure the SAP B-MAC that will be used for the MAC notification messages. The **source-bmac-lsb** (source backbone MAC least significant bits) command has been added to the MC-LAG branch so that the operator can decide the two last octets to be used in the SAP B-MAC. Those two last octets can be derived from the LACP key (if the **use-lacp-key** statement is used) or can be specifically defined.

```
*A:MTU-4# configure redundancy multi-chassis peer 192.0.2.5 mc-lag lag ?
  - lag <lag-id> lacp-key <admin-key> system-id <system-id> [remote-lag <remote-lag-    id>]
 system-priority <system-priority> source-bmac-lsb use-lacp-key
  - lag <lag-id> lacp-key <admin-key> system-id <system-id> [remote-lag <remote-lag-    id>]
 system-priority <system-priority> source-bmac-lsb <MAC-Lsb>
  - lag <lag-id> lacp-key <admin-key> system-id <system-id> [remote-lag <remote-lag-    id>]
 system-priority <system-priority>
  - lag <lag-id> [remote-lag <remote-lag-id>]
  - no lag <lag-id>

 <lag-id>              : [1..800]
 <admin-key>           : [1..65535]
 <system-id>           : xx:xx:xx:xx:xx:xx     - xx [00..FF]
 <remote-lag-id>       : [1..800]
 <system-priority>     : [1..65535]
 <MAC-Lsb>             : [1..65535] or xx-xx or xx:xx
```

There must be a different SAP B-MAC per MC-LAG. The use of the LACP key as a default for two least significant octets makes the operations simpler. In this example, the sap-bmac last two octets will come from the lacp-key. The configuration on MTU-4 is as follows:

```
# on MTU-4:
configure
    redundancy
        multi-chassis
            peer 192.0.2.5 create
                mc-lag
                    lag 1 lacp-key 15 system-id 00:00:00:00:00:01
                          system- priority 65535 source-bmac-lsb use-lacp-key
                    no shutdown
                exit
                no shutdown
```

Therefore, the SAP B-MAC will be formed in the following way:

[sap-bmac = 4 first bytes of the source bmac + 2 bytes from source-bmac-lsb]

MAC notification in B-VPLS 100 is enabled on all MTUs, as follows:

```
# on MTU-4, MTU-5, MTU-6:
configure
    service
        vpls "B-VPLS 100"
            mac-notification
                no shutdown
```

The **mac-notification** command activates the described mechanism and has the following parameters:

```
*A:MTU-4# configure service vpls "B-VPLS 100" mac-notification ?
  - mac-notification

 [no] count           - Configure count for MAC-notification messages
 [no] interval        - Configure interval for MAC-notification messages
 [no] renotify        - Configure re-notify interval for MAC-notification messages
 [no] shutdown        - Configure admin state for MAC-notification messages
```

Where:

- **interval** <*value*> controls how often the subsequent MAC notification messages are sent. Default = 100 ms. Required values: 100 ms – 10 sec, in increments of 100 ms.
- **count** <*value*> controls how often the MAC notification messages are sent. Default: 3. Range: 1–10.

The “count” and “interval” parameters can also be configured at the service context. The settings configured at the B-VPLS service context take precedence though.

```
*A:MTU-4# configure service mac-notification ?
  - mac-notification

 [no] count           - Configure count for MAC-notification messages
 [no] interval        - Configure interval for MAC-notification messages
```

Finally, the B-VPLS is instructed to use the SAP B-MAC. The **use-sap-bmac** statement enables the use of the source B-MAC allocated to the multi-homed SAPs (assigned to the MC-LAG) in the related I-VPLS service (could be Epipe service as well). The command will fail if the value of the source B-MAC assigned to the B-VPLS is the hardware (chassis) B-MAC. In other words, the source B-MAC must be a configured one. The **use-sap-bmac** statement is by default off.

```
# on MTU-4:
configure
    service
        vpls "B-VPLS 100"
            pbb
                source-bmac 00:aa:aa:aa:aa:04
                use-sap-bmac
            exit
```

```
*A:MTU-5# configure
    service
        vpls "B-VPLS 100"
            pbb
                source-bmac 00:aa:aa:aa:aa:05
                use-sap-bmac
            exit
```

```
*A:MTU-6# show service id 2 fdb pbb
```

```
==============================================================================
Forwarding Database, i-Vpls Service 2
==============================================================================
MAC               Source-Identifier     B-Svc     b-Vpls MAC        Type/Age
 Transport:Tnl-Id
------------------------------------------------------------------------------
00:08:00:00:00:00 b-sdp:63:100          100       00:aa:aa:aa:00:0f L/0
00:10:00:00:00:00 sap:1/1/1:10          100       N/A               L/0
==============================================================================
```

As soon as the **mac-notification** is enabled, an Ethernet CFM notification message is sent from MTU-4, which is the node where the active MC-LAG link resides. The CFM message will have the source mac "00:aa:aa:aa:00:0f" (4 first bytes of the configured source bmac + 2 bytes from the configured **source-bmac-lsb**, which is 15 in hex) and will be flooded throughout the B-VPLS domain. Should the link between CE-8 and MTU-4 fail, the MC-LAG protocol will activate the redundant link and MTU-5 will immediately issue a CFM message with the shared sourced SAP B-MAC that will be flooded in the B-VPLS domain.

PBB and IGMP snooping

IGMP snooping can be enabled on I-VPLS SAPs and SDPs (it cannot be enabled on B-VPLS). SR OS can keep track of IGMP joins received over individual B-SDPs or B-SAPs, and it starts flooding the multicast group (and only the multicast group) to all B-components (using the group B-MAC for I-SID) as soon as the first IGMP join for that multicast group is received in one of the B-SAP/SDP components.

The first IGMP join message received over the local B-VPLS will add all the B-VPLS SAP/SDP components into the related multicast table associated with the I-VPLS context. When the querier is connected to a remote I-VPLS instance, over the B-VPLS infrastructure, its location is identified by the B-VPLS SDP/SAP on which the query was received and also by the source B-MAC address used in the PBB header for the query message, the B-MAC associated with the B-VPLS instance on the remote PBB PE.

The following configuration on MTU-4 enables IGMP snooping in I-VPLS 1 and adds some static groups on a SAP. The location of the querier is configured by adding the B-MAC where the querier is connected to (in this example, MTU-6) and adding the two B-VPLS spoke-SDPs as mrouter ports (B-VPLS mrouter ports are added in the I-VPLS backbone-vpls context).

The **mac-name** command translates MAC address into strings so that the names can be used instead of typing the entire MAC address every time we need to.

```
# on MTU-4:
configure
    service
       pbb
            source-bmac 00:04:04:04:04:04
            mac-name "MTU-4" 00:04:04:04:04:04
            mac-name "MTU-5" 00:05:05:05:05:05
            mac-name "MTU-6" 00:06:06:06:06:06
        exit
        vpls 1 name "I-VPLS 1" customer 1 i-vpls create
            pbb
                backbone-vpls 100
                    igmp-snooping
                        mrouter-dest "MTU-6"
                    exit
                    sdp 41:100
                        igmp-snooping
                            mrouter-port
                        exit
                    exit
                    sdp 42:100
                        igmp-snooping
                            mrouter-port
                        exit
                    exit
                exit
            exit
            igmp-snooping
                no shutdown
            exit
            sap 1/1/1:7 create
                igmp-snooping
                    static
                        group 228.0.0.1
                            starg
                        exit
                        group 228.0.0.2
                            starg
                        exit
                        group 239.0.0.1
                            source 172.16.99.99
                        exit
                    exit
                exit
            exit
            no shutdown
```

As in regular VPLS instances, mrouter ports are added to all the multicast groups:

```
*A:MTU-4# show service id 1 mfib

===============================================================================
Multicast FIB, Service 1
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
*               *                     b-sdp:41:100                 100      Fwd
                                      b-sdp:42:100                 100      Fwd
*               228.0.0.1             sap:1/1/1:7                  Local    Fwd
                                      b-sdp:41:100                 100      Fwd
                                      b-sdp:42:100                 100      Fwd
*               228.0.0.2             sap:1/1/1:7                  Local    Fwd
                                      b-sdp:41:100                 100      Fwd
                                      b-sdp:42:100                 100      Fwd
172.16.99.99    239.0.0.1             sap:1/1/1:7                  Local    Fwd
                                      b-sdp:41:100                 100      Fwd
                                      b-sdp:42:100                 100      Fwd
-------------------------------------------------------------------------------
Number of entries: 4
===============================================================================
```

When the **show service id x mfib** command is issued in an I-VPLS as in the preceding output, the IGMP (S,G) and (*,G) entries for the I and B components are shown if IGMP snooping is enabled. However, when the same command is launched in a B-VPLS as in the following output, the group B-MAC entries are shown.

```
*A:MTU-4# show service id 100 mfib

===============================================================================
Multicast FIB, Service 100
```

```
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
*               01:1e:83:00:00:01     b-sdp:41:100                 Local    Fwd
*               01:1e:83:00:00:02     b-sdp:41:100                 Local    Fwd
-------------------------------------------------------------------------------
Number of entries: 2
===============================================================================
```

MMRP policies and ISID-based filtering for PBB inter-domain expansion

As described in the MMRP for flooding optimization section, MMRP is used in the backbone VPLS instances to build per I-VPLS flooding trees. Each I-VPLS has an associated group B-MAC in the B-VPLS, which is derived from the ISID, and is advertised by MMRP throughout the whole B-VPLS context, regardless of whether a certain I-VPLS is present in one or all the B-VPLS PEs.

In an inter-domain environment, the same B-VPLS can be defined in different domains and as such MMRP will advertise all the group B-MACs in every domain. The group B-MACs are consuming resources in all the PEs no matter if a particular ISID—and therefore its group B-MAC—is required in one of the domains or not. When MMRP is enabled in a particular PE, data plane and control plane resources are consumed and they must be taken into consideration when designing PBB-VPLS networks:

- Control plane – MRRP processing takes CPU cycles and the number of attributes that can be advertised is not unlimited
- Data plane – each group B-MAC registration takes one MFIB entry (the MFIB is shared between MMRP and IGMP/PIM snooping)

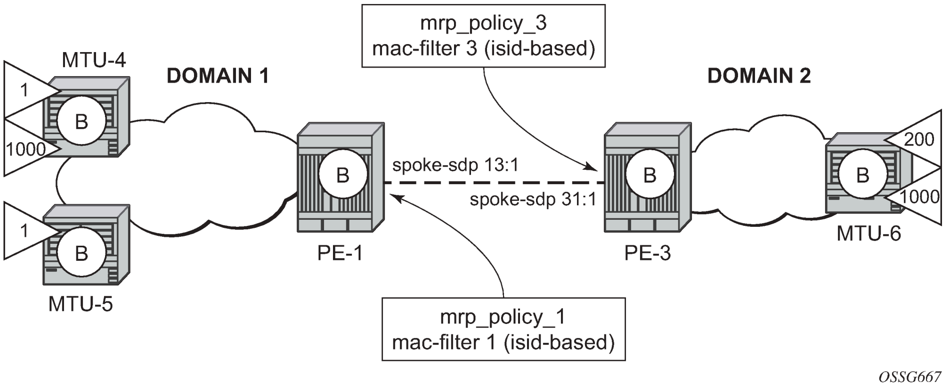
SR OS routers support MMRP policies and ISID-based filters so that control plane and data plane resources can be saved when I-VPLS instances are not defined in all the domains.

Figure 255: Inter-domain B-VPLS and MMRP policies/ISID-based filters example illustrates an example of usage for MMRP policies and ISID-based filters that will be configured in this section. “Domain 1” and “domain 2” will have a range of local ISIDs each and a range of “inter-domain” ISIDs:

- Domain 1 local ISIDs: from 1 to 100
- Domain 2 local ISIDs: from 101 to 200
- Inter-domain ISIDs: from 1000 to 2000

By applying the MMRP policies indicated in Figure 255: Inter-domain B-VPLS and MMRP policies/ISID-based filters example, domain 1 attributes will be prevented from being declared and registered in domain 2 and vice versa, domain 2 attributes from being declared and registered in domain 1. The egress MAC filters will drop any traffic sourced from a local ISID preventing it to be transmitted to the remote domain.

Figure 255: Inter-domain B-VPLS and MMRP policies/ISID-based filters example



MMRP policies

The following shows the MMRP policy configuration on node PE-1. This policy will block any registration/declaration except those for ISIDs 1000-2000. Packets will be compared against the configured matching ISIDs as long as the PBB Etype matches the one configured on the port or SDP.

```
# on PE-1:
configure
    service
        mrp
            mrp-policy "mrp_policy_1" create
                description "allow-inter-domain-isids"
                default-action block
                entry 10 create
                    action allow
                    match
                        isid 1000 to 2000
                    exit
                exit
            exit
        exit
    exit
```

Once the MMRP policy is configured, it must be applied on the corresponding SAP or SDP-binding. An MRP policy can be applied to a B-VPLS SAP, B-VPLS spoke-SDP or B-VPLS mesh-SDP:

```
# on PE-1:
configure
    service
        vpls "B-VPLS 100"
            spoke-sdp 14:100 create
                mrp
                    mrp-policy "mrp_policy_1"
                exit
            exit
            spoke-sdp 15:100 create
                mrp
                    mrp-policy "mrp_policy_1"
                exit
            exit
            no shutdown
```

```
         exit
```

In the same way, mrp_policy_3 will be configured in PE-3.

Some additional considerations about the MMRP policies:

- Different entries within the same MRP policy can have overlapping ISID ranges. The entries will be evaluated in the order of their IDs and the first match will cause the implementation to execute the associated action for that entry and then to exit the MRP policy.
- If no ISID is specified in the match condition then:
  - If the action is "end-station", no entry is added and the action is block.
  - If the action is different from "end-station", every ISID is considered for that action.
- The MRP policy specifies either a forward or a drop action for the group B-MAC attributes associated with the ISIDs specified in the match criteria.

```
      *A:PE-1# configure service mrp mrp-policy "mrp_policy_1" entry 10 action ?
        - action <action>
        - no action

       <action>             : none|block|allow|end-station
```

- There is an additional action called **end-station**. This action specifies that an end-station emulation is present on the SAP/SDP-binding where the policy has been applied. The matching ISIDs will not get declared/registered in the SAP/SDP-binding (just like the **block** action). However, those attributes will get mapped as static MMRP entries on the SAP/SDP-binding, which implicitly get instantiated in the data plane as MFIB entries associated with that SAP/SDP-binding for the related group B-MAC. When the action is "end-station", the default action must be block:

```
*A:PE-3>config>serv>mrp>mrp-policy# default-action allow
MINOR: SVCMGR #5904 Mrp-policy default-action must be block when end-station action
exists
```

- The **end-station** action can be used in the inter-domain gateways when, for instance, we do not want MMRP control plane exchanges between domains. The following output shows how to define the static MMRP entries 1000-2000 in PE-3 without receiving any declaration for any of those attributes or having any of those locally configured.

```
# on PE-3:
configure
    service
        mrp
            mrp-policy "mrp_policy_3" create
                default-action block
                entry 10 create
                    action end-station
                    match
                        isid 1000 to 2000
                    exit
                exit
            exit
        exit
    exit
```

```
*A:PE-3# show service id 100 mfib

===============================================================================
Multicast FIB, Service 100
===============================================================================
Source Address  Group Address         Port Id                       Svc Id   Fwd
                                                                             Blk
-------------------------------------------------------------------------------
*               01:1e:83:00:00:01     b-sdp:36:100                  Local    Fwd
*               01:1e:83:00:03:e8     b-sdp:36:100                  Local    Fwd
*               01:1e:83:00:03:e9     b-sdp:31:4294967294           Local    Fwd
                                      b-sdp:36:100                  Local    Fwd
---snip---
*               01:1e:83:00:07:ce     b-sdp:36:100                  Local    Fwd
*               01:1e:83:00:07:cf     b-sdp:36:100                  Local    Fwd
*               01:1e:83:00:07:d0     b-sdp:36:100                  Local    Fwd
-------------------------------------------------------------------------------
Number of entries: 1002
===============================================================================
```

- The MRP policy can be applied to multiple B-VPLS services as long as the scope of the policy is *template* (the scope can also be *exclusive*).
- Any changes made to the existing policy will be applied immediately to all services where this policy is applied. For this reason, when many changes are required on a MRP policy, Nokia recommends copying the policy to a work-in-progress policy. That work-in-progress policy can be modified until complete and then written over the original MRP policy. You can use the **configure service mrp copy** command to work with the policies in this manner. The **renum** command can also help to change the entries sequence order.

```
*A:PE-3# configure service mrp copy ?
  - copy <src-mrp-policy> to <dst-mrp-policy>

 <src-mrp-policy>     : [32 chars max]
 <dst-mrp-policy>     : [32 chars max]
```

```
*A:PE-3# configure service mrp mrp-policy "mrp_policy_3" renum ?
  - renum <src-entry-id> to <dst-entry-id>

 <src-entry-id>       : [1..65535]
 <dst-entry-id>       : [1..65535]
```

- The **no** form of the **mrp-policy** command deletes the MRP policy. An MRP policy cannot be deleted until it is removed from all the SAPs/SDP-bindings where it is applied.

ISID-based filters

The MMRP policies help to control the exchange of group B-MAC attributes across domains. Based on the registration state of a specific group B-MAC on a SAP/SDP-binding, the BUM traffic for a particular I-VPLS will be allowed or dropped. However, to avoid that any local ISID packet is flooded to the remote B-VPLS domain, all the packets tagged with the local ISIDs at the gateway PEs need to be filtered at the data plane. ISID- based filters will prevent the local ISIDs from sending any packet with unicast B-MAC to the remote domain. This is particularly useful for PBB-Epipe services across domains, where all the frames use unicast B-MACs and MMRP policies cannot help because they only act on group B-MAC packets.

The following CLI output shows how to configure an ISID-based filter that drops all the traffic sourced from the local ISIDs on PE-1 (the default action is drop and it does not show up in the configuration).

```
# on PE-1:
configure
```

```
    filter
        mac-filter 1 name "MAC 1" create
            description "drop_local_isids"
            type isid
            entry 10 create
                match frame-type 802dot3
                    isid 1000 to 2000
                exit
                log 101
                action
                    forward
                exit
            exit
```

Once the filter is configured, it must be applied on a B-VPLS SAP or SDP-binding and always at egress.

```
# on PE-1:
configure
    service
        vpls 100
            spoke-sdp 14:100 create
                egress
                    filter mac 1
                exit
            exit
            spoke-sdp 15:100 create
                egress
                    filter mac 1
                exit
            exit
```

Some additional comments about ISID-based filters:

- The **type isid** statement must be added before introducing any ISID in the match command, otherwise the system will show an error:

  ```
  *A:PE-1>config>filter>mac-filter>entry>match$ isid 1000 to 2000
  MINOR: FILTER #1533 The match criteria entered are not compatible with the Mac
  filter type - On a normal filter no ISID or VID match criteria are allowed
  ```

  ```
  *A:PE-1>config>filter>mac-filter$ type isid
  MINOR: FILTER #1561 Cannot change filter type when filter contains entries
  ```

- Once the operator sets the “type isid”, the filter cannot be applied at ingress. Only egress ISID-based filters are allowed:

  ```
  *A:PE-1>config>service>vpls>spoke-sdp# ingress filter mac 1
  MINOR: SVCMGR #2050 Can not apply filter of type 'isid' on ingress
  ```

- Like any filter or MMRP policy, the filter can be applied to multiple B-VPLS services as long as the scope of the policy is “template” (the scope can also be “exclusive”).
- The following command shows the filter configuration and packets that have matched the filter (field “Egr. Matches”):

  ```
  *A:PE-1# show filter mac 1

  ===============================================================================
  Mac Filter
  ===============================================================================
  ```

```
Filter Id           : 1                              Applied        : Yes
Scope               : Template                       Def. Action    : Drop
Entries             : 1                              Type           : isid
Description         : drop_local_isids
Filter Name         : MAC 1
-------------------------------------------------------------------------------
Filter Match Criteria : Mac
-------------------------------------------------------------------------------
Entry               : 10                             FrameType      : Ethernet
Description         : (Not Specified)
Log Id              : 101
ISID                : 1000..2000
Primary Action      : Forward
Ing. Matches        : 0 pkts
Egr. Matches        : 5 pkts (580 bytes)

===============================================================================
```

- Like any other filter, the matching packets can be logged. An example follows (the Ethertype is 0x88e7, which is the default standard Ethertype for PBB):

```
*A:PE-1# show filter log 101

===============================================================================
Filter Log
===============================================================================
Admin state : Enabled
Description : Default filter log
Destination : Memory
Wrap        : Enabled
-------------------------------------------------------------------------------
Maximum entries configured : 1000
Number of entries logged   : 5
-------------------------------------------------------------------------------
2021/01/12 17:13:40  Mac Filter: 1:10  Desc:
Interface: int-PE-1-MTU-4  Direction: Egress  Action: Forward
VID match: 0
Src MAC: 00-06-06-06-06-06  Dst MAC: 00-aa-aa-aa-00-0f  EtherType: 88e7
Hex: 00 00 03 e9 00 08 00 00 00 00 00 10 00 00 00 00
     08 00 45 00 00 54 27 97 00 00 40 01 22 ee ac 10
     6c 02 ac 10 6c 01 00 00 f1 ff 00 fb 80 01 5f fd*

2021/01/12 17:13:41  Mac Filter: 1:10  Desc:
Interface: int-PE-1-MTU-4  Direction: Egress  Action: Forward
VID match: 0
Src MAC: 00-06-06-06-06-06  Dst MAC: 00-aa-aa-aa-00-0f  EtherType: 88e7
Hex: 00 00 03 e9 00 08 00 00 00 00 00 10 00 00 00 00
     08 00 45 00 00 54 27 99 00 00 40 01 22 ec ac 10
     6c 02 ac 10 6c 01 00 00 41 05 00 fb 80 02 5f fd*

---snip---
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

B-VPLS and I-VPLS show and debug commands

For the following output, the MRP policies and ISID-based MAC filters have been removed from the spoke-SDPs on PE-1 and PE-3. The following commands can help to check the B-VPLS and I-VPLS configuration and their related parameters. The first is for the B-VPLS on MTU-4:

```
*A:MTU-4# show service id 100 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 100                 Vpn Id            : 0
Service Type      : b-VPLS
MACSec enabled    : no
Name              : B-VPLS 100
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 01/12/2021 16:08:29
Last Mgmt Change  : 01/12/2021 17:03:38
Etree Mode        : Disabled
Admin State       : Up                  Oper State        : Up
MTU               : 2000
SAP Count         : 0                   SDP Bind Count    : 2
Snd Flush on Fail : Disabled            Host Conn Verify  : Disabled
SHCV pol IPv4     : None
Propagate MacFlush: Disabled            Per Svc Hashing   : Disabled
Allow IP Intf Bind: Disabled
Fwd-IPv4-Mcast-To*: Disabled            Fwd-IPv6-Mcast-To*: Disabled
Mcast IPv6 scope  : mac-based
Temp Flood Time   : Disabled            Temp Flood        : Inactive
Temp Flood Chg Cnt: 0
SPI load-balance  : Disabled
TEID load-balance : Disabled
Src Tep IP        : N/A
Vxlan ECMP        : Disabled
MPLS ECMP         : Disabled
VSD Domain        : <none>
Oper Backbone Src : 00:aa:aa:aa:aa:04
Use SAP B-MAC     : Enabled
i-Vpls Count      : 2
Epipe Count       : 0
Use ESI B-MAC     : Disabled

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sdp:41:100 S(192.0.2.1)                  Spok         8000    8000    Up   Up
sdp:42:100 S(192.0.2.2)                  Spok         8000    8000    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

For the I-VPLS on MTU-4:

```
*A:MTU-4# show service id 1 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : i-VPLS
MACSec enabled    : no
Name              : I-VPLS 1
Description       : (Not Specified)
Customer Id       : 1                  Creation Origin   : manual
Last Status Change: 01/12/2021 16:17:52
Last Mgmt Change  : 01/12/2021 16:17:52
Etree Mode        : Disabled
Admin State       : Up                 Oper State        : Up
MTU               : 1514
SAP Count         : 1                  SDP Bind Count    : 0
Snd Flush on Fail : Disabled           Host Conn Verify  : Disabled
SHCV pol IPv4     : None
Propagate MacFlush: Disabled           Per Svc Hashing   : Disabled
Allow IP Intf Bind: Disabled
Fwd-IPv4-Mcast-To*: Disabled           Fwd-IPv6-Mcast-To*: Disabled
Mcast IPv6 scope  : mac-based
Temp Flood Time   : Disabled           Temp Flood        : Inactive
Temp Flood Chg Cnt: 0
SPI load-balance  : Disabled
TEID load-balance : Disabled
Src Tep IP        : N/A
Vxlan ECMP        : Disabled
MPLS ECMP         : Disabled
VSD Domain        : <none>
b-Vpls Id         : 100                Oper ISID         : 1
b-Vpls Status     : Up
Snd Flush in bVpls: None
Flsh On bVpls Fail: Disabled           Prop Flsh fr bVpls: Disabled
Force QTag Fwd    : Disabled
SendBvplsEvpnFlush: Disabled

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/1:7                              q-tag        1518    1518    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

The following command shows all the I-VPLS instances multiplexed into a particular B-VPLS.

```
*A:MTU-4# show service id 100 i-vpls

===============================================================================
Related i-Vpls services for b-Vpls service 100
===============================================================================
i-Vpls SvcId        Oper ISID           Admin               Oper
-------------------------------------------------------------------------------
1                   1                   Up                  Up
2                   2                   Up                  Up
-------------------------------------------------------------------------------
Number of Entries : 2
-------------------------------------------------------------------------------
===============================================================================
```

Some useful commands to check the I and B VPLS FIBs correlating C-MACs and B-MACs:

```
*A:MTU-4# show service id 1 fdb pbb

==============================================================================
Forwarding Database, i-Vpls Service 1
```

```
==============================================================================
MAC               Source-Identifier     B-Svc     b-Vpls MAC        Type/Age
 Transport:Tnl-Id
------------------------------------------------------------------------------
00:07:00:00:00:00 sap:1/1/1:7           100       N/A               L/0
00:09:00:00:00:00 b-sdp:41:100          100       00:06:06:06:06:06 L/0
==============================================================================
```

```
*A:MTU-4# show service id 100 fdb pbb

=======================================================================
Forwarding Database, b-Vpls Service 100
=======================================================================
MAC               Source-Identifier     iVplsMACs  Epipes     Type/Age
 Transport:Tnl-Id
-----------------------------------------------------------------------
00:06:06:06:06:06 sdp:41:100            2          0          L/0
04:0f:ff:00:00:00 sdp:41:100            0          0          L/0
=======================================================================
```

If **mac-name** is used in the configuration, the following commands can show the translations:

```
*A:MTU-4# show service pbb mac-name

======================================================================
MAC Name Table
======================================================================
MAC-Name                           MAC-Address
----------------------------------------------------------------------
MTU-4                              00:04:04:04:04:04
MTU-5                              00:05:05:05:05:05
MTU-6                              00:06:06:06:06:06
======================================================================
```

```
*A:MTU-4# show service pbb mac-name "MTU-6" detail

======================================================================
Services Using MAC name='MTU-6' addr='00:06:06:06:06:06'
======================================================================
Svc-Id                             ISID
----------------------------------------------------------------------
1                                  N/A
----------------------------------------------------------------------
Number of services: 1
======================================================================
```

The following command shows the base MAC notification parameters as well as the source B-MAC configured at the service PBB level. Those values are overridden by any potential MAC notification or source B-MAC values configured under the B-VPLS service context.

```
*A:MTU-4# show service pbb base

======================================================================
PBB MAC Information
======================================================================
MAC-Notif Count                   : 3
MAC-Notif Interval                : 1
Source BMAC                       : 00:04:04:04:04:04
Leaf Source BMAC                  : Default
======================================================================
```

If MAC notification is used in a particular B-VPLS, the configured least significant bits for the SAP B-MAC on a particular MC-LAG can be shown by using the detailed view of the **show lag** command:

```
*A:MTU-4# show lag 1 detail

===============================================================================
LAG Details
===============================================================================
Description        : N/A
-------------------------------------------------------------------------------
Details
-------------------------------------------------------------------------------
Lag-id             : 1                     Mode                : access
Adm                : up                    Opr                 : up

---snip---

MC Peer Address     : 192.0.2.5             MC Peer Lag-id       : 1
MC System Id        : 00:00:00:00:00:01     MC System Priority   : 65535
MC Admin Key        : 15                    MC Active/Standby    : active
MC Lacp ID in use   : true                  MC extended timeout  : false
MC Selection Logic  : local master decided
MC Config Mismatch  : no mismatch
Source BMAC LSB     : use-lacp-key          Oper Src BMAC LSB    : 00:0f

---snip---
===============================================================================
```

The following debug commands allow the operator to check the LDP label mapping, label withdrawal, messages and also the MAC-flush messages for regular VPLS, for I-VPLS and B-VPLS including the PBB extensions and TLVs.

```
*A:MTU-4# show debug
debug
    router "Base"
        ldp
            peer 192.0.2.1
                event
                exit
                packet
                    init detail
                    label detail
                exit
            exit
            peer 192.0.2.2
                event
                exit
                packet
                    init detail
                    label detail
                exit
            exit
        exit
    exit
exit
```

The following debug commands can help the operator to troubleshoot MMRP.

```
*A:MTU-4# debug service id 100 mrp ?
  - mrp
  - no mrp
```

```
     all-events      - Enable/disable MRP debugging for all events
[no] applicant-sm    - Enable/disable MRP debugging for applicant state machine
                       changes
[no] leave-all-sm    - Enable/disable MRP debugging for leave all state machine
                       changes
[no] mmrp-mac        - Enable/disable MRP debugging for a particular MAC address
[no] mrpdu           - Enable/disable MRP debugging for Rx/Tx MRP PDUs
[no] mvrp-vlan       - Enable/disable debugging for a particular vlan
[no] periodic-sm     - Enable/disable MRP debugging for periodic state machine
                       changes
[no] registrant-sm   - Enable/disable MRP debugging for registrant state machine
                       changes
[no] sap             - Enable/disable MRP debugging for a particular SAP
[no] sdp             - Enable/disable MRP debugging for a particular SDP
```

Conclusion

PBB-VPLS allows the service providers to scale VPLS services by multiplexing customer I-VPLS instances into one or more B-VPLS instances. This multiplexing dramatically reduces the number of services, pseudowires, and MAC addresses in the core and therefore allows the service provider to scale Layer 2 multi-point networks and provide services across international backbones.

The example used in this chapter shows the configuration of the customer and backbone VPLS instances as well as all the related features which are required for this environment. Show and debug commands have also been suggested so that the operator can verify and troubleshoot the service.

PIM Snooping for IPv4 in EVPN-MPLS Services

This chapter provides information about PIM snooping for IPv4 in EVPN-MPLS services.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

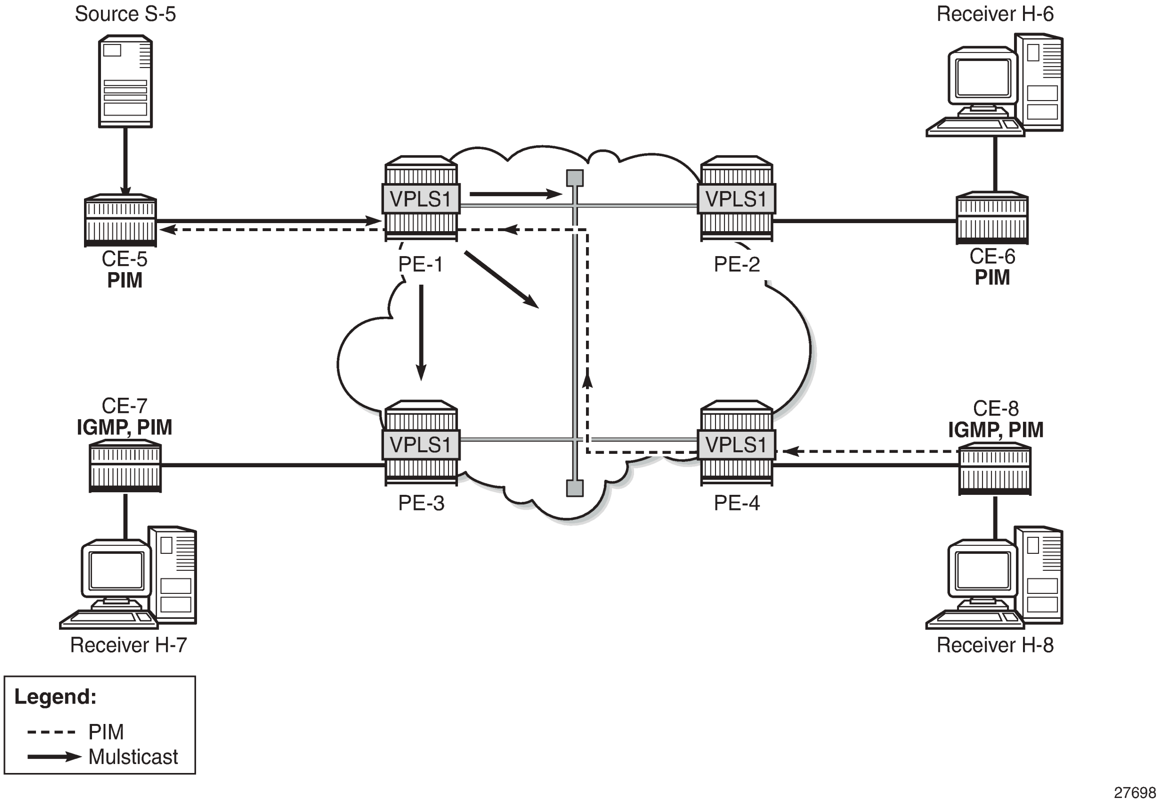
Applicability

This chapter was initially written based on SR OS Release 15.0.R5, but the CLI in the current edition corresponds to SR OS Release 23.7.R1. PIM snooping for IPv4 is supported in EVPN-MPLS services in SR OS Release 15.0.R1, and later. PIM snooping in single-active multi-homing mode without ESI label is supported in SR OS Release 15.0.R1, and later, whereas PIM snooping in single-active multi-homing mode with ESI label is supported in SR OS Release 15.0.R4, and later. PIM snooping in all-active multi-homing mode is supported in SR OS Release 15.0.R4, and later. Data-driven PIM state synchronization is supported in SR OS Release 15.0.R4, and later.

Overview

Figure 256: Multicast in VPLS without PIM Snooping shows the example topology with four CEs that have IGMP and PIM enabled (L3) and four PEs configured with VPLS 1 (L2). Source-specific multicast is used in this example. The following description applies to all VPLSs, with or without EVPN.

Figure 256: Multicast in VPLS without PIM Snooping



The VPLS emulates a LAN interconnecting sites with L3-capable devices that use PIM to join or leave multicast groups. When receiver H-8 sends an IGMP report message to join a multicast group, CE-8 sends a PIM join message to CE-5. The PEs forward the PIM message without learning any PIM-related information, such as which CE sent the PIM join and for which multicast group.

The source S-5 is sending the multicast stream to CE-5. When CE-5 receives a PIM join message for this multicast group from CE-8, it forwards the multicast stream to CE-8. By default, all PEs flood the multicast stream on all their connections in the VPLS domain, regardless of whether a PIM join was received from that connection. L2 flooding is not aware of the PIM join/prune messages from the L3 edge routers, resulting in an inefficient use of network resources. To avoid this L2 flooding, PIM snooping can be enabled in the VPLS by the following command:

```
*A:PE-1# configure service vpls 1 pim-snooping ?
  - no pim-snooping
  - pim-snooping

 [no] group-policy    - Enable/Disable PIM snooping group policy
 [no] hold-time       - Configure PIM snooping hold time
 [no] ipv4-multicast* - Administratively disable/enable PIM operation for IPv4
 [no] ipv6-multicast* - Administratively disable/enable PIM operation for IPv6
      mode            - Administratively enable/disable PIM Snooping mode
```

The default mode is proxy, but PIM snooping can also use snooping mode, depending on the information in the received PIM hello messages. In snooping mode, the PE does not modify the PIM messages; in proxy mode, the PE terminates incoming PIM messages and generates its own PIM messages.

PIM snooping is used for router multicast registration, whereas IGMP snooping is used for host/client multicast registration. IGMP snooping in EVPN-MPLS services is described in chapter P2MP mLDP Tunnels for BUM Traffic in EVPN-MPLS Services. Optionally, PIM snooping and IGMP snooping can be enabled simultaneously.

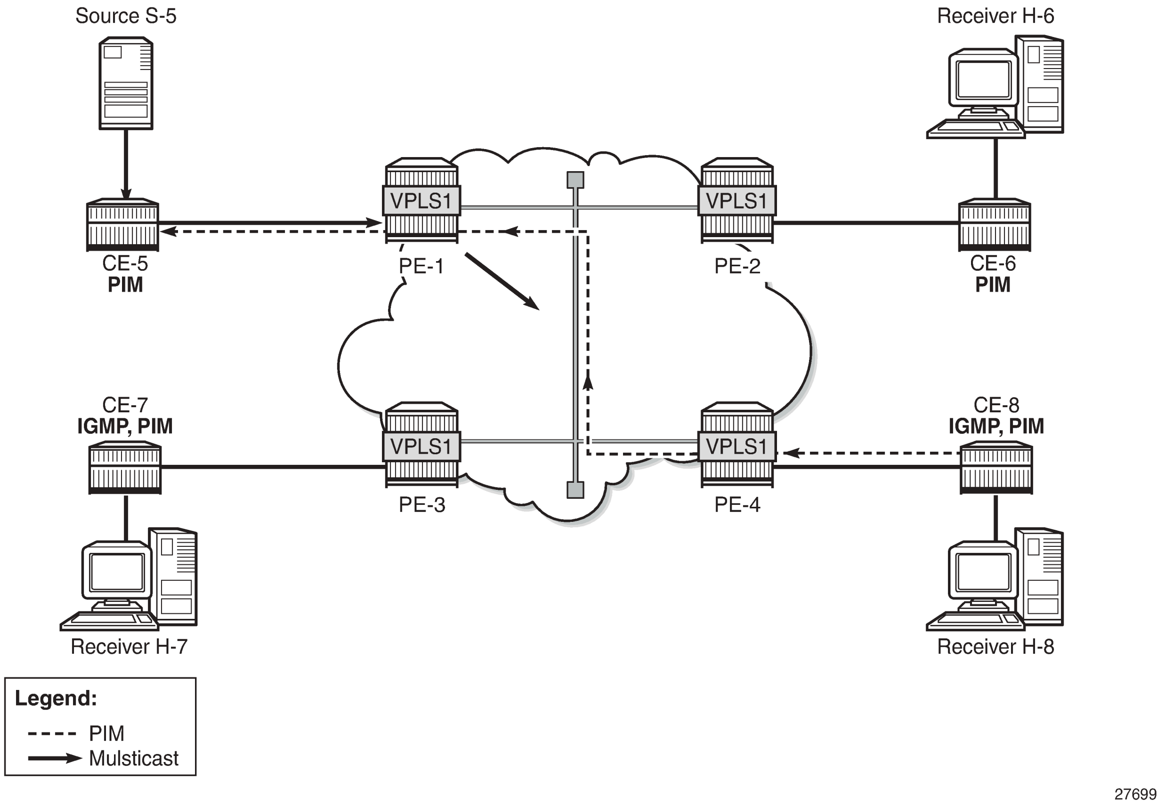
With PIM enabled, the CEs send PIM hello messages to the well-known multicast address for PIM, 224.0.0.13. PIM hello messages are used to form PIM neighbors and can be used to form the Forwarding Database (FDB). With PIM snooping enabled in the VPLS in the PEs, the PEs snoop PIM messages. The PEs only forward multicast traffic downstream when required, as determined from the received PIM messages. This provides a more efficient use of network resources.

PIM snooping states in a PE are maintained per VPLS instance. When PIM snooping is enabled, IP multicast traffic to a multicast group that is not learned via snooping is dropped by default, unless it is received from a directly connected source.

PIM Snooping in Snooping Mode

Figure 257: Multicast in VPLS with PIM Snooping in Snooping Mode shows that the multicast stream is not flooded in PE-1 when PIM snooping is enabled and operating in snooping mode.

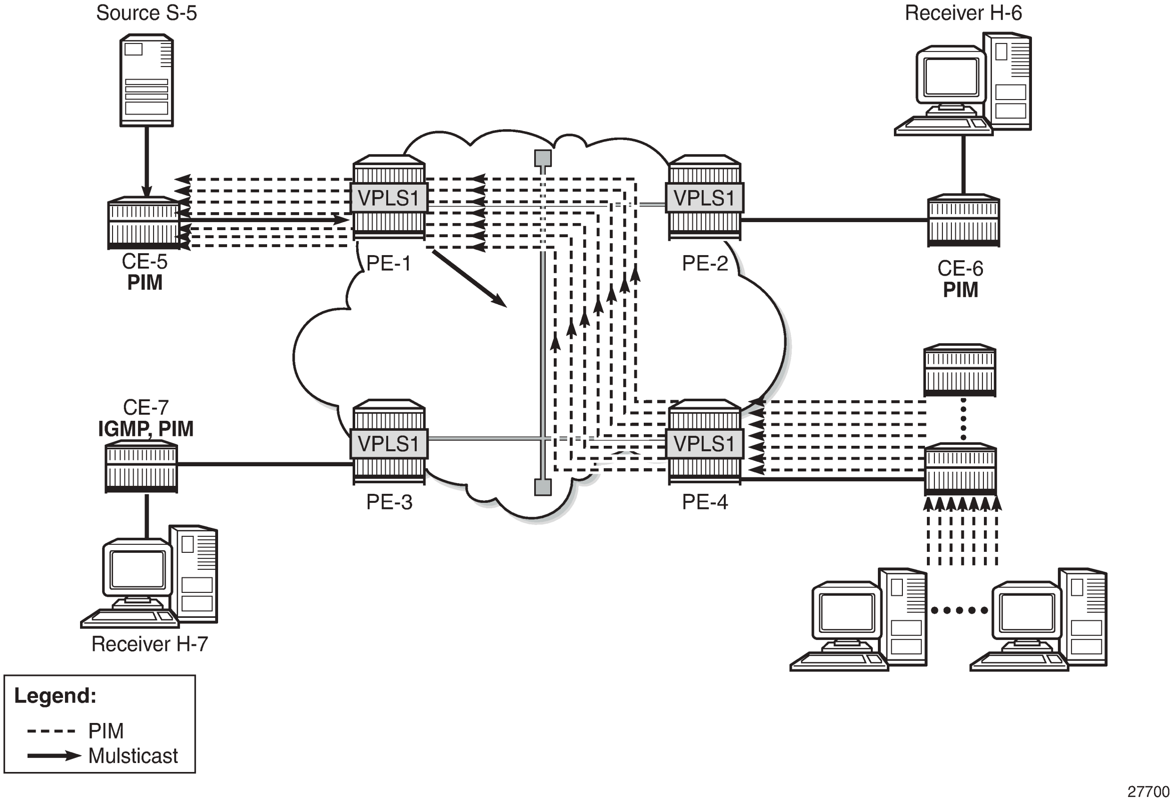
Figure 257: Multicast in VPLS with PIM Snooping in Snooping Mode



When H-8 sends an IGMP report message to join the multicast stream from source S-5 to CE-8, CE-8 sends a PIM join message to CE-5. PE-4 snoops the PIM join message and builds the FDB. PE-4 forwards the PIM join message to PE-1 by matching the upstream neighbor address in the join with the neighbor database. PE-1 snoops the PIM join message, builds its Multicast Forwarding Information Base (MFIB), and performs a similar lookup in its FDB. PE-1 forwards the PIM join to CE-5. The Source Path Tree (SPT) between receiver CE-8 and sender CE-5 is now built and CE-5 forwards multicast data frames to CE-8. PE-1 does not flood multicast frames, but forwards them to CE-8 only, based on the MFIB.

Figure 258: Multicast in VPLS with PIM Snooping in Snoop Mode – Multiple CEs shows how the number of PIM messages in the control plane increases when multiple client CEs are connected to PE-4.

Figure 258: Multicast in VPLS with PIM Snooping in Snoop Mode – Multiple CEs



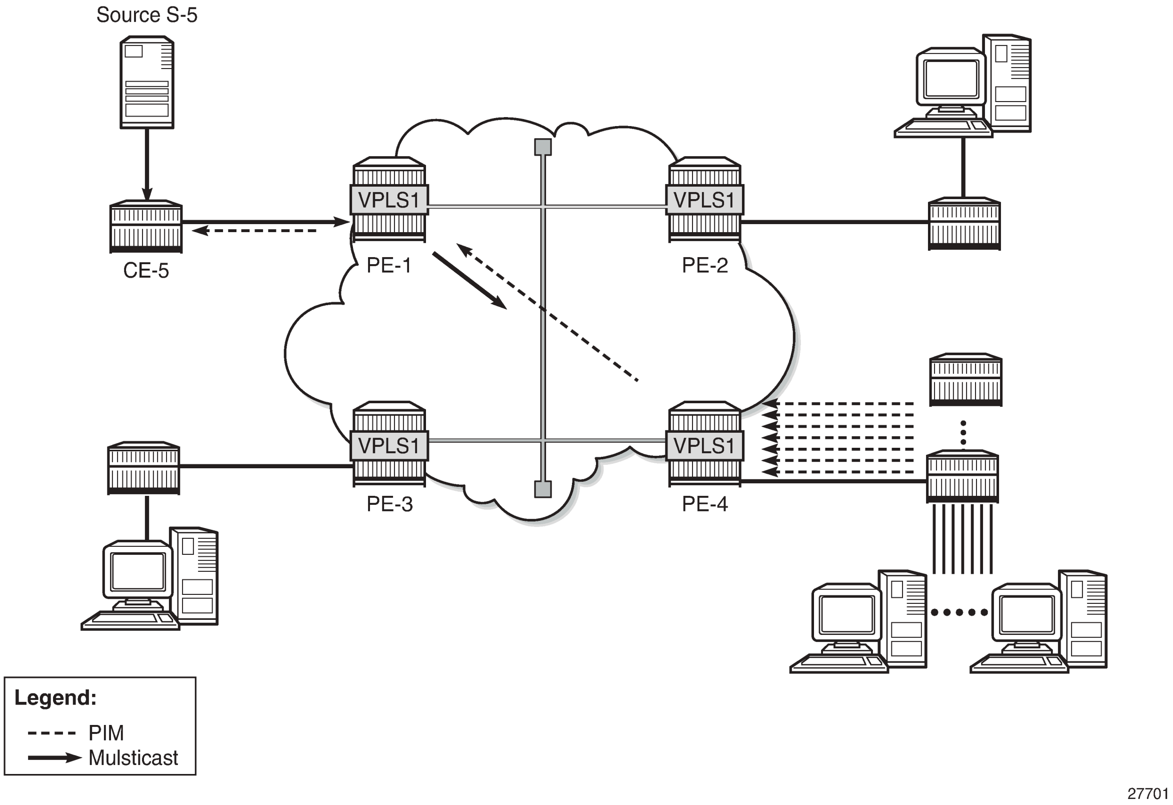
PIM Snooping in Proxy Mode

When H-8 sends an IGMP report message to join a multicast stream, CE-8 again sends a PIM join message to CE-5. PE-4 terminates the incoming PIM join message and generates its own PIM join message using CE-5 as the source address, learned from the PIM hello messages. PE-4 builds its MFIB and sends a new PIM join message to S-5. PE-1 terminates the incoming PIM join message and builds its MFIB. PE-1 generates its own PIM join message using CE-5 as the source address. PE-1 forwards the PIM join to CE-5. The SPT between CE-8 and CE-5 is now built and the multicast stream flows from source S-5 to receiver H-8. No multicast traffic is sent to CE-6 and CE-7, because they do not have receivers attached that joined the multicast stream.

The default mode for PIM snooping is proxy mode.

Figure 259: Multicast in VPLS with PIM Snooping in Proxy Mode - Multiple CEs shows that the number of PIM messages in the control plane does not increase when multiple client CEs are connected to PE-4, compared to snooping mode.

Figure 259: Multicast in VPLS with PIM Snooping in Proxy Mode - Multiple CEs



PIM snooping in proxy mode can be configured with a delay to avoid existing traffic interruption. PIM snooping in proxy mode does not program the MFIB until a hold timer has expired. This hold time is useful in the following cases:

- PIM snooping being enabled on the VPLS
- PIM snooping states being manually cleared by an operator

When the hold timer is started, but not expired yet, multicast traffic is flooded in the VPLS as if PIM snooping was not enabled. VPLS flooding ensures flow delivery during the hold time.

PIM Snooping in VPLS with EVPN-MPLS

PIM snooping in an EVPN-MPLS service supports the following:

- Regular PIM snooping on SAPs/SDP-bindings
  - PIM messages received on EVPN-MPLS endpoints are forwarded to SAPs/SDP-bindings.
  - IP multicast traffic received on an EVPN-MPLS binding is forwarded to SAPs/SDP-bindings from which a PIM join was received, or to ports configured as mrouter ports.

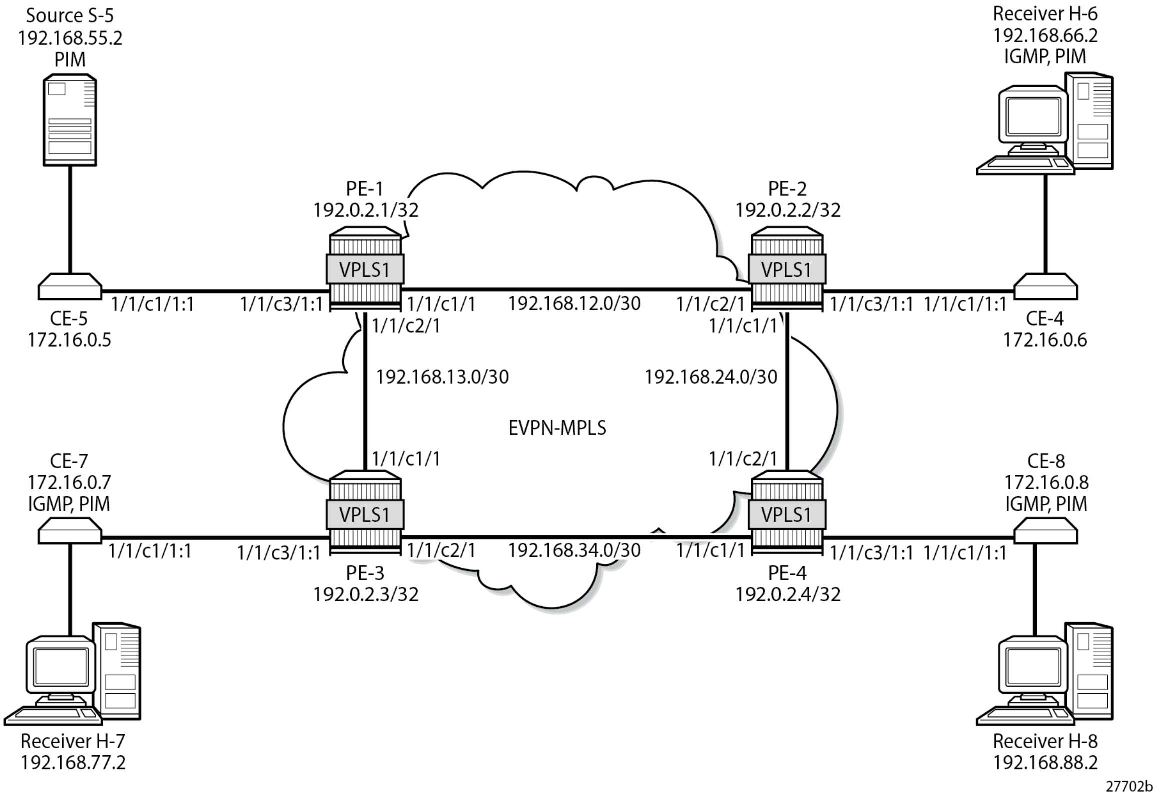
- The EVPN-MPLS endpoints are treated as a single PIM interface:
  - IP multicast traffic and PIM messages received on an EVPN-MPLS endpoint are not forwarded to other EVPN-MPLS endpoints (split-horizon).
  - Hello and join/prune messages from SAPs/SDP-bindings are forwarded to all EVPN-MPLS destinations.
  - When a hello message is received from one PIM neighbor on an EVPN-MPLS destination, the single interface representing all EVPN-MPLS destinations has that neighbor.
    - Individual destinations appear in the MFIB, but the information for each EVPN-MPLS destination entry is identical.
- If a Point-to-Multipoint (P2MP) mLDP provider tunnel is configured:
  - If the PE is the root node of a P2MP LSP that is up, PIM messages and IP multicast traffic are only forwarded over the P2MP Label Switched Path (LSP) instead of being sent to the EVPN-MPLS endpoints. Therefore, the P2MP leaves must match the EVPN-MPLS endpoints, in this case, PE-2, PE-3, and PE-4.
  - If the PE is a leaf node of a P2MP LSP, it sends PIM messages and IP multicast traffic over its EVPN-MPLS endpoints.
  - The PEs can expect to receive IP multicast traffic and PIM messages from leaf nodes over their EVPN-MPLS endpoints, or over the P2MP LSPs for traffic from root nodes.
- PIM snooping is supported in inter-AS model B and inter-AS model C, as for IGMP snooping.
- All-active and single-active EVPN multi-homing are supported.
- Multi-chassis Synchronization (MCS) of PIM snooping state is supported on SAPs and spoke-SDPs in dual-homing.
  - The active (Designated Forwarder (DF)) PE sends the PIM states to the backup non-DF (NDF) PE.
  - In case of failure, the backup PE has the PIM states already, and the multicast traffic path can be re-established fast without any need to wait for PIM states to be snooped.
  - A sync-tag is configured on the ports or SDPs that need to be synchronized on both PEs.
  - MCS PIM snooping is restricted to two peers, even though MCS supports more peers for other types of information. An error is raised when attempting to configure a sync-tag on the same port or SDP to more than one peer.
- PIM snooping is supported for both IPv4 and IPv6 multicast. PIM snooping for IPv6 uses MAC-based forwarding by default, and can be configured to use (S,G)-based forwarding.
- PIM snooping is transparent to the underlying tunnel. PIM snooping works with RSVP, LDP, SR-ISIS, SR-OSPF, SR-TE, BGP, and MPLSoUDP.
- PIM snooping is not supported with routed VPLS with EVPN-MPLS, and its configuration is blocked.

Configuration

Figure 260: Example Topology shows the example topology. Source S-5 sends multicast streams to CE-5, which forwards those only after a PIM join message has been received. An mLDP P2MP LSP is used to distribute the multicast from the root node PE-1 to the other PEs. All CEs have PIM enabled and the receiving CEs (CE-6, CE-7, and CE-8) have IGMP configured on the interface toward the receivers (H-6,

H-7, and H-8). EVPN-MPLS VPLS 1 is configured on the PEs. Initially, PIM snooping is disabled in the VPLS. Receiver H-8 joins multicast group 232.1.1.1 from source S-5.

Figure 260: Example Topology



The initial configuration includes the following:

- Cards, MDAs
- Ports
  - Ports between PEs are network ports with null encapsulation
  - Ports between CEs and PEs are hybrid ports with dot1q encapsulation
- IS-IS as IGP between the PEs (alternatively, OSPF can be used)
- LDP between the PEs
- BGP with address family EVPN between the PEs. PE-2 is the route reflector (RR). The BGP configuration on RR PE-2 is as follows:

```
On PE-2:
configure
    router "Base"
        autonomous-system 64496
        bgp
            rapid-withdrawal
            rapid-update evpn
            group INTERNAL
                family evpn
                type internal
                cluster 192.0.2.2
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
            exit
        exit
    exit all
```

EVPN-MPLS VPLS without PIM Snooping

VPLS 1 is configured with EVPN-MPLS in the PEs. By default, PIM snooping is disabled. PE-1 is configured as **root-and-leaf** node for the P2MP mLDP multicast tree, while the other three PEs have the default **no root-and-leaf** configured, so they are leaf-only nodes. The configuration of VPLS 1 on PE-1 is as follows:

```
On PE-1:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    ingress-replication-bum-label
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            provider-tunnel
                inclusive
                    owner bgp-evpn-mpls
                    root-and-leaf
                    mldp
                    no shutdown
                exit
            exit
            sap 1/1/c3/1:1 create
            exit
            no shutdown
        exit all
```

A P2MP mLDP multicast tree is created from root node PE-1 to the leaf nodes. On the root node PE-1, an SDP of type **VplsPmsi** is auto-created:

```
*A:PE-1# show service id 1 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : VPLS
---snip---
```

```
-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/1/c3/1:1                           q-tag        8936    8936    Up   Up
sdp:32767:4294967294 SB(not applicable)  VplsPmsi     9782    9782    Up   Up
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

The following inclusive provider tunnel is created on root node PE-1:

```
*A:PE-1# show service id 1 provider-tunnel

===============================================================================
Service Provider Tunnel Information
===============================================================================
Type               : inclusive          Root and Leaf      : enabled
Admin State        : enabled            Data Delay Intvl   : 15 secs
PMSI Type          : ldp                LSP Template       :
Remain Delay Intvl : 0 secs             LSP Name used      : 8193
PMSI Owner         : bgpEvpnMpls
Oper State         : up                 Root Bind Id       : 32767
-------------------------------------------------------------------------------
Type               : selective          Wildcard SPMSI     : disabled
Admin State        : disabled           Data Delay Intvl   : 3 secs
PMSI Type          : none               Max P2MP SPMSI     : 10
PMSI Owner         : none
===============================================================================
```

When a P2MP mLDP provider tunnel is configured, the root node forwards PIM messages and IP multicast traffic over the provider tunnel instead of over the EVPN-MPLS endpoints. However, the leaf nodes of a P2MP mLDP provider tunnel send PIM messages and IP multicast traffic over the EVPN-MPLS endpoints.

The following P2MP mLDP bindings are active on root node PE-1: one toward PE-2 via port 1/1/c1/1 and one toward PE-3 via port 1/1/c2/1.

```
*A:PE-1# show router ldp bindings active p2mp opaque-type generic ipv4

===============================================================================
LDP Bindings (IPv4 LSR ID 192.0.2.1)
             (IPv6 LSR ID ::)
===============================================================================
Label Status:
        U - Label In Use, N - Label Not In Use, W - Label Withdrawn
        WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
        e - Label ELC
FEC Flags:
        LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
        BA - ASBR Backup FEC
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                     Interface
RootAddr                                    Op
IngLbl                                      EgrLbl
EgrNH                                       EgrIf/LspId
-------------------------------------------------------------------------------
8193                                        73728
192.0.2.1                                   Push
  --                                        524281
192.168.12.2                                1/1/c1/1
```

```
8193                                        73728
192.0.2.1                                   Push
  --                                        524281
192.168.13.2                                1/1/c2/1

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 2
===============================================================================
```

The following P2MP mLDP bindings are active on PE-2. PE-2 is a leaf node (pop operation) and a transit node for traffic toward PE-4 (swap operation):

```
*A:PE-2# show router ldp bindings active p2mp opaque-type generic ipv4

===============================================================================
---snip---
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                     Interface
RootAddr                                    Op
IngLbl                                      EgrLbl
EgrNH                                       EgrIf/LspId
-------------------------------------------------------------------------------
8193                                        73728
192.0.2.1                                   Pop
524281                                        --
  --                                          --

8193                                        73728
192.0.2.1                                   Swap
524281                                      524281
192.168.24.2                                1/1/c1/1

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 2
===============================================================================
```

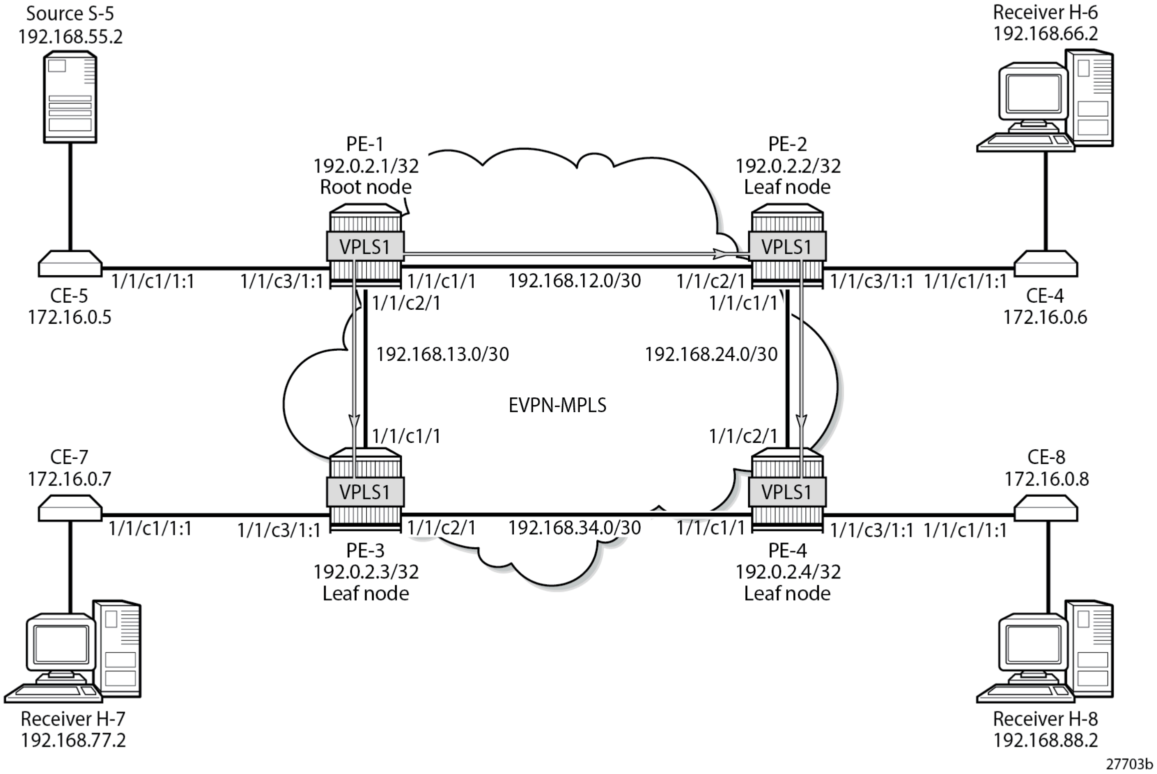
PE-3 and PE-4 are leaf nodes, so there is a pop operation. The active P2MP LDP binding on PE-4 is the following. A similar P2MP LDP binding occurs on PE-3.

```
*A:PE-4# show router ldp bindings active p2mp opaque-type generic ipv4

===============================================================================
---snip---
===============================================================================
LDP Generic IPv4 P2MP Bindings (Active)
===============================================================================
P2MP-Id                                     Interface
RootAddr                                    Op
IngLbl                                      EgrLbl
EgrNH                                       EgrIf/LspId
-------------------------------------------------------------------------------
8193                                        73728
192.0.2.1                                   Pop
524281                                        --
  --                                          --

-------------------------------------------------------------------------------
No. of Generic IPv4 P2MP Active Bindings: 1
```

```
===============================================================================
```

Figure 261: P2MP mLDP Multicast Tree shows the mLDP multicast tree. Multicast traffic from source S-5 uses the mLDP multicast tree from PE-1 to both PE-2 and PE-3. PE-2 is a transit node for multicast traffic to PE-4, and also a leaf node. PE-3 and PE-4 are leaf nodes.

Figure 261: P2MP mLDP Multicast Tree



CE-6, CE-7, and CE-8 have IGMP enabled on the interface toward the receiver and PIM enabled on all interfaces. The configuration on CE-8 is as follows:

```
On CE-8:
configure
    router "Base"
        interface "int-CE-8-H-8"
            address 192.168.88.1/24
            port 1/1/c2/1
        exit
        interface "int-CE-8-PE-4"
            address 172.16.0.8/16
            port 1/1/c1/1:1
        exit
        interface "system"
            address 192.0.2.8/32
        exit
        static-route-entry 192.168.55.0/30
            next-hop 172.16.0.5
                no shutdown
            exit
        exit
        igmp
            interface "int-CE-8-H-8"
            exit
        exit
        pim
            apply-to all
        exit
    exit all
```

The static route is required on the receiving CEs for the PIM join/prune messages to reach the multicast source S-5 with IP address 192.168.55.2; only IP subnet 172.16.0.0/16 can be reached via the VPLS.

CE-5 has PIM enabled and static routes configured to reach the receiving hosts, as follows:

```
On CE-5:
configure
    router "Base"
        interface "int-CE-5-PE-1"
            address 172.16.0.5/16
            port 1/1/c1/1:1
        exit
        interface "int-CE-5-S-5"
            address 192.168.55.1/30
            port 1/1/c3/1
        exit
        interface "system"
            address 192.0.2.5/32
        exit
        static-route-entry 192.168.66.0/24
            next-hop 172.16.0.6
                no shutdown
            exit
        exit
        static-route-entry 192.168.77.0/24
            next-hop 172.16.0.7
                no shutdown
            exit
        exit
        static-route-entry 192.168.88.0/24
            next-hop 172.16.0.8
                no shutdown
            exit
        exit
        pim
            apply-to all
        exit
    exit all
```

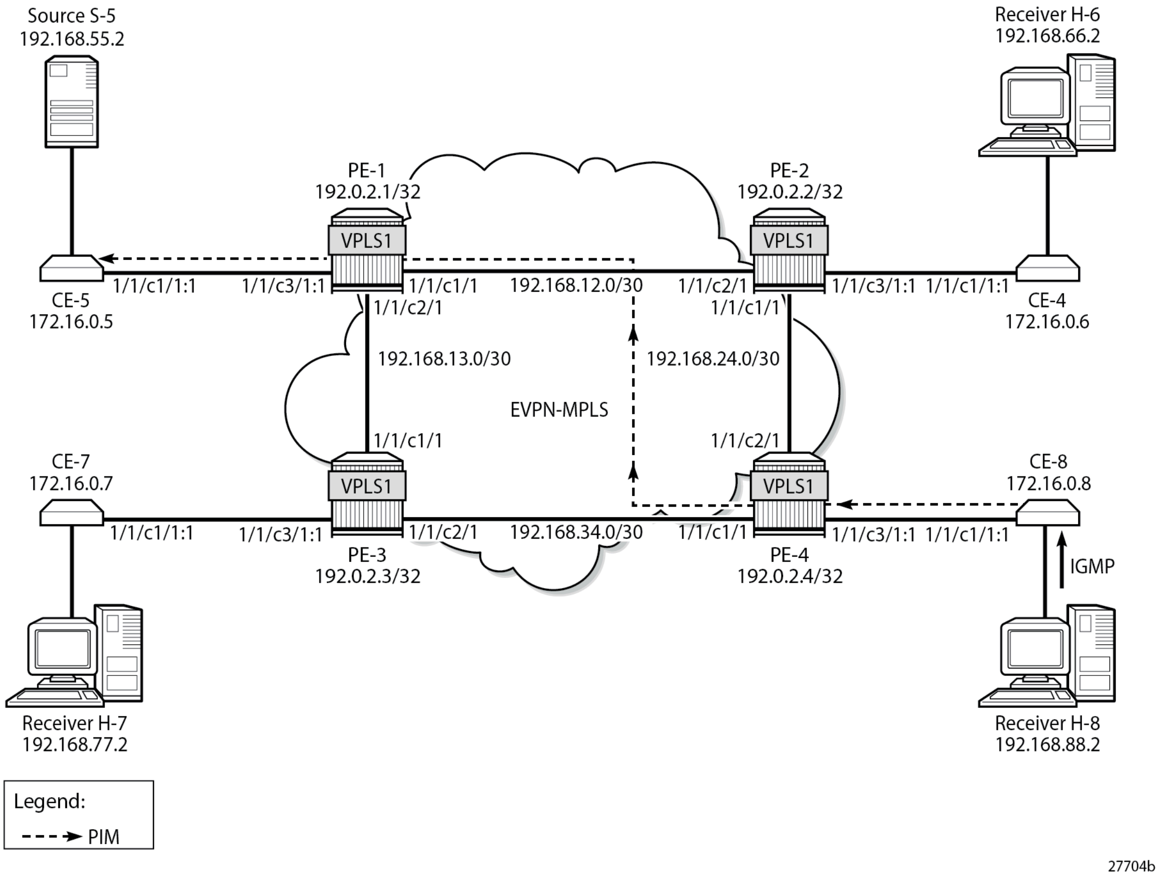
The PIM neighbors of CE-5 are the receiving CEs: CE-6, CE-7, and CE-8, as follows:

```
*A:CE-5# show router pim neighbor

===============================================================================
PIM Neighbor ipv4
===============================================================================
Interface               Nbr DR Prty     Up Time       Expiry Time     Hold Time
   Nbr Address
-------------------------------------------------------------------------------
int-CE-5-PE-1           1               0d 00:03:50   0d 00:01:27     105
   172.16.0.6
int-CE-5-PE-1           1               0d 00:03:34   0d 00:01:42     105
   172.16.0.7
int-CE-5-PE-1           1               0d 00:03:11   0d 00:01:36     105
```

```
   172.16.0.8
-------------------------------------------------------------------------------
Neighbors : 3
===============================================================================
```

Figure 262: H-8 Joins Group (192.168.55.2, 232.1.1.1) and PIM Snooping is Disabled shows that receiver H-8 sends an IGMP report to CE-8 and CE-8 sends a PIM join message to CE-5 via PE-4. PE-4 floods the PIM join message to all PEs, and the message is not snooped by any intermediate PE.

Figure 262: H-8 Joins Group (192.168.55.2, 232.1.1.1) and PIM Snooping is Disabled



Alternatively, a static multicast group can be configured on IGMP interface int-CE-8-H-8 for multicast group (192.168.55.2, 232.1.1.1), as follows:

```
On CE-8:
configure
    router "Base"
        igmp
            interface "int-CE-8-H-8"
                ssm-translate
                    grp-range 232.0.0.0 232.255.255.255
                        source 192.168.55.2
                    exit
                exit
                static
                    group 232.1.1.1 source 192.168.55.2
                exit
                no shutdown
            exit
            no shutdown
        exit all
```

CE-8 sends the following PIM join message for multicast group (192.168.55.2, 232.1.1.1) to upstream IP address 172.16.0.5 on CE-5:

```
3 2023/08/08 09:26:05.675 UTC MINOR: DEBUG #2001 Base PIM[Instance 1 Base]
"PIM[Instance 1 Base]: Join/Prune
[000 00:23:23.670] PIM-TX ifId 3 ifName int-CE-8-PE-4 0.0.0.0 -> 224.0.0.13 Length: 34
PIM Version: 2 Msg Type: Join/Prune Checksum: 0x4828
Upstream Nbr IP : 172.16.0.5 Resvd: 0x0, Num Groups 1, HoldTime 210
Group: 232.1.1.1/32 Num Joined Srcs: 1, Num Pruned Srcs: 0
Joined Srcs:
192.168.55.2/32 Flag S  <S,G>

"
```

Multicast stream 232.1.1.1 is sent from source S-5 to CE-5. When CE-5 has received the PIM join message, it floods the multicast stream to PE-1. Root node PE-1 sends the multicast stream to both PE-2 and PE-3. PE-2 forwards the multicast stream to PE-4 and to CE-6; PE-3 forwards the stream to CE-7, and PE-4 forwards to CE-8. The following PIM group for group address 232.1.1.1 is joined on CE-8:

```
*A:CE-8# show router pim group detail

===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 232.1.1.1
Source Address     : 192.168.55.2
RP Address         : 0
Advt Router        :
Flags              :                    Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 172.16.0.5
MRIB Src Flags     : remote
Keepalive Timer    : Not Running
Up Time            : 0d 00:07:51        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:09
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 172.16.0.5
Incoming Intf      : int-CE-8-PE-4
Outgoing Intf List : int-CE-8-H-8

Curr Fwding Rate   : 9751.560 kbps
Forwarded Packets  : 370106             Discarded Packets  : 0
Forwarded Octets   : 548497092          RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

CE-8 forwards the multicast stream to outgoing interface int-CE-8-H-8 toward receiver H-8, while CE-6 and CE-7 drop the traffic.

The following port statistics show that the incoming traffic on port 1/1/c3/1 on PE-1 is forwarded to port 1/1/c1/1 to PE-2 and to port 1/1/c2/1 to PE-3:

```
*A:PE-1# show port 1/1/c1/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c1/1                                 25                            2501
                                      16474                        24976325
===============================================================================
*A:PE-1# show port 1/1/c2/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c2/1                                 22                            2242
                                      16474                        24976361
===============================================================================
*A:PE-1# show port 1/1/c3/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c3/1                              16454                        24745388
                                          4                             304
===============================================================================
```

Besides the multicast traffic, signaling messages (such as IS-IS or BGP) are sent, which explains the other counters on the ports being different from zero.

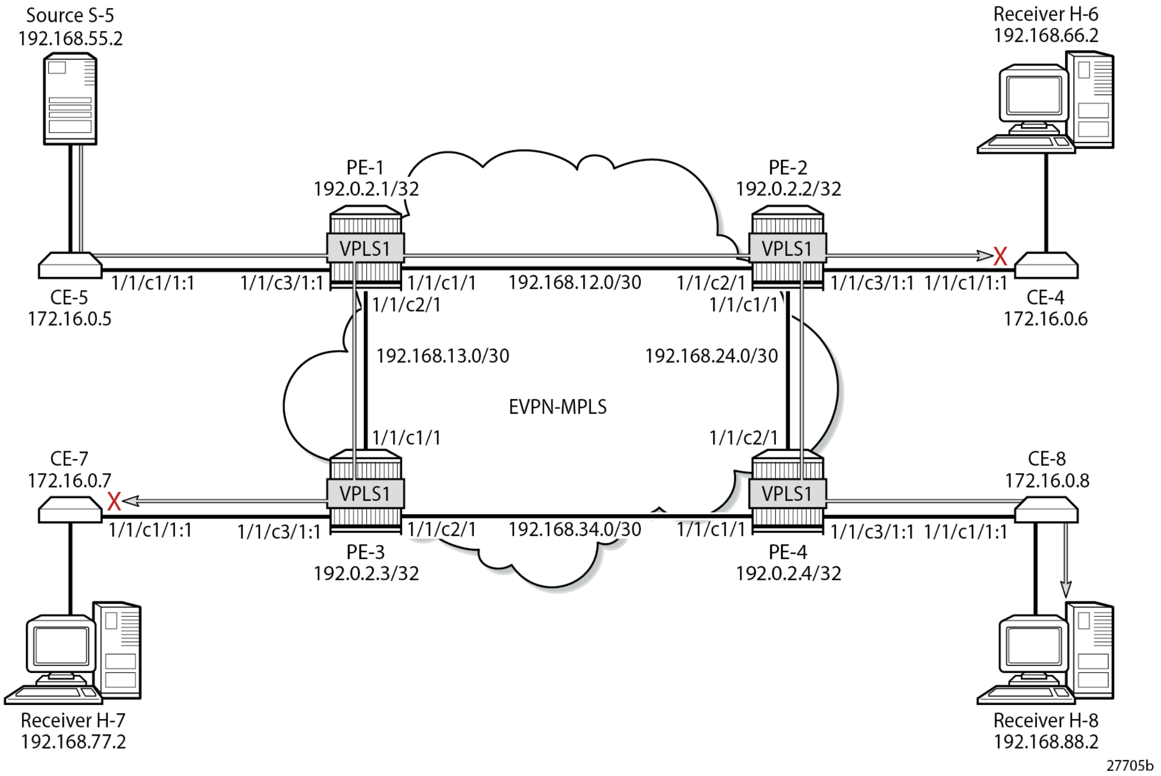
A similar result occurs on PE-2, where incoming traffic from PE-1 is forwarded to PE-4 and to CE-6.

The following port statistics on CE-6 show that the incoming traffic on port 1/1/c1/1 from PE-2 is not forwarded to port 1/1/c2/1 to H-6:

```
*A:CE-6# show port 1/1/c1/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c1/1                              16462                        24755992
                                          0                               0
===============================================================================
*A:CE-6# show port 1/1/c2/1 statistics
```

Without PIM snooping, multicast streams are forwarded to CEs that drop them, which wastes resources. Figure 263: Multicast Stream (192.168.55.2, 232.1.1.1) with PIM Snooping Disabled shows the multicast data streams with receiver H-8 joined and PIM snooping disabled.

Figure 263: Multicast Stream (192.168.55.2, 232.1.1.1) with PIM Snooping Disabled



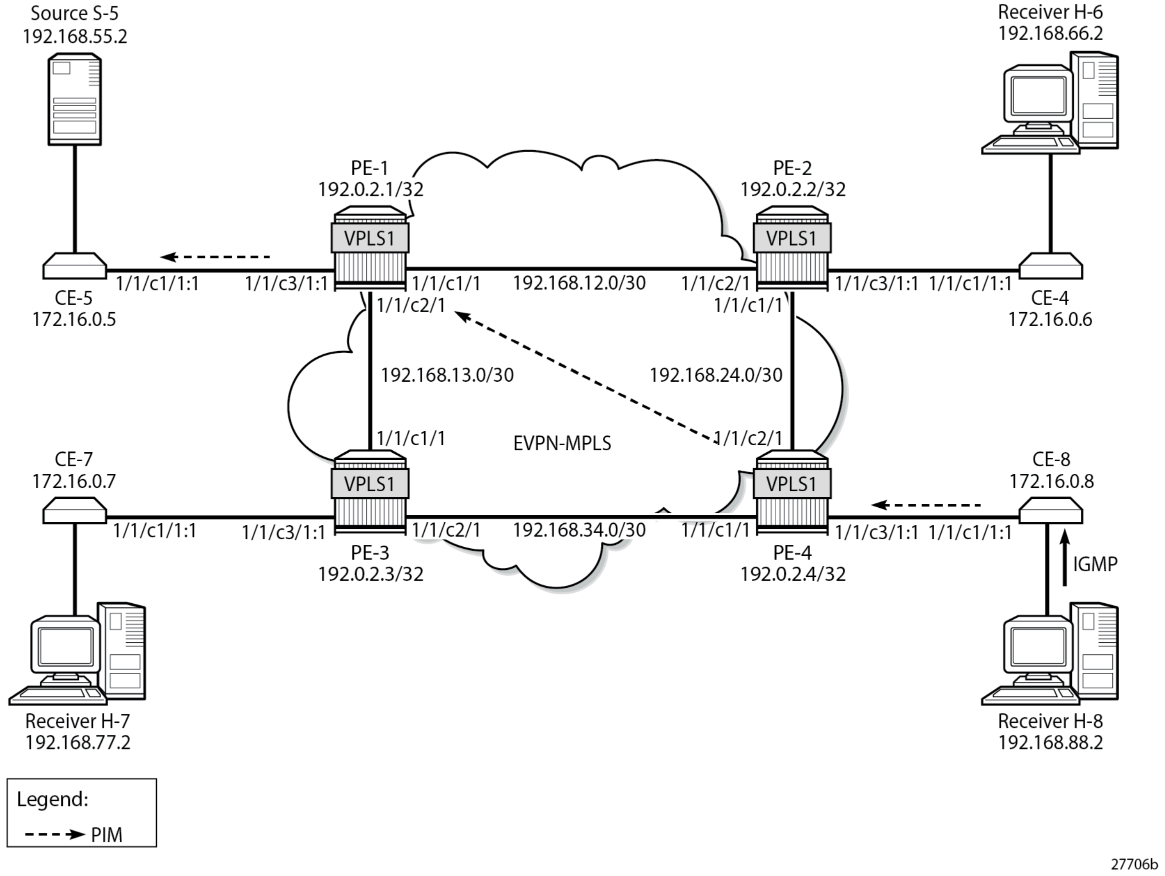
EVPN-MPLS VPLS with PIM Snooping Enabled

PIM snooping is enabled on all PEs as follows:

```
configure service vpls 1 pim-snooping
```

The default mode for PIM snooping is proxy mode, which allows the intermediate PEs to terminate the incoming PIM join or prune messages and create their own PIM join or prune message to be sent toward CE-5, as shown in Figure 264: H-8 Joins (192.168.55.2, 232.1.1.1) and PIM Snooping is Enabled in Proxy Mode:

Figure 264: H-8 Joins (192.168.55.2, 232.1.1.1) and PIM Snooping is Enabled in Proxy Mode



PE-4 receives the following PIM join message for multicast group (192.168.55.2, 232.1.1.1) from CE-8 to CE-5 on SAP 1/1/c3/1:1:

```
17 2023/08/08 09:31:26.206 UTC MINOR: DEBUG #2001 Base PIM[vpls 1 ]
"PIM[vpls 1 ]: Join/Prune
[000 00:28:47.290] PIM-RX ifId 1 ifName SAP:1/1/c3/1:1 172.16.0.8 -> 224.0.0.13 Length: 34
PIM Version: 2 Msg Type: Join/Prune Checksum: 0x4828
Upstream Nbr IP : 172.16.0.5 Resvd: 0x0, Num Groups 1, HoldTime 210
Group: 232.1.1.1/32 Num Joined Srcs: 1, Num Pruned Srcs: 0
Joined Srcs:
192.168.55.2/32 Flag S  <S,G>

"
```

PE-4 sends the following PIM join message for multicast group (192.168.55.2, 232.1.1.1) to CE-5 on interface EVPN-MPLS:

```
18 2023/08/08 09:31:26.206 UTC MINOR: DEBUG #2001 Base PIM[vpls 1 ]
"PIM[vpls 1 ]: Join/Prune
[000 00:28:47.290] PIM-TX ifId 1071394 ifName EVPN-MPLS 0.0.0.0 -> 224.0.0.13 Length: 34
PIM Version: 2 Msg Type: Join/Prune Checksum: 0x4828
Upstream Nbr IP : 172.16.0.5 Resvd: 0x0, Num Groups 1, HoldTime 210
Group: 232.1.1.1/32 Num Joined Srcs: 1, Num Pruned Srcs: 0
Joined Srcs:
```

```
192.168.55.2/32 Flag S  <S,G>

"
```

In a similar way, PE-1 terminates this PIM join message and sends the following PIM join message for multicast group (192.168.55.2, 232.1.1.1) to CE-5 on SAP 1/1/c3/1:1.

```
16 2023/08/08 09:31:26.559 UTC MINOR: DEBUG #2001 Base PIM[vpls 1 ]
"PIM[vpls 1 ]: Join/Prune
[000 00:28:59.640] PIM-TX ifId 2 ifName SAP:1/1/c3/1:1 0.0.0.0 -> 224.0.0.13 Length: 34
PIM Version: 2 Msg Type: Join/Prune Checksum: 0x4828
Upstream Nbr IP : 172.16.0.5 Resvd: 0x0, Num Groups 1, HoldTime 210
Group: 232.1.1.1/32 Num Joined Srcs: 1, Num Pruned Srcs: 0
Joined Srcs:
192.168.55.2/32 Flag S  <S,G>

"
```

The following command shows the status of PIM snooping in VPLS 1 on PE-1:

```
*A:PE-1# show service id 1 pim-snooping status

===============================================================================
PIM Snooping Status ipv4
===============================================================================
Admin State                       : Up
Oper State                        : Up
Mode Admin                        : Proxy
Mode Oper                         : Proxy
Hold Time                         : 90
Designated Router                 : 172.16.0.8
J/P Tracking                      : Inactive
Up Time                           : 0d 00:03:39
Group Policy                      : None
===============================================================================
```

The following PIM snooping statistics show the number of received and transmitted PIM messages, and the source group statistics: one (S,G) group is joined and no (*,G) group.

```
*A:PE-1# show service id 1 pim-snooping statistics

=================================================================
PIM Snooping Statistics ipv4
=================================================================
Message Type        Received       Transmitted    Rx Errors
-----------------------------------------------------------------
Hello               29             -              0
Join Prune          2              2              0
Total Packets       31             2
-------------------------------------------------------------------------------
General Statistics
-------------------------------------------------------------------------------
Rx Neighbor Unknown               : 0
Rx Bad Checksum Discard           : 0
Rx Bad Encoding                   : 0
Rx Bad Version Discard            : 0
Join Policy Drops                 : 0
-------------------------------------------------------------------------------
Source Group Statistics
-------------------------------------------------------------------------------
(S,G)                             : 1
```

```
(*,G)                             : 0
=================================================================
```

PE-4 has four neighbors for PIM snooping: the local SAP toward CE-8 and the EVPN-MPLS destinations toward the other CEs, as follows:

```
*A:PE-4# show service id 1 pim-snooping neighbor

===============================================================================
PIM Snooping Neighbors ipv4
===============================================================================
Port Id                 Nbr DR Prty     Up Time       Expiry Time     Hold Time
  Nbr Address
-------------------------------------------------------------------------------
SAP:1/1/c3/1:1          1               0d 00:03:19   0d 00:01:26     105
  172.16.0.8
EVPN-MPLS               1               0d 00:03:24   0d 00:01:21     105
  172.16.0.5
EVPN-MPLS               1               0d 00:03:28   0d 00:01:17     105
  172.16.0.6
EVPN-MPLS               1               0d 00:03:12   0d 00:01:33     105
  172.16.0.7
-------------------------------------------------------------------------------
Neighbors : 4
===============================================================================
```

The EVPN-MPLS destinations appear as a single entry with port ID "EVPN-MPLS" in the following **show** command:

```
*A:PE-4# show service id 1 pim-snooping port

===============================================================================
PIM Snooping Port ipv4
===============================================================================
Port Id                                                         Opr   PW Fwding
-------------------------------------------------------------------------------
SAP:1/1/c3/1:1                                                  Up    Actv
EVPN-MPLS                                                       Up    Actv
===============================================================================
```

In the MFIB output on PE-1 and PE-4, each EVPN-MPLS destination is shown individually, but the information for each EVPN-MPLS destination is identical, as follows:

```
*A:PE-1# show service id 1 mfib

===============================================================================
Multicast FIB, Service 1
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
192.168.55.2    232.1.1.1             sap:1/1/c3/1:1               Local    Fwd
                                      mpls:192.0.2.2:524282        Local    Fwd
                                      mpls:192.0.2.3:524282        Local    Fwd
                                      mpls:192.0.2.4:524282        Local    Fwd
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

On PE-2 and PE-3, the MFIB has no entries, as follows:

```
*A:PE-2# show service id 1 mfib

===============================================================================
Multicast FIB, Service 1
===============================================================================
Source Address  Group Address         Port Id                       Svc Id   Fwd
                                                                             Blk
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Number of entries: 0
===============================================================================
```

The MFIB statistics for VPLS 1 on PE-1 show the number of matched packets and matched octets for multicast group (192.168.55.2, 232.1.1.1), as follows:

```
*A:PE-1# show service id 1 mfib statistics

===============================================================================
Multicast FIB Statistics, Service 1
===============================================================================
Source Address  Group Address       Matched Pkts          Matched Octets
                                                          Forwarding Rate
-------------------------------------------------------------------------------
192.168.55.2    232.1.1.1           82698                 124047000
                                                          9866.817 kbps
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

The following **show** command of the PIM group snooped on PE-1 shows the SAP toward the source as incoming interface, and the EVPN-MPLS interface as outgoing interface (traffic coming in from the source is not sent back to the SAP toward the source):

```
*A:PE-1# show service id 1 pim-snooping group 232.1.1.1 detail

===============================================================================
PIM Snooping Source Group ipv4
===============================================================================
Group Address      : 232.1.1.1
Source Address     : 192.168.55.2
Up Time            : 0d 00:01:22

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:38
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

RPF Neighbor       : 172.16.0.5
Incoming Intf      : SAP:1/1/c3/1:1
Outgoing Intf List : EVPN-MPLS, SAP:1/1/c3/1:1

Forwarded Packets  : 68202              Forwarded Octets   : 102303000
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

The following identical **show** command of the PIM group snooped on PE-4 shows the EVPN-MPLS interface as incoming interface. Even though the EVPN-MPLS interface is also listed as outgoing interface, traffic coming from that interface is not forwarded on that interface (all EVPN-MPLS destinations are treated as one single EVPN-MPLS interface), so the traffic is forwarded to the SAP toward the receiving CE only.

```
*A:PE-4# show service id 1 pim-snooping group 232.1.1.1 detail

===============================================================================
PIM Snooping Source Group ipv4
===============================================================================
Group Address      : 232.1.1.1
Source Address     : 192.168.55.2
Up Time            : 0d 00:01:30

Up JP State        : Joined            Up JP Expiry       : 0d 00:00:30
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

RPF Neighbor       : 172.16.0.5
Incoming Intf      : EVPN-MPLS
Outgoing Intf List : EVPN-MPLS, SAP:1/1/c3/1:1

Forwarded Packets  : 74349             Forwarded Octets   : 111226104
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

The following port statistics on PE-2 show that the multicast stream coming in from PE-1 on port 1/1/c2/1 is forwarded to port 1/1/c1/1 toward PE-4 only, but not to port 1/1/c3/1 toward CE-6:

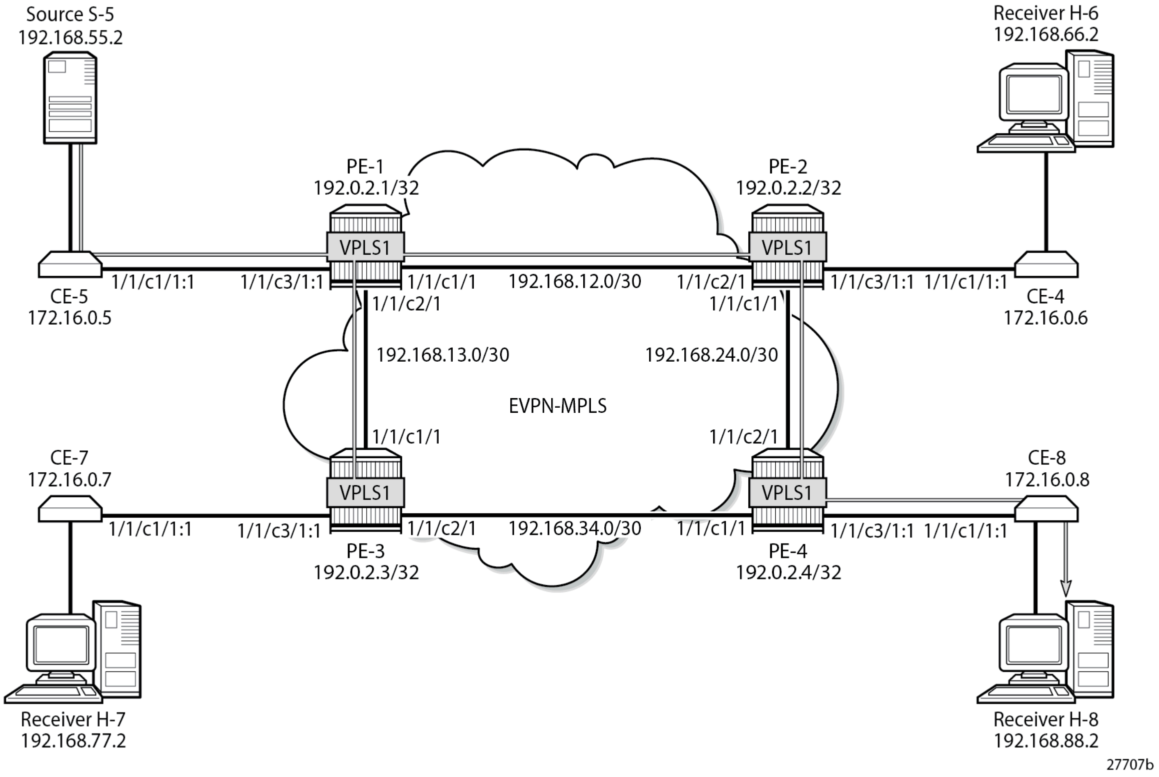
```
*A:PE-2# show port 1/1/c1/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c1/1                                 15                            1567
                                      16467                        24972851
===============================================================================
*A:PE-2# show port 1/1/c2/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c2/1                              16473                        24973338
                                         23                            2286
===============================================================================
*A:PE-2# show port 1/1/c3/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c3/1                                  1                              76
                                          1                              76
===============================================================================
```

In a similar way, the multicast traffic on PE-3 that comes in from PE-1 via port 1/1/c1/1 is not forwarded to any port, as follows:

```
*A:PE-3# show port 1/1/c1/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c1/1                              16487                        24996095
                                         20                            2054
===============================================================================
*A:PE-3# show port 1/1/c2/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c2/1                                 17                            1786
                                         16                            1687
===============================================================================
*A:PE-3# show port 1/1/c3/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c3/1                                  0                               0
                                          3                             228
===============================================================================
```

Figure 265: Multicast Stream (192.168.55.2, 232.1.1.1) with PIM Snooping Enabled shows that the multicast stream still flows from the source S-5 to the receiver H-8, but is not forwarded to CE-6 and CE-7 when PIM snooping is enabled. The root node PE-1 sends the multicast traffic received on the SAP to all EVPN-MPLS destinations over the P2MP mLDP provider tunnel. The EVPN-MPLS interface is treated as a single interface.

Figure 265: Multicast Stream (192.168.55.2, 232.1.1.1) with PIM Snooping Enabled

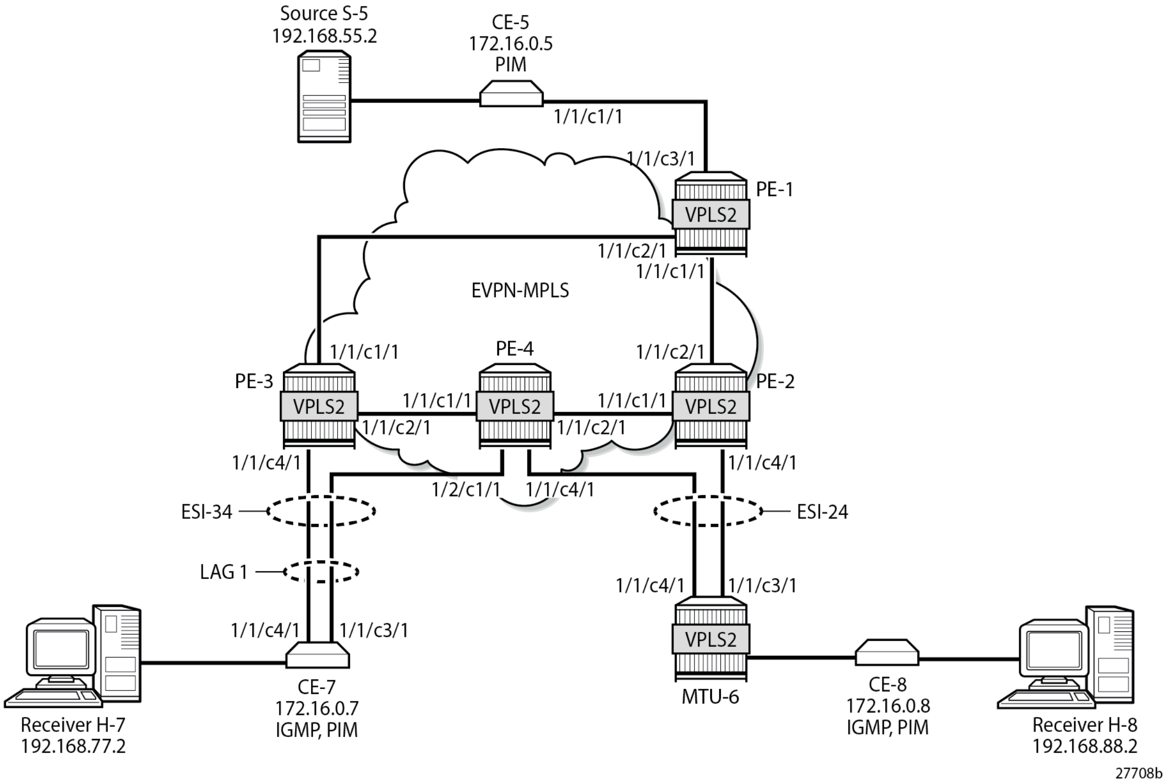


Multi-homed EVPN-MPLS VPLS without PIM Snooping

When CE-5 receives a PIM join message, it forwards the multicast stream to PE-1. All multicast traffic in VPLS 2 is sent to all receiving CEs, regardless of the received PIM join messages.

Figure 266: Example Topology with Multi-homing ESs shows the example topology with an all-active multi-homing virtual Ethernet Segment (ES) "ESI-34_2" between PE-3 and PE-4 using a LAG, and a single-active multi-homing ES "ESI-24" between PE-2 and PE-4 using SDPs.

Figure 266: Example Topology with Multi-homing ESs



The configuration of VPLS 2 is similar to the configuration of VPLS 1 on all PEs. An identical P2MP mLDP provider tunnel is established on the PEs for VPLS 2: PE-1 is the root node, PE-2 is a leaf node and a transit node, PE-3 is a leaf node, and PE-4 is also a leaf node.

On PE-2, PE-3, and PE-4, one or more ESs are configured. The service configuration on PE-2 is as follows. An SDP is configured toward MTU-6 that is associated with a single-active multi-homing ES "ESI-24". Spoke-SDP 26:2 is associated with VPLS 2.

```
On PE-2:
configure
    service
        sdp 26 mpls create
            far-end 192.0.2.6
            ldp
            no shutdown
        exit
        system
            bgp-evpn
                ethernet-segment "ESI-24" create
                    esi 01:00:00:00:00:24:00:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode manual
                        manual
                            preference non-revertive create
                                value 10000
                            exit
                        exit
                    exit
                    multi-homing single-active
                    sdp 26
                    no shutdown
                exit
            exit
        exit
        vpls 2 name "VPLS 2" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 2
                mpls bgp 1
                    ingress-replication-bum-label
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            provider-tunnel
                inclusive
                    owner bgp-evpn-mpls
                    no root-and-leaf        # default
                    mldp
                    no shutdown
                exit
            exit
            spoke-sdp 26:2 create
            exit
            no shutdown
        exit
    exit all
```

The same ES is configured on PE-4, together with another ES-an all-active multi-homing virtual ES that applies to VPLS 2 only (**q-tag-range 2**); see chapter Virtual Ethernet Segments. The preference for the DF election is configured manually to a value of 5000 (which is lower than the preference 10000 on PE-3); see chapter Preference-based and Non-revertive EVPN DF Election. VPLS 2 has a SAP and a spoke-SDP configured. The service configuration on PE-4 is as follows:

```
On PE-4:
configure
    service
        sdp 46 mpls create
            far-end 192.0.2.6
            ldp
            no shutdown
        exit
        system
            bgp-evpn
                ethernet-segment "ESI-24" create
                    esi 01:00:00:00:00:24:00:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode manual
                        manual
                            preference non-revertive create
                                value 5000
                            exit
                        exit
                    exit
                    multi-homing single-active
                    sdp 46
                    no shutdown
                exit
                ethernet-segment "ESI-34_2" virtual create
                    esi 01:00:00:00:00:34:02:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode manual
                        manual
                            preference non-revertive create
                                value 5000
                            exit
                        exit
                    exit
                    multi-homing all-active
                    lag 1
                    dot1q
                        q-tag-range 2
                    exit
                    no shutdown
                exit
            exit
        exit
        vpls 2 name "VPLS 2" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 2
                mpls bgp 1
                    ingress-replication-bum-label
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            provider-tunnel
                inclusive
                    owner bgp-evpn-mpls
                    no root-and-leaf        # default
                    mldp
                    no shutdown
                exit
            exit
            sap lag-1:2 create
            exit
            spoke-sdp 46:2 create
            exit
            no shutdown
        exit
    exit all
```

The service configuration on PE-3 includes the same all-active multi-homing virtual ES with preference 10000, as follows:

```
On PE-3:
configure
    service
        system
            bgp-evpn
                ethernet-segment "ESI-34_2" virtual create
                    esi 01:00:00:00:00:34:02:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode manual
                        manual
                            preference non-revertive create
                                value 10000
                            exit
                        exit
                    exit
                    multi-homing all-active
                    lag 1
                    dot1q
                        q-tag-range 2
                    exit
                    no shutdown
                exit
            exit
        exit
        vpls 2 name "VPLS 2" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 2
                mpls bgp 1
                    ingress-replication-bum-label
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            provider-tunnel
                inclusive
                    owner bgp-evpn-mpls
                    no root-and-leaf        # default
                    mldp
                    no shutdown
                exit
            exit
            sap lag-1:2 create
            exit
            no shutdown
        exit
    exit all
```

The following is the service configuration on MTU-6:

```
On MTU-6:
configure
    service
        sdp 62 mpls create
            far-end 192.0.2.2
            ldp
            no shutdown
        exit
        sdp 64 mpls create
            far-end 192.0.2.4
            ldp
            no shutdown
        exit
        vpls 2 name "VPLS 2" customer 1 create
            endpoint "x" create
            exit
            sap 1/2/c1/1:2 create
            exit
            spoke-sdp 62:2 endpoint "x" create
            exit
            spoke-sdp 64:2 endpoint "x" create
            exit
            no shutdown
        exit
    exit all
```

For VPLS 2, PE-2 is the DF in ES "ESI-24", as follows:

```
*A:PE-2# show service id 2 ethernet-segment
No sap entries

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
26:2                  ESI-24                           DF
===============================================================================
No vxlan instance entries
```

PE-3 is the DF in ES "ESI-34_2", as follows:

```
*A:PE-3# show service id 2 ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
lag-1:2               ESI-34_2                         DF
===============================================================================
No sdp entries
No vxlan instance entries
```

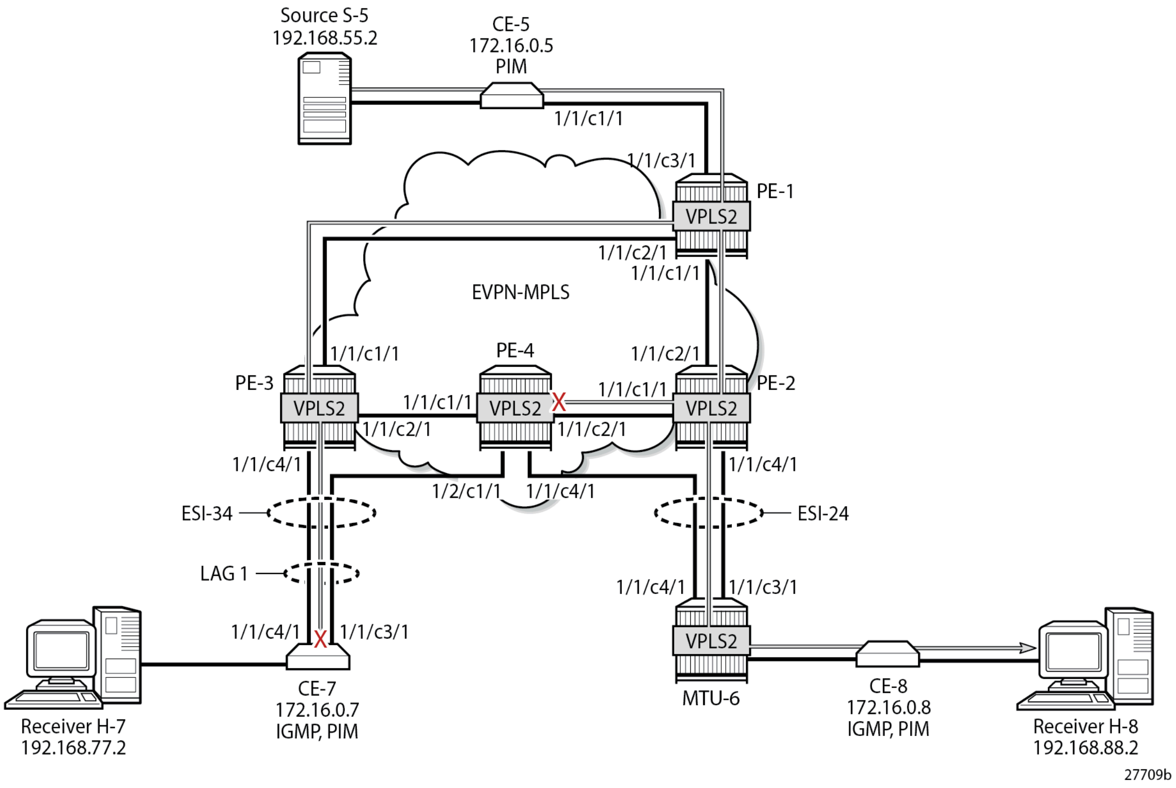
PE-4 is NDF for both ESI-24 and ESI-34_2, as follows:

```
*A:PE-4# show service id 2 ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
lag-1:2               ESI-34_2                         NDF
===============================================================================

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
46:2                  ESI-24                           NDF
===============================================================================
No vxlan instance entries
```

When H-8 sends an IGMP report to join multicast group 232.1.1.1 from source 192.168.55.2, CE-5 forwards the multicast stream after receiving the corresponding PIM join message. PE-1 forwards the multicast traffic on the P2MP mLDP tree to PE-2, PE-3, and PE-4. The DF PE-2 forwards the traffic to

MTU-6, and DF PE-3 forwards it to CE-7, even though a PIM join for this group has not been received from CE-7. PE-4 is NDF, so it does not forward the traffic to MTU-6 or CE-7. MTU-6 forwards the traffic to CE-8, which sends it to H-8. CE-7 drops the multicast traffic because no attached receiver has joined the multicast group. Figure 267: EVPN-MPLS with Multi-homing – Receiver H-8 Joined shows how this multicast is forwarded when PIM snooping is disabled.

Figure 267: EVPN-MPLS with Multi-homing – Receiver H-8 Joined



The static IGMP multicast is removed to emulate an IGMPv3 report from receiver H-8 to exclude multicast group 232.1.1.1 from source 192.168.55.2, as follows:

```
On CE-8:
configure
    router "Base"
        igmp
            interface "int-CE-8-H-8"
                static
                    no group 232.1.1.1
                exit
                no shutdown
            exit
            no shutdown
        exit all
```

Multi-homed EVPN-MPLS VPLS with PIM Snooping

PIM snooping is enabled in VPLS 2 on all PEs, including PE-1, which is not part of an ES-with the following command:

```
configure service vpls 2 pim-snooping
```

All PEs have three PIM snooping neighbors: CE-5, CE-7, and CE-8. The list of PIM snooping neighbors on PE-1 is as follows:

```
*A:PE-1# show service id 2 pim-snooping neighbor

===============================================================================
PIM Snooping Neighbors ipv4
===============================================================================
Port Id                 Nbr DR Prty     Up Time       Expiry Time     Hold Time
  Nbr Address
-------------------------------------------------------------------------------
SAP:1/1/c3/1:2          1               0d 00:00:56   0d 00:01:18     105
  172.16.0.5
EVPN-MPLS               1               0d 00:01:05   0d 00:01:40     105
  172.16.0.7
EVPN-MPLS               1               0d 00:01:05   0d 00:01:40     105
  172.16.0.8
-------------------------------------------------------------------------------
Neighbors : 3
===============================================================================
```

On PE-2, the same PIM snooping neighbors are listed: CE-5, CE-7, and CE-8, as follows:

```
*A:PE-2# show service id 2 pim-snooping neighbor

===============================================================================
PIM Snooping Neighbors ipv4
===============================================================================
Port Id                 Nbr DR Prty     Up Time       Expiry Time     Hold Time
  Nbr Address
-------------------------------------------------------------------------------
SPOKE_SDP:26:2          1               0d 00:01:07   0d 00:01:37     105
  172.16.0.8
EVPN-MPLS               1               0d 00:00:59   0d 00:01:15     105
  172.16.0.5
EVPN-MPLS               1               0d 00:01:08   0d 00:01:37     105
  172.16.0.7
-------------------------------------------------------------------------------
Neighbors : 3
===============================================================================
```

PE-3 and PE-4 also have these three CEs as PIM snooping neighbors.

All-active MH EVPN-MPLS VPLS with PIM Snooping

On CE-7, the following static IGMP membership is configured on interface int-CE-7-H-7:

```
On CE-7:
configure
    router "Base"
```

```
        igmp
            interface "int-CE-7-H-7"
                ssm-translate
                    grp-range 232.0.0.0 232.255.255.255
                        source 192.168.55.2
                    exit
                exit
                static
                    group 232.1.1.1 source 192.168.55.2
                exit
                no shutdown
            exit
            no shutdown
        exit all
```

When H-7 joins the multicast group 232.1.1.1 via source 192.168.55.2, the PIM join messages are snooped by the PEs and the MFIB is built. The MFIB on PE-1 contains one entry for group address 232.1.1.1 and source address 192.168.55.2, with four port IDs: the local SAP to CE-5 and the EVPN-MPLS destinations, as follows:

```
*A:PE-1# show service id 2 mfib

===============================================================================
Multicast FIB, Service 2
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
192.168.55.2    232.1.1.1             sap:1/1/c3/1:2               Local    Fwd
                                      mpls:192.0.2.2:524279        Local    Fwd
                                      mpls:192.0.2.3:524279        Local    Fwd
                                      mpls:192.0.2.4:524278        Local    Fwd
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

The MFIB on PE-2 is empty because no locally attached node has sent a PIM join for any multicast group:

```
*A:PE-2# show service id 2 mfib

===============================================================================
Multicast FIB, Service 2
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Number of entries: 0
===============================================================================
```

The MFIB on PE-3 contains an entry for the (S,G) with the local SAP lag-1:2 and the EVPN-MPLS destination, as follows:

```
*A:PE-3# show service id 2 mfib

===============================================================================
Multicast FIB, Service 2
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
```

```
192.168.55.2    232.1.1.1            sap:lag-1:2                 Local    Fwd
                                     mpls:192.0.2.1:524280       Local    Fwd
                                     mpls:192.0.2.2:524279       Local    Fwd
                                     mpls:192.0.2.4:524278       Local    Fwd
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

Data-driven PIM state synchronization between PE-3 and PE-4 in the ESI-34_2 results in the following MFIB entry on PE-4:

```
*A:PE-4# show service id 2 mfib

===============================================================================
Multicast FIB, Service 2
===============================================================================
Source Address  Group Address         Port Id                     Svc Id   Fwd
                                                                           Blk
-------------------------------------------------------------------------------
192.168.55.2    232.1.1.1             sap:lag-1:2                 Local    Fwd
                                      mpls:192.0.2.1:524280       Local    Fwd
                                      mpls:192.0.2.2:524279       Local    Fwd
                                      mpls:192.0.2.3:524279       Local    Fwd
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

When debugging is enabled on the PEs as follows, the synchronization between peers in ES "ESI-34_2" is logged:

```
debug
    service
        id 2
            pim-snooping
                jp
                packet evpn-mpls
            exit all
```

For example, PE-4 sends the following PIM message to its remote peer PE-3 in ESI-34_2:

```
87 2023/08/08 09:47:08.842 UTC MINOR: DEBUG #2001 Base PIM[vpls 2 ]
"PIM[vpls 2 ]: pimVplsFwdJPToEvpn
Forwarding to remote peer on bgp-evpn ethernet-segment ESI-34_2"
```

PE-3 receives the following PIM message from its remote peer PE-4 in ESI-34_2:

```
65 2023/08/08 09:45:40.009 UTC MINOR: DEBUG #2001 Base PIM[vpls 2 ]
"PIM[vpls 2 ]: pimProcessPdu
Received from remote peer on bgp-evpn ethernet-segment ESI-34_2, will be applied on lag-1:2
"
```

On PE-1, the PIM snooping group (192.168.55.2, 232.1.1.1) has incoming interface SAP 1/1/c3/1:2 toward CE-5 and the EVPN-MPLS interface as outgoing interface, as follows:

```
*A:PE-1# show service id 2 pim-snooping group detail

===============================================================================
PIM Snooping Source Group ipv4
===============================================================================
Group Address      : 232.1.1.1
Source Address     : 192.168.55.2
Up Time            : 0d 00:00:51

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:08
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

RPF Neighbor       : 172.16.0.5
Incoming Intf      : SAP:1/1/c3/1:2
Outgoing Intf List : EVPN-MPLS, SAP:1/1/c3/1:2

Forwarded Packets  : 42672              Forwarded Octets   : 64008000
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

On PE-2, no PIM join messages are received and no groups are listed, as follows:

```
*A:PE-2# show service id 2 pim-snooping group detail

===============================================================================
PIM Snooping Source Group ipv4
===============================================================================
No Matching Entries
===============================================================================
```

On PE-3, the same PIM snooping group has the EVPN-MPLS as incoming interface and the SAP lag-1:2 as outgoing interface. The split-horizon mechanism ensures that the multicast traffic that enters through the EVPN-MPLS interface is not forwarded on the EVPN-MPLS interface, which is regarded as a single interface.

```
*A:PE-3# show service id 2 pim-snooping group detail

===============================================================================
PIM Snooping Source Group ipv4
===============================================================================
Group Address      : 232.1.1.1
Source Address     : 192.168.55.2
Up Time            : 0d 00:00:55

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:14
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

RPF Neighbor       : 172.16.0.5
Incoming Intf      : EVPN-MPLS
Outgoing Intf List : EVPN-MPLS, SAP:lag-1:2

Forwarded Packets  : 45745              Forwarded Octets   : 68434520
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

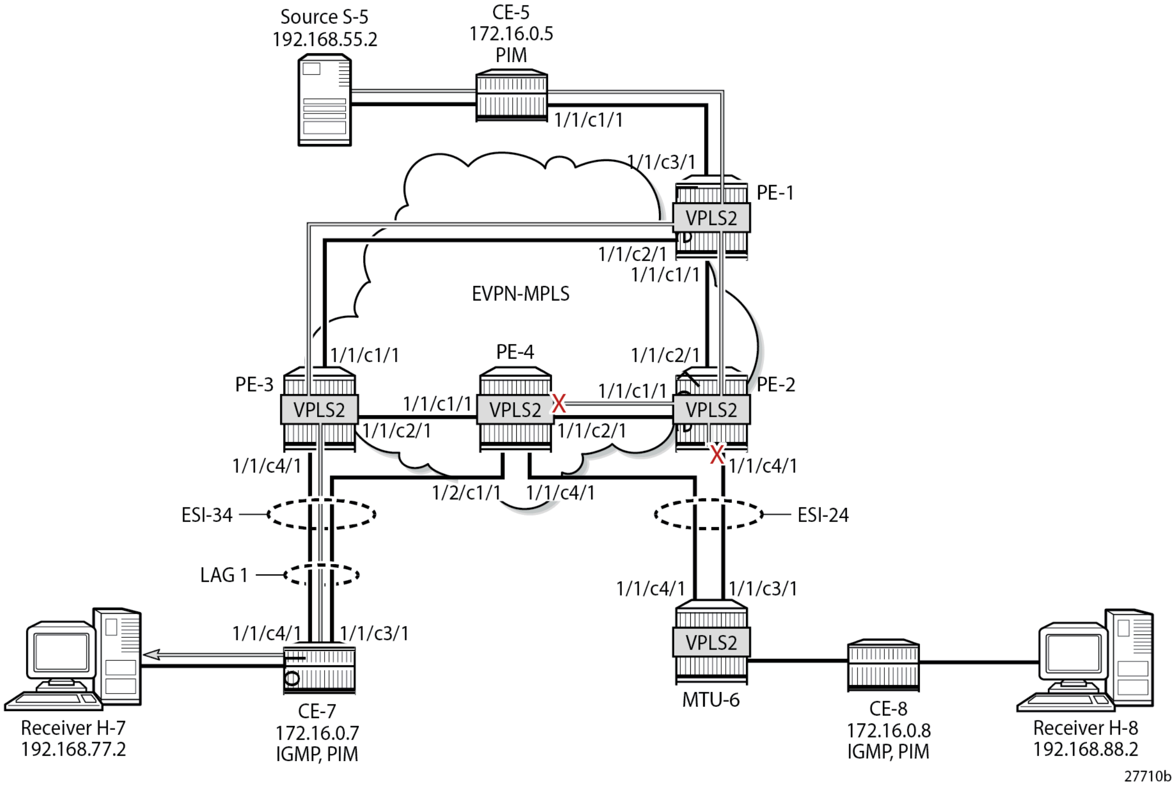
On PE-4, the same PIM snooping information is available, because of the data-driven PIM state synchronization between PE-3 and PE-4 in ESI-34_2, as follows:

```
*A:PE-4# show service id 2 pim-snooping group detail

===============================================================================
PIM Snooping Source Group ipv4
===============================================================================
Group Address      : 232.1.1.1
```

```
Source Address      : 192.168.55.2
Up Time             : 0d 00:00:57

Up JP State         : Joined              Up JP Expiry       : 0d 00:00:13
Up JP Rpt           : Not Joined StarG    Up JP Rpt Override : 0d 00:00:00

RPF Neighbor        : 172.16.0.5
Incoming Intf       : EVPN-MPLS
Outgoing Intf List : EVPN-MPLS, SAP:lag-1:2

Forwarded Packets   : 46387               Forwarded Octets   : 69394952
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

Figure 268: EVPN-MPLS with All-active Multi-homing and PIM Snooping Enabled – Receiver H-7 Joined shows how the multicast traffic is forwarded when H-7 joins the multicast group and PIM snooping is enabled. DF PE-3 forwards the traffic toward CE-7. The multicast stream also reaches PE-2 and PE-4, where it is dropped.

Figure 268: EVPN-MPLS with All-active Multi-homing and PIM Snooping Enabled – Receiver H-7 Joined

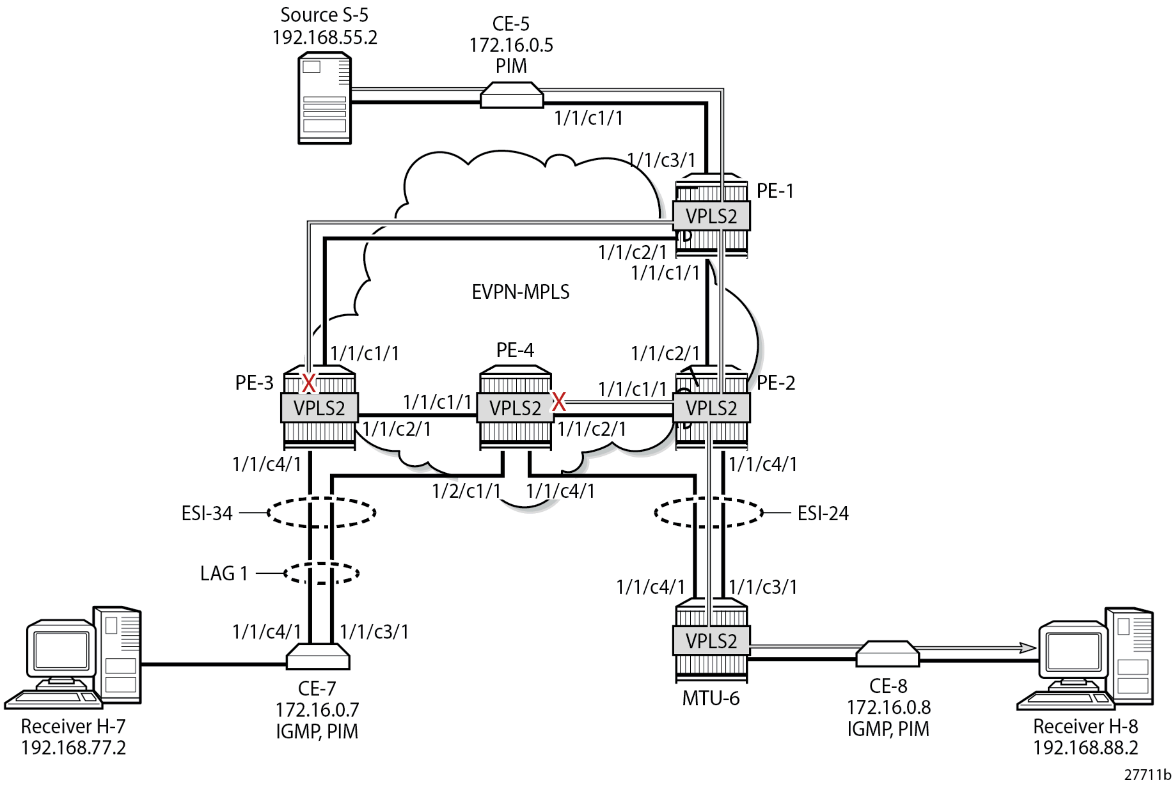


H-7 leaves the multicast group and H-8 joins it instead.

Single-active MH EVPN-MPLS VPLS with PIM Snooping

When H-8 joins the multicast group and PIM snooping is enabled, only DF PE-2 forwards traffic from the EVPN-MPLS toward a receiver. PE-3 does not forward traffic to CE-7 because no PIM join message was received from CE-7. Figure 269: EVPN-MPLS with Single-active Multi-homing and PIM Snooping Enabled – Receiver H-8 Joined shows how the multicast traffic is forwarded when H-8 joins the multicast group and PIM snooping is enabled.

Figure 269: EVPN-MPLS with Single-active Multi-homing and PIM Snooping Enabled – Receiver H-8 Joined



On PE-1, the MFIB looks the same as in the preceding case, as follows:

```
*A:PE-1# show service id 2 mfib

===============================================================================
Multicast FIB, Service 2
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
192.168.55.2    232.1.1.1             sap:1/1/c3/1:2               Local    Fwd
                                      mpls:192.0.2.2:524279        Local    Fwd
                                      mpls:192.0.2.3:524279        Local    Fwd
                                      mpls:192.0.2.4:524278        Local    Fwd
-------------------------------------------------------------------------------
Number of entries: 1
```

```
===============================================================================
```

On PE-2, the MFIB contains an entry for source address 192.168.55.2 and group address 232.1.1.1 with spoke-SDP 26:2 and the EVPN-MPLS destinations to the other PEs, as follows:

```
*A:PE-2# show service id 2 mfib

===============================================================================
Multicast FIB, Service 2
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
192.168.55.2    232.1.1.1             sdp:26:2                     Local    Fwd
                                      mpls:192.0.2.1:524280        Local    Fwd
                                      mpls:192.0.2.3:524279        Local    Fwd
                                      mpls:192.0.2.4:524278        Local    Fwd
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

The MFIB on PE-3 is empty, because multicast traffic toward H-8 is not sent via PE-3, as follows:

```
*A:PE-3# show service id 2 mfib

===============================================================================
Multicast FIB, Service 2
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Number of entries: 0
===============================================================================
```

The data-driven PIM state synchronization ensures that DF PE-2 sends updates to NDF PE-4. With debugging enabled, the following debug message is displayed at PE-2:

```
203 2023/08/08 09:48:51.571 UTC MINOR: DEBUG #2001 Base PIM[vpls 2 ]
"PIM[vpls 2 ]: pimVplsFwdJPToEvpn
Forwarding to remote peer on bgp-evpn ethernet-segment ESI-24"
```

The following debug message is displayed at PE-4:

```
74 2023/08/08 09:45:40.499 UTC MINOR: DEBUG #2001 Base PIM[vpls 2 ]
"PIM[vpls 2 ]: pimProcessPdu
Received from remote peer on bgp-evpn ethernet-segment ESI-24, will be applied on 46:2
"
```

As a result, the MFIB on PE-4 is not empty, as follows:

```
*A:PE-4# show service id 2 mfib

===============================================================================
Multicast FIB, Service 2
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
192.168.55.2    232.1.1.1             sdp:46:2                     Local    Fwd
```

```
                                      mpls:192.0.2.1:524280        Local    Fwd
                                      mpls:192.0.2.2:524279        Local    Fwd
                                      mpls:192.0.2.3:524279        Local    Fwd
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

On PE-1, the PIM snooping group (192.168.55.2, 232.1.1.1) has incoming interface SAP 1/1/c3/1:2 toward CE-5 and the EVPN-MPLS interface as outgoing interface, as follows:

```
*A:PE-1# show service id 2 pim-snooping group detail

===============================================================================
PIM Snooping Source Group ipv4
===============================================================================
Group Address      : 232.1.1.1
Source Address     : 192.168.55.2
Up Time            : 0d 00:00:43

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:16
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

RPF Neighbor       : 172.16.0.5
Incoming Intf      : SAP:1/1/c3/1:2
Outgoing Intf List : EVPN-MPLS, SAP:1/1/c3/1:2

Forwarded Packets  : 36064              Forwarded Octets   : 54096000
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

On PE-2, the same PIM snooping group has the EVPN-MPLS as incoming interface and the spoke-SDP 26:2 as outgoing interface. Again, the split-horizon mechanism ensures that the multicast traffic that enters through the EVPN-MPLS interface is not forwarded on the EVPN-MPLS interface, which is regarded as a single interface.

```
*A:PE-2# show service id 2 pim-snooping group detail

===============================================================================
PIM Snooping Source Group ipv4
===============================================================================
Group Address      : 232.1.1.1
Source Address     : 192.168.55.2
Up Time            : 0d 00:00:46

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:21
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

RPF Neighbor       : 172.16.0.5
Incoming Intf      : EVPN-MPLS
Outgoing Intf List : EVPN-MPLS, SPOKE_SDP:26:2

Forwarded Packets  : 38275              Forwarded Octets   : 57259400
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

On PE-3, no PIM join messages are received and no groups are listed, as follows:

```
*A:PE-3# show service id 2 pim-snooping group detail
```

```
===============================================================================
PIM Snooping Source Group ipv4
===============================================================================
No Matching Entries
===============================================================================
```

On PE-4, the same PIM snooping information is available, because of the data-driven PIM state synchronization between PE-2 and PE-4 in ESI-24, as follows. The incoming interface is the EVPN-MPLS interface and the outgoing interface is spoke-SDP 46:2.

```
*A:PE-4# show service id 2 pim-snooping group detail

===============================================================================
PIM Snooping Source Group ipv4
===============================================================================
Group Address      : 232.1.1.1
Source Address     : 192.168.55.2
Up Time            : 0d 00:00:50

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:23
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

RPF Neighbor       : 172.16.0.5
Incoming Intf      : EVPN-MPLS
Outgoing Intf List : EVPN-MPLS, SPOKE_SDP:46:2

Forwarded Packets  : 41792              Forwarded Octets   : 62520832
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

PIM state synchronization is data-driven, so the PIM states are not stored in a database. Therefore, the ESs must be configured as **non-revertive** to avoid reverting back to the preferred PE while this PE is unaware of the PIM states.

PIM Snooping with Multi-chassis Synchronization

Data-driven PIM state synchronization is supported in SR OS Release 15.0.R4, and later. The ES must be configured as non-revertive, so that after a failover, the new DF remains the DF even when the original DF is operational again. When data-driven PIM state synchronization cannot be used, for example, when the service carving is configured in auto mode, or when the SR OS Release is an earlier release of 15.0, Multi-chassis synchronization (MCS) can be configured for a faster failover. MCS of the PIM snooping state on SAPs and spoke-SDPs is supported between an active and a standby PE and the PIM states are stored in a synchronization database. This can be configured in case of single-active multi-homing (MH), for example on PE-2 for peer PE-4, with PIM snooping on spoke-SDPs, as follows:

```
On PE-2:
configure
    redundancy
        multi-chassis
            peer 192.0.2.4 create
                sync
                    pim-snooping spoke-sdps
                    sdp 26 create
                        range 2-2 sync-tag "syncSA"
                    exit
                    no shutdown
                exit
                no shutdown
            exit all
```

On PE-4, MCS is configured for peer PE-2, as follows:

```
On PE-4:
configure
    redundancy
        multi-chassis
            peer 192.0.2.2 create
                sync
                    pim-snooping spoke-sdps
                    sdp 46 create
                        range 2-2 sync-tag "syncSA"
                    exit
                    no shutdown
                exit
                no shutdown
            exit all
```

When H-8 joins the multicast group, the following entries are in the MCS synchronization database of the PEs. The MCS sync-database on PE-2 shows the PIM snooping entries on the spoke-SDP 26:2 of the single-active MH ESI-24, as follows:

```
*A:PE-2# tools dump redundancy multi-chassis sync-database detail

If no entries are present for an application, no detail will be displayed.

FLAGS LEGEND: ld - local delete; da - delete alarm; pd - pending global delete;
              oal - omcr alarmed; ost - omcr standby

Peer Ip 192.0.2.4


Application pim-snooping-sdp
Sdp-id             Client Key
 SyncTag                          DLen  Flags            timeStamp
  deleteReason code and description                  #ShRec
-------------------------------------------------------------------------------
26:2               Adj 172.16.0.8
 syncSA                           72    -- -- -- --- --- 08/08/2023 09:57:40
  0x0                                                   0
26:2               IfSG SG 192.168.55.2 232.1.1.1
 syncSA                           69    -- -- -- --- --- 08/08/2023 09:57:51
  0x0                                                   0

The following totals are for:
 peer ip ALL, port/lag/sdp ALL, sync-tag ALL, application ALL
Valid Entries:                    2
Locally Deleted Entries:          0
Locally Deleted Alarmed Entries: 0
Pending Global Delete Entries:   0
Omcr Alarmed Entries:            0
Omcr Standby Entries:            0
Associated Shared Records (ALL): 0
Associated Shared Records (LD):  0
```

The MCS sync-database on PE-4 is similar, with SDP ID 46:2 instead of 26:2.

On PE-4, the MFIB is populated as follows:

```
*A:PE-4# show service id 2 mfib
```

```
================================================================================
Multicast FIB, Service 2
================================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
--------------------------------------------------------------------------------
192.168.55.2    232.1.1.1            sdp:46:2                     Local    Fwd
                                     mpls:192.0.2.1:524280        Local    Fwd
                                     mpls:192.0.2.2:524279        Local    Fwd
                                     mpls:192.0.2.3:524279        Local    Fwd
--------------------------------------------------------------------------------
Number of entries: 1
================================================================================
```

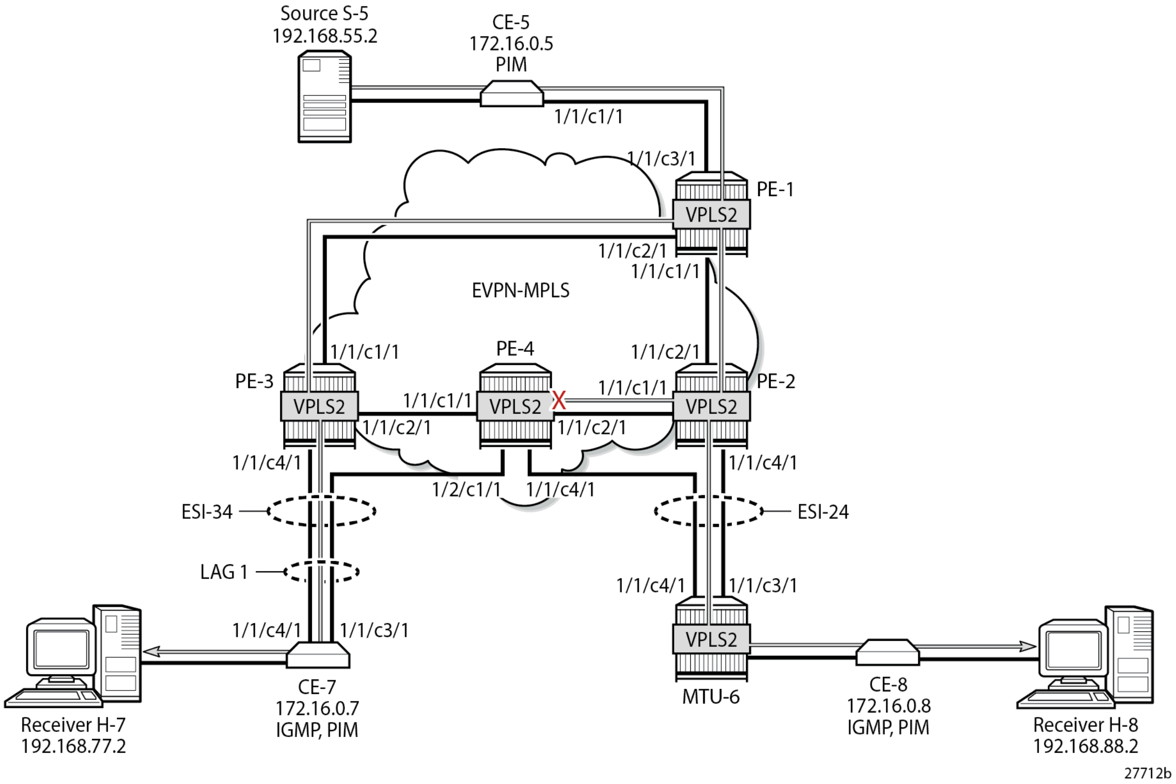
The PIM snooping group information on PE-4 shows the EVPN-MPLS as incoming interface and the spoke-SDP as outgoing interface, as follows. The split-horizon mechanism does not allow forwarding traffic from the EVPN-MPLS back to the EVPN-MPLS.

```
*A:PE-4# show service id 2 pim-snooping group detail

================================================================================
PIM Snooping Source Group ipv4
================================================================================
Group Address      : 232.1.1.1
Source Address     : 192.168.55.2
Up Time            : 0d 00:03:33

Up JP State        : Joined              Up JP Expiry       : 0d 00:00:55
Up JP Rpt          : Not Joined StarG    Up JP Rpt Override : 0d 00:00:00

RPF Neighbor       : 172.16.0.5
Incoming Intf      : EVPN-MPLS
Outgoing Intf List : EVPN-MPLS, SPOKE_SDP:46:2

Forwarded Packets  : 102854              Forwarded Octets   : 153869584
--------------------------------------------------------------------------------
Groups : 1
================================================================================
```

Failover

Figure 270: EVPN-MPLS with Multi-homing and PIM Snooping - Receivers H-7 and H-8 Joined shows the multicast traffic flow in the case where both receivers H-7 and H-8 joined multicast group 232.1.1.1 from source 192.168.55.2 and no failures have occurred. For SR OS Release 15.0.R4, and later, MCS need not be configured for faster failover in single-active MH when the ES is non-revertive.

Figure 270: EVPN-MPLS with Multi-homing and PIM Snooping - Receivers H-7 and H-8 Joined



NDF PE-4 has an MFIB table with the required information for a fast failover, as follows:

```
*A:PE-4# show service id 2 mfib

===============================================================================
Multicast FIB, Service 2
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
192.168.55.2    232.1.1.1             sap:lag-1:2                  Local    Fwd
                                      sdp:46:2                     Local    Fwd
                                      mpls:192.0.2.1:524280        Local    Fwd
                                      mpls:192.0.2.2:524279        Local    Fwd
                                      mpls:192.0.2.3:524279        Local    Fwd
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

In SR OS Release 15.0.R4, and later, data-driven PIM state synchronization ensures that NDF PE-4 has the following PIM snooping information for group 232.1.1.1.

```
*A:PE-4# show service id 2 pim-snooping group detail

===============================================================================
PIM Snooping Source Group ipv4
===============================================================================
Group Address      : 232.1.1.1
Source Address     : 192.168.55.2
Up Time            : 0d 00:02:40

Up JP State        : Joined             Up JP Expiry       : 0d 00:01:12
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

RPF Neighbor       : 172.16.0.5
Incoming Intf      : EVPN-MPLS
Outgoing Intf List : EVPN-MPLS, SAP:lag-1:2, SPOKE_SDP:46:2

Forwarded Packets  : 131682             Forwarded Octets   : 196996272
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

The following failures are introduced to force a failover from PE-2 to PE-4 and from PE-3 to PE-4. On MTU-6, SDP 62 is disabled, as follows:

```
*A:MTU-6# configure service sdp 62 shutdown
```

On CE-7, port 1/1/c4/1 toward PE-3 is disabled, as follows:

```
*A:CE-7# configure port 1/1/c4/1 shutdown
```

Log 99 on PE-3 shows that the DF state in ESI-34_2 changed to false:

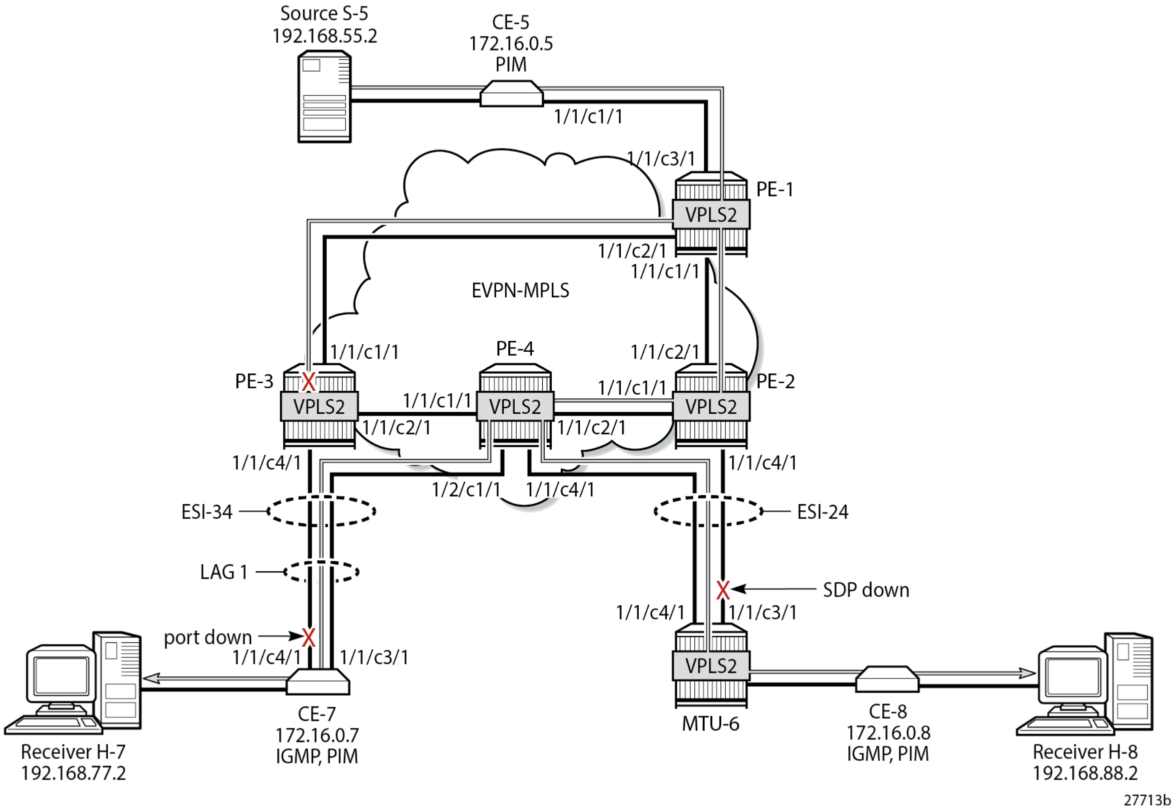
```
156 2023/08/08 09:51:51.762 UTC MINOR: SVCMGR #2094 Base
"Ethernet Segment:ESI-34_2, EVI:2, Designated Forwarding state changed to:false"
```

PE-4 becomes the DF for both ESs, as follows:

```
*A:PE-4# show service id 2 ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
lag-1:2               ESI-34_2                         DF
===============================================================================

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
46:2                  ESI-24                           DF
===============================================================================
No vxlan instance entries
```

Figure 271: EVPN-MPLS with Multi-homing and PIM Snooping - Multicast Flow after Failover shows the traffic flow after failover to new DF PE-4.

Figure 271: EVPN-MPLS with Multi-homing and PIM Snooping - Multicast Flow after Failover



PE-1 receives the multicast on port 1/1/c3/1 and forwards it on port 1/1/c1/1 to PE-2, and on port 1/1/c2/1 to PE-3, as follows:

```
*A:PE-1# show port 1/1/c1/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c1/1                                 20                            1990
                                      16472                        24976189
===============================================================================
*A:PE-1# show port 1/1/c2/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c2/1                                 17                            1758
                                      16470                        24976030
===============================================================================
*A:PE-1# show port 1/1/c3/1 statistics
```

```
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c3/1                              16452                        24743808
                                          1                              76
===============================================================================
```

PE-2 receives the multicast stream from PE-1 on port 1/1/c2/1 and forwards it to port 1/1/c1/1 to PE-4; it does not forward to port 1/1/c4/1 because SDP 26 is down, as follows:

```
*A:PE-2# show port 1/1/c1/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c1/1                                 22                            2263
                                      16476                        24985127
===============================================================================
*A:PE-2# show port 1/1/c2/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c2/1                              16477                        24985161
                                         22                            2241
===============================================================================
*A:PE-2# show port 1/1/c4/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c4/1                                 17                            1750
                                         16                            1691
===============================================================================
```

PE-4 receives the multicast traffic on port 1/1/c2/1 and forwards it on port 1/1/c4/1 toward MTU-6, and on port 1/2/c1/1 to CE-7, as follows:

```
*A:PE-4# show port 1/1/c1/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c1/1                                 17                            1977
                                         18                            2044
===============================================================================
*A:PE-4# show port 1/1/c2/1 statistics
```

```
===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c2/1                              16476                        24985255
                                         21                            2138
===============================================================================
*A:PE-4# show port 1/1/c4/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c4/1                                 15                            1611
                                      16474                        25116645
===============================================================================
*A:PE-4# show port 1/2/c1/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/2/c1/1                                 21                            2636
                                      16479                        24755468
===============================================================================
```

MTU-6 forwards the traffic to CE-8, which forwards it to H-8. CE-7 forwards the traffic to H-7. PE-3 drops the multicast traffic because LAG-1 is down because of the failure that was introduced at CE-7 (port disabled).

Conclusion

PIM snooping in EVPN-MPLS services results in a more efficient use of network resources because multicast traffic no longer needs to be flooded. PIM snooping can be used in EVPN-MPLS services with all-active and single-active multi-homing with data-driven PIM state synchronization. Alternatively, MCS synchronization of the PIM snooping state on SAPs and spoke-SDPs is supported with single-active MH.

PIM Snooping for IPv4 in PBB-EVPN Services

This chapter describes PIM Snooping for IPv4 in PBB-EVPN Services.

Topics in this chapter include:

Applicability

This chapter was initially written based on SR OS Release 15.0.R5, but the CLI in the current edition corresponds to SR OS Release 23.7.R1. Protocol Independent Multicast (PIM) snooping for IPv4 is supported in Provider Backbone Bridging - Ethernet Virtual Private Network (PBB-EVPN) services in SR OS Release 15.0.R1, and later. PIM snooping in single-active multi-homing (MH) mode without Ethernet Segment Identifier (ESI) label is supported in SR OS Release 15.0.R1, and later, whereas PIM snooping in single-active MH mode with ESI label is supported in SR OS Release 15.0.R4, and later. PIM snooping for IPv4 in all-active MH mode is supported in SR OS Release 15.0.R4, and later. Data-driven PIM state synchronization is supported in SR OS Release 15.0.R4, and later.

Overview

PBB-EVPN services have EVPN-MPLS enabled in the B-VPLS. PIM snooping in PBB-EVPN I-VPLS provides the following:

- PIM snooping in SAPs and SDP-bindings: PIM messages received from SAPs, SDP-bindings, or the B-VPLS are forwarded to SAPs or SDP-bindings according to the PIM snooping.
- Multicast flooding between I-VPLS and B-VPLS is the same for a PBB-EVPN B-VPLS as for a B-VPLS without EVPN. The first PIM join message received over the local B-VPLS from a B-VPLS SAP/SDP-binding or EVPN endpoint results in adding the B-VPLS SAP/SDP-binding or EVPN interface into the Multicast Forwarding Information Base (MFIB) associated with the I-VPLS context. Multicast traffic is flooded throughout the B-VPLS on a per-ISID single tree.
- When the PIM router is connected to a remote I-VPLS instance over the B-VPLS infrastructure, its location is identified by the B-VPLS SAP/SDP-binding or by the set of all EVPN endpoints on which PIM hellos are received. The location is also identified by the source BMAC address in the PBB header for the PIM hello message, which is the BMAC address associated with the B-VPLS instance on the remote PBB PE.
- The set of all EVPN endpoints in the B-VPLS is treated as a single PIM interface.
    - Hello and join/prune messages from I-VPLS SAPs/SDP-bindings are always sent to all B-VPLS PBB-EVPN destinations.

  - When a hello message is received from one B-VPLS PBB-EVPN destination PIM neighbor, the single interface representing all B-VPLS PBB-EVPN destinations will have that PIM neighbor.
  - All individual B-VPLS PBB-EVPN destinations appear in the MFIB, but the information for each B-VPLS PBB-EVPN destination entry is identical.
  - The EVPN split-horizon logic ensures that IP multicast traffic and PIM messages received on a PBB-EVPN endpoint are not forwarded back to other PBB-EVPN endpoints.
- When a point-to-multipoint (P2MP) mLDP provider tunnel is configured in the B-VPLS, the provider tunnel only works for the default multicast list. Ingress Replication (IR) is used for the per-ISID MFIB trees. ISID policies can be configured to specify ISID ranges that will use the default multicast list. ISID policies can help reduce the per-ISID MFIB resources used.
- PIM snooping for IPv4 within a PBB-EVPN I-VPLS is supported with single-active MH and with all-active MH in the associated I-VPLS.
- Data-driven PIM state synchronization between remote peers in an all-active MH Ethernet Segment (ES) is supported.
- Multi-Chassis Synchronization (MCS) of PIM snooping state on SAPs and spoke-SDPs is supported in active/standby scenarios.

The following command enables PIM snooping in an I-VPLS:

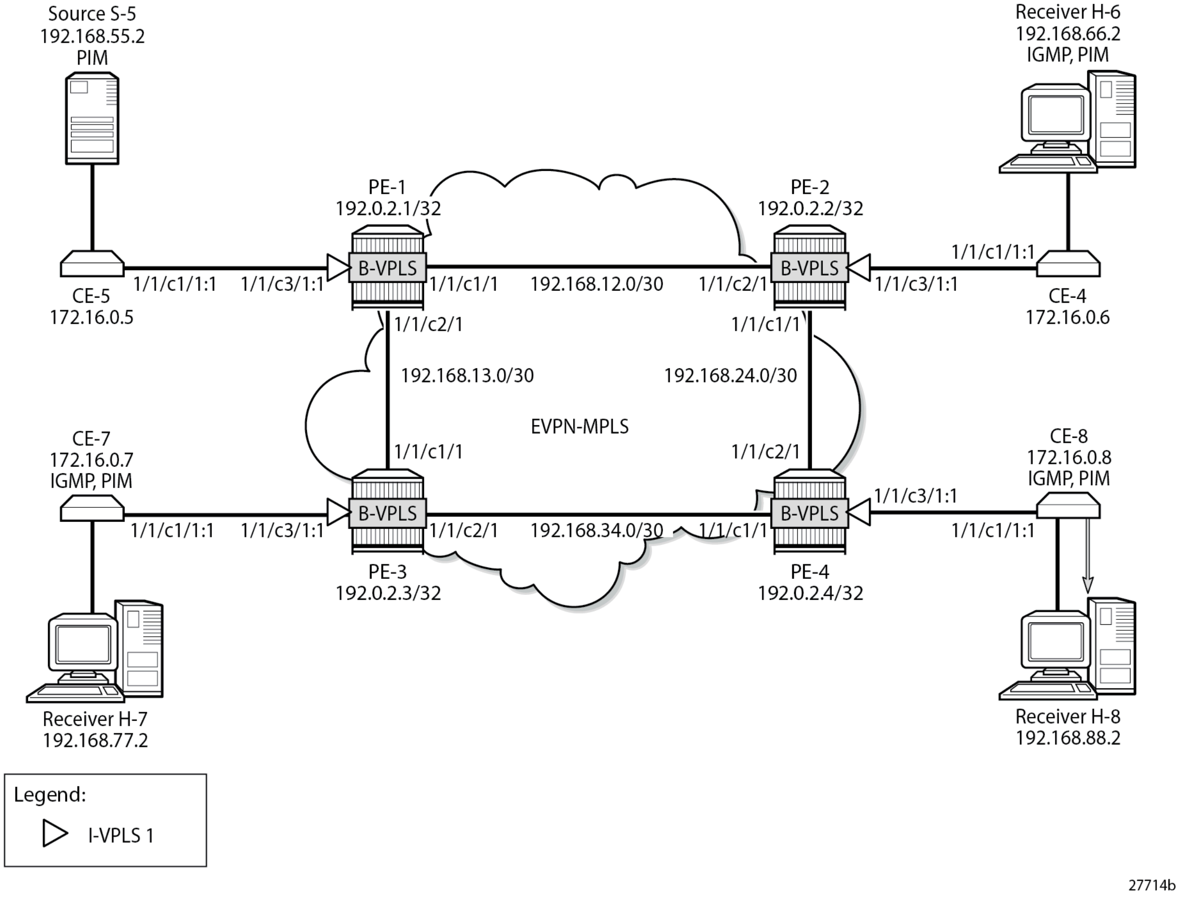
```
configure service vpls 1 pim-snooping
```

The default PIM snooping mode is proxy mode, which implies that the PE will terminate the PIM join/prune messages and generate its own PIM join/prune messages with the same (S,G). The advantage is that the number of PIM messages to be sent can be reduced: regardless of the number of PIM join messages received for a certain (S,G), the node only needs to send one PIM join message toward the source. PIM snooping can also use snooping mode based on the information in the received PIM hello messages; in snooping mode, the PE does not modify the PIM messages.

Configuration

Figure 272: Example Topology for PBB-EVPN without MH shows the example topology with source S-5 and receivers H-6, H-7, and H-8 attached to CEs that are connected to PEs. On the PEs, B-VPLS 100 is configured and I-VPLS 1 is associated with it. B-VPLS 100 has EVPN-MPLS enabled. An mLDP P2MP provider tunnel is used to distribute multicast traffic from PE-1 to the other PEs.

Figure 272: Example Topology for PBB-EVPN without MH



The initial configuration includes:

- Cards, MDAs, ports
- Router interfaces
- IS-IS enabled on the PEs (alternatively, OSPF can be used)
- LDP enabled on the PEs

BGP is configured on the PEs with address family EVPN, and PE-2 is configured as route reflector (RR). The BGP configuration on PE-2 is as follows:

```
On PE-2:
configure
    router "Base"
        bgp
            rapid-withdrawal
            rapid-update evpn
            group INTERNAL
                family evpn
                type internal
                cluster 192.0.2.2
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
            exit
        exit all
```

PBB-EVPN without MH – No PIM Snooping

B-VPLS 100 is configured with EVPN-MPLS enabled on all PEs. Multicast LDP is configured in B-VPLS 100 with PE-1 as the P2MP tunnel root node (**root-and-leaf**) and the other PEs as leaf nodes (**no root-and-leaf** is default). An (optional) ISID policy defines that the default multicast tree -which is used by the P2MP mLDP tunnel- is used for ISIDs 1 and 2 (range 1 to 2). The configuration of B-VPLS 100 on PE-1 is as follows:

```
On PE-1:
configure
    service
        vpls 100 name "B-VPLS 100" customer 1 b-vpls create
            description "B-VPLS 100"
            service-mtu 2000
            pbb
                source-bmac 00:00:00:00:00:01
            exit
            split-horizon-group "CORE" create
            exit
            bgp
            exit
            bgp-evpn
                evi 100
                mpls
                    split-horizon-group "CORE"
                    ingress-replication-bum-label
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            provider-tunnel
                inclusive
                    owner bgp-evpn-mpls
                    root-and-leaf
                    mldp
                    no shutdown
                exit
            exit
            isid-policy
                entry 1 create
                    use-def-mcast
                    no advertise-local
                    range 1 to 2
                exit
            exit
            no shutdown
        exit all
```

The configuration of B-VPLS on the other PEs is similar, but without the root-and-leaf option.

In B-VPLS 100 on root node PE-1, the following mLDP provider tunnel is created with Provider Multicast Service Interface (PMSI) owner bgpEvpnMpls. PE-1 is configured as root-and-leaf node.

```
*A:PE-1# show service id 100 provider-tunnel

===============================================================================
Service Provider Tunnel Information
===============================================================================
Type               : inclusive          Root and Leaf      : enabled
Admin State        : enabled            Data Delay Intvl   : 15 secs
PMSI Type          : ldp                LSP Template       :
Remain Delay Intvl : 0 secs             LSP Name used      : 8193
PMSI Owner         : bgpEvpnMpls
Oper State         : up                 Root Bind Id       : 32767
-------------------------------------------------------------------------------
Type               : selective          Wildcard SPMSI     : disabled
Admin State        : disabled           Data Delay Intvl   : 3 secs
PMSI Type          : none               Max P2MP SPMSI     : 10
PMSI Owner         : none
===============================================================================
```

When the B-VPLS is created, I-VPLS 1 can be associated with it, as follows:

```
On PE-1:
configure
    service
        vpls 1 name "I-VPLS 1" customer 1 i-vpls create
            pbb
                backbone-vpls 100
                exit
            exit
            sap 1/1/c3/1:1 create
            exit
            no shutdown
        exit all
```

The configuration of I-VPLS 1 on the other PEs is identical.

CE-6, CE-7, and CE-8 have IGMP enabled on the interface toward the receiver and PIM enabled on all interfaces. Source-specific multicast is used in this example. The configuration on CE-8 is as follows:

```
On CE-8:
configure
    router "Base"
        interface "int-CE-8-H-8"
            address 192.168.88.1/24
            port 1/1/c2/1
        exit
        interface "int-CE-8-PE-4"
            address 172.16.0.8/16
            port 1/1/c1/1:1
        exit
        interface "system"
            address 192.0.2.8/32
        exit
        static-route-entry 192.168.55.0/30
            next-hop 172.16.0.5
                no shutdown
            exit
        exit
        igmp
            interface int-CE-8-H-8
            exit
        exit
        pim
            apply-to all
        exit
    exit all
```

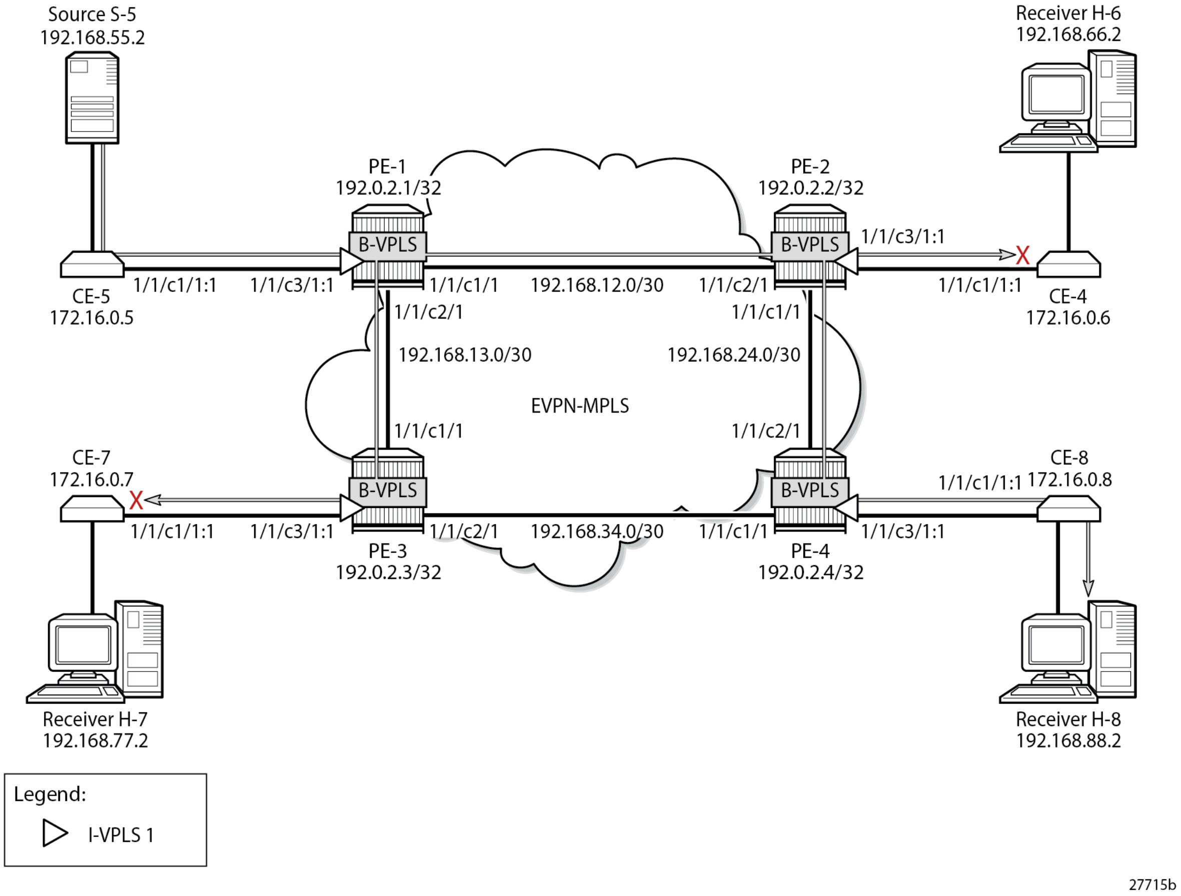
The static route is required on the receiving CEs for the PIM join/prune messages to reach the multicast source S-5 with IP address 192.168.55.2; only IP subnet 172.16.0.0/16 can be reached via the VPLS.

CE-5 has PIM enabled and static routes configured to reach the receiving hosts, as follows:

```
On CE-5:
configure
    router "Base"
        interface "int-CE-5-PE-1"
            address 172.16.0.5/16
            port 1/1/c1/1:1
        exit
        interface "int-CE-5-S-5"
            address 192.168.55.1/30
            port 1/1/c3/1
        exit
        interface "system"
            address 192.0.2.5/32
        exit
        static-route-entry 192.168.66.0/24
            next-hop 172.16.0.6
                no shutdown
            exit
        exit
        static-route-entry 192.168.77.0/24
            next-hop 172.16.0.7
                no shutdown
            exit
        exit
        static-route-entry 192.168.88.0/24
            next-hop 172.16.0.8
                no shutdown
            exit
        exit
        pim
            apply-to all
        exit
    exit all
```

When receiver H-8 sends an IGMP report to join multicast group (S,G), CE-8 sends a PIM join message to CE-5. This PIM join message is flooded by the PEs. When CE-5 receives the PIM join message, it forwards the multicast stream to receiver H-8. PIM snooping is disabled by default and the MFIB on each of the PEs remains empty, so the multicast stream is not only sent to CE-8, but also to CE-6 and CE-7. CE-6 and CE-7 drop this stream when no receiver is active, while CE-8 forwards the multicast stream to receiver H-8, as shown in Figure 273: Multicast Stream to Receiver H-8 with PIM Snooping Disabled.

Figure 273: Multicast Stream to Receiver H-8 with PIM Snooping Disabled



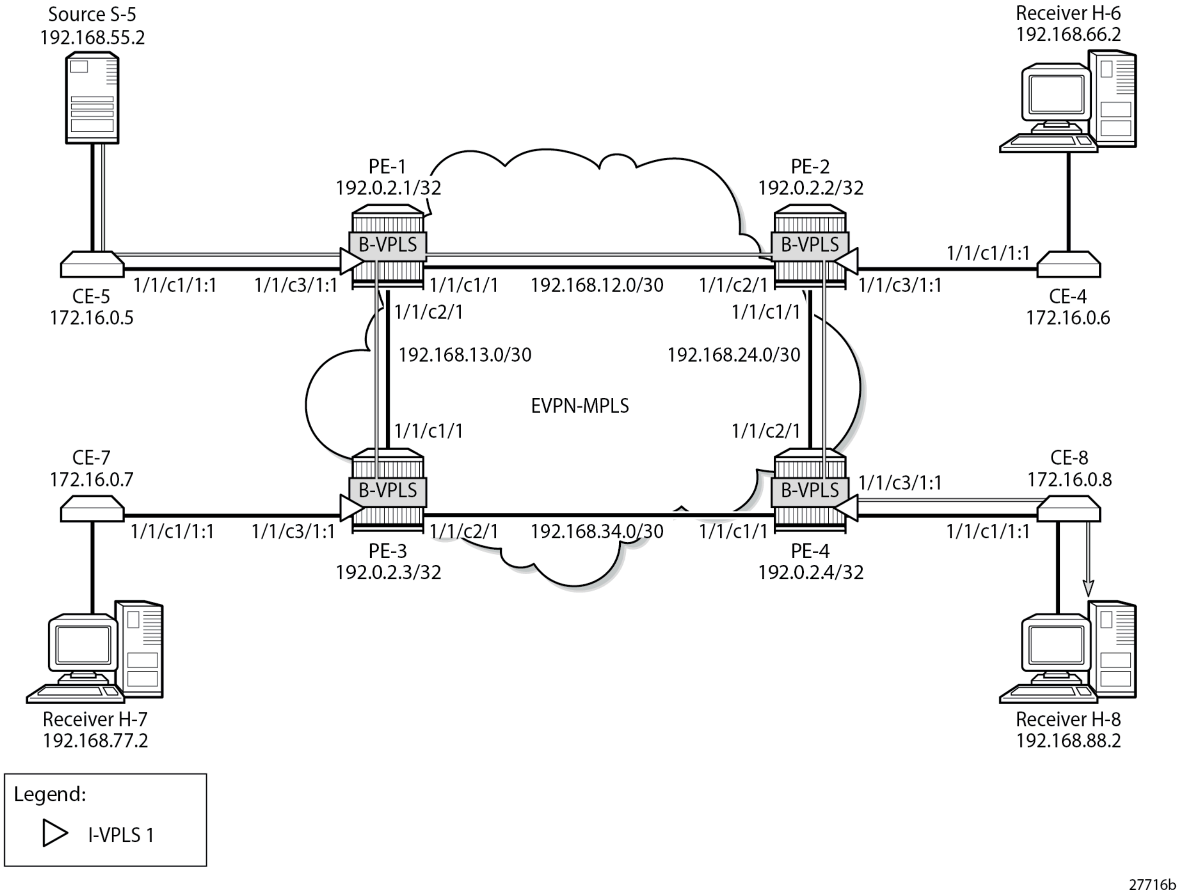
PBB-EVPN without MH – PIM Snooping for IPv4 Enabled

PIM snooping for IPv4 is enabled in I-VPLS 1 on all PEs as follows:

```
On all PEs:
configure service vpls 1 pim-snooping
```

When PIM snooping for IPv4 is enabled, the PEs only forward the multicast traffic to those CEs that have sent PIM join messages for that multicast group. This implies that PE-2 and PE-3 do not forward traffic to the CEs; only PE-4 forwards traffic toward CE-8 and CE-8 forwards to receiver H-8, as shown in Figure 274: Multicast Stream to Receiver H-8 with PIM Snooping Enabled.

Figure 274: Multicast Stream to Receiver H-8 with PIM Snooping Enabled



When PIM snooping for IPv4 is enabled, PE-1 has the following two PIM snooping ports: the SAP toward the source and the backbone b-EVPN-MPLS interface, which is treated as one entity for all PBB-EVPN destinations.

```
*A:PE-1# show service id 1 pim-snooping port

===============================================================================
PIM Snooping Port ipv4
===============================================================================
Port Id                                                       Opr   PW Fwding
-------------------------------------------------------------------------------
SAP:1/1/c3/1:1                                                Up    Actv
b-EVPN-MPLS                                                   Up    Actv
===============================================================================
```

PE-1 has the following PIM snooping neighbors: CE-5 with IP address 172.16.0.5 is attached via SAP 1/1/c3/1:1, and the other CEs are attached to the b-EVPN-MPLS. Even though this b-EVPN-MPLS is treated as one entity, individual entries are shown for each B-VPLS PBB-EVPN destination, as follows:

```
*A:PE-1# show service id 1 pim-snooping neighbor
```

```
===============================================================================
PIM Snooping Neighbors ipv4
===============================================================================
Port Id                 Nbr DR Prty     Up Time       Expiry Time     Hold Time
  Nbr Address
-------------------------------------------------------------------------------
SAP:1/1/c3/1:1          1               0d 00:01:21   0d 00:01:24     105
  172.16.0.5
b-EVPN-MPLS             1               0d 00:01:30   0d 00:01:45     105
  172.16.0.6
b-EVPN-MPLS             1               0d 00:01:18   0d 00:01:27     105
  172.16.0.7
b-EVPN-MPLS             1               0d 00:01:34   0d 00:01:41     105
  172.16.0.8
-------------------------------------------------------------------------------
Neighbors : 4
===============================================================================
```

Receiver H-8 joins the multicast stream and the PIM group with group address 232.1.1.1, and source address 192.168.55.2 is shown on CE-8 with incoming interface toward PE-4 and outgoing interface toward H-8. The Reverse Path Forwarding (RPF) neighbor is CE-5 with IP address 172.16.0.5, as follows:

```
*A:CE-8# show router pim group detail

===============================================================================
PIM Source Group ipv4
===============================================================================
Group Address      : 232.1.1.1
Source Address     : 192.168.55.2
RP Address         : 0
Advt Router        :
Flags              :                    Type               : (S,G)
Mode               : sparse
MRIB Next Hop      : 172.16.0.5
MRIB Src Flags     : remote
Keepalive Timer    : Not Running
Up Time            : 0d 00:08:19        Resolved By        : rtable-u

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:41
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 172.16.0.5
Incoming Intf      : int-CE-8-PE-4
Outgoing Intf List : int-CE-8-H-8

Curr Fwding Rate   : 9751.560 kbps
Forwarded Packets  : 75064              Discarded Packets  : 0
Forwarded Octets   : 111244848          RPF Mismatches     : 0
Spt threshold      : 0 kbps             ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

With PIM snooping for IPv4 enabled, and after receiving a PIM join message for multicast (192.168.55.2, 232.1.1.1), the MFIB on PE-1 has an entry for group address 232.1.1.1 and source address 192.168.55.2, as follows. The local SAP connects to CE-5; the other port IDs correspond to the b-EVPN-MPLS interface.

```
*A:PE-1# show service id 1 mfib
```

```
===============================================================================
Multicast FIB, Service 1
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
192.168.55.2    232.1.1.1             sap:1/1/c3/1:1               Local    Fwd
                                      b-mpls:192.0.2.2:524282      100      Fwd
                                      b-mpls:192.0.2.3:524282      100      Fwd
                                      b-mpls:192.0.2.4:524282      100      Fwd
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

The MFIB on PE-4 is similar, with a local SAP connecting to CE-8 and three b-eMpls port IDs for each of the PE peers. In contrast, the MFIBs on PE-2 and PE-3 are empty, because no multicast traffic needs to be forwarded to the attached CEs, as follows:

```
*A:PE-2# show service id 1 mfib

===============================================================================
Multicast FIB, Service 1
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Number of entries: 0
===============================================================================
```

The following MFIB statistics on PE-1 show the number of matched packets and matched octets for group address 232.1.1.1 and source address 192.168.55.2:

```
*A:PE-1# show service id 1 mfib statistics

===============================================================================
Multicast FIB Statistics, Service 1
===============================================================================
Source Address  Group Address       Matched Pkts          Matched Octets
                                                          Forwarding Rate
-------------------------------------------------------------------------------
192.168.55.2    232.1.1.1           60968                 91452000
                                                          398.559 kbps
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

The following shows that PE-2 receives the multicast packets on port 1/1/c2/1 and forwards them to PE-4 on port 1/1/c1/1. With PIM snooping for IPv4 enabled, PE-2 does not forward the traffic to CE-6 on port 1/1/c3/1 because no PIM join message was received from CE-6. Besides the multicast traffic, some signaling messages (such as PIM, IS-IS, and so on) are sent on the ports, which explains why all counters have non-zero values.

```
*A:PE-2# show port 1/1/c1/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
```

```
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c1/1                                 18                            1796
                                      16474                        25275378
===============================================================================

*A:PE-2# show port 1/1/c2/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c2/1                              16477                        25275662
                                         23                            2299
===============================================================================

*A:PE-2# show port 1/1/c3/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c3/1                                  1                              76
                                          2                             152
===============================================================================
```

The following PIM snooping group with group address 232.1.1.1 and source address 192.168.55.2 is shown on PE-1. The incoming interface is the SAP toward CE-5 and the outgoing interface is the b-EVPN-MPLS interface. A single b-EVPN-MPLS interface is shown in the outgoing interface list, regardless of the B-VPLS PBB-EVPN destination. The split-horizon mechanism ensures that all traffic from the incoming interface SAP 1/1/c3/1:1 is only forwarded on the b-EVPN-MPLS interface, not sent back on the SAP.

```
*A:PE-1# show service id 1 pim-snooping group detail

===============================================================================
PIM Snooping Source Group ipv4
===============================================================================
Group Address      : 232.1.1.1
Source Address     : 192.168.55.2
Up Time            : 0d 00:01:14

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:46
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

RPF Neighbor       : 172.16.0.5
Incoming Intf      : SAP:1/1/c3/1:1
Outgoing Intf List : b-EVPN-MPLS, SAP:1/1/c3/1:1

Forwarded Packets  : 60968              Forwarded Octets   : 91452000
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

On PE-2 and PE-3, there are no PIM snooping groups.

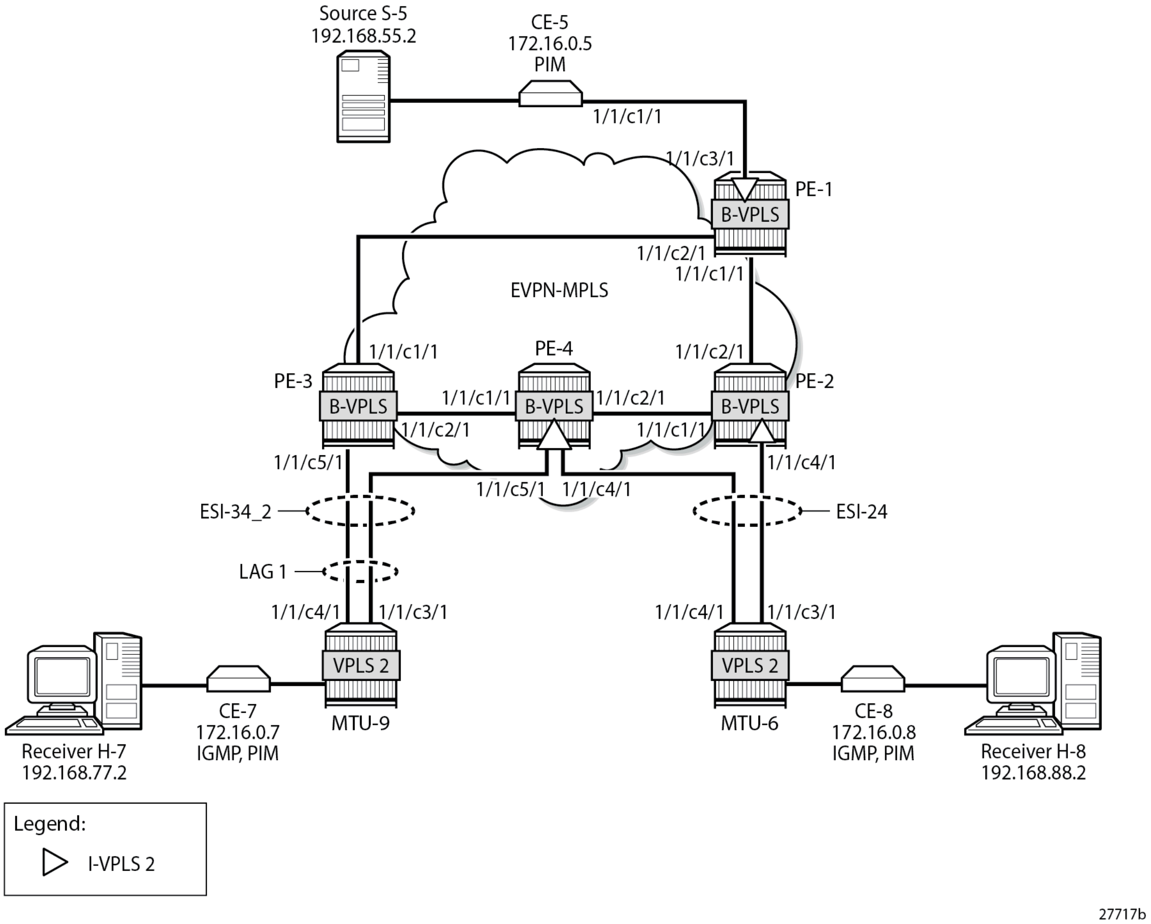
On PE-4, the PIM snooping group with group address 232.1.1.1 and source address 192.168.55.2 has the b-EVPN-MPLS interface as incoming interface and SAP 1/1/c3/1:1 toward CE-8 as outgoing interface, as follows. The split-horizon mechanism ensures that traffic received from the b-EVPN-MPLS interface is not forwarded on the b-EVPN-MPLS interface to the other PEs, so it is only forwarded on the SAP 1/1/c3/1:1 toward CE-8.

```
*A:PE-4# show service id 1 pim-snooping group 232.1.1.1 detail

===============================================================================
PIM Snooping Source Group ipv4
===============================================================================
Group Address      : 232.1.1.1
Source Address     : 192.168.55.2
Up Time            : 0d 00:01:21

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:38
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

RPF Neighbor       : 172.16.0.5
Incoming Intf      : b-EVPN-MPLS
Outgoing Intf List : b-EVPN-MPLS, SAP:1/1/c3/1:1

Forwarded Packets  : 66623              Forwarded Octets   : 100867222
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

PBB-EVPN with MH – No PIM Snooping

Figure 275: Example Topology for PBB-EVPN with MH shows the example topology with CE-7 attached to MTU-9, which is connected to both PE-3 and PE-4 via LAG 1. Virtual ES (vES) ESI-34_2 is configured in all-active MH mode using LAG 1 for dot1q value 2. MTU-6 is connected to PE-2 and PE-4 with SDPs. These SDPs are associated with a single-active MH ES ESI-24.

Figure 275: Example Topology for PBB-EVPN with MH



The configuration of I-VPLS 2 is similar to the preceding configuration of I-VPLS 1 on all PEs.

On PE-2, PE-3, and PE-4, one or more ESs are configured. The service configuration on PE-2 is as follows. An SDP is configured toward MTU-6 that is associated with a single-active MH ES ESI-24, that is non-restrictive -after failover, it does not restore to the initial designated forwarder (DF) if available again; see chapter Preference-based and Non-revertive EVPN DF Election. The manually configured preference is 200 on PE-2, which is higher than preference 50 at PE-3, so PE-2 is the DF when no failover has occurred. Spoke-SDP 26:2 is associated with I-VPLS 2. The B-VPLS 100 remains unchanged and is not repeated here.

```
On PE-2:
configure
    service
        sdp 26 mpls create
            far-end 192.0.2.6
            ldp
            no shutdown
        exit
        system
            bgp-evpn
                ethernet-segment "ESI-24" create
                    esi 01:00:00:00:00:24:00:00:00:01
                    source-bmac-lsb 24-02 es-bmac-table-size 8
                    es-activation-timer 3
                    service-carving
                        mode manual
                        manual
                            preference non-revertive create
                                value 200
                            exit
                        exit
                    exit
                    multi-homing single-active
                    sdp 26
                    no shutdown
                exit
            exit
        exit
        vpls 2 name "I-VPLS 2" customer 1 i-vpls create
            pbb
                backbone-vpls 100
                exit
            exit
            spoke-sdp 26:2 create
            exit
            no shutdown
        exit
    exit all
```

On PE-4, LAG 1 is configured in access mode with dot1q encapsulation on the port to MTU-9, as follows. The LAG configuration is similar on PE-3.

```
On PE-4:
configure
    lag 1
        mode access
        encap-type dot1q
        port 1/1/c5/1
        lacp active administrative-key 1 system-id 00:00:00:00:01:34
        no shutdown
    exit all
```

Single-active ES ESI-24 is configured on PE-4, together with a virtual ES ESI-34_2, which is an all-active MH virtual ES that applies to LAG 1 for I-VPLS 2 only (**q-tag-range 2**); see chapter Virtual Ethernet Segments. The preference for the DF election is configured manually to a value of 50 (which is lower than preference 200 on the remote peer in the ES). I-VPLS 2 has a SAP and a spoke-SDP configured. The service configuration on PE-4 is as follows:

```
On PE-4:
configure
    service
        sdp 46 mpls create
            far-end 192.0.2.6
            ldp
            no shutdown
        exit
        system
            bgp-evpn
                ethernet-segment "ESI-24" create
                    esi 01:00:00:00:00:24:00:00:00:01
                    source-bmac-lsb 24-04 es-bmac-table-size 8
                    es-activation-timer 3
                    service-carving
```

```
                        mode manual
                        manual
                            preference non-revertive create
                                value 50
                            exit
                        exit
                    exit
                    multi-homing single-active
                    sdp 46
                    no shutdown
                exit
                ethernet-segment "ESI-34_2" virtual create
                    esi 01:00:00:00:00:34:02:00:00:01
                    source-bmac-lsb 34-34 es-bmac-table-size 8
                    es-activation-timer 3
                    service-carving
                        mode manual
                        manual
                            preference non-revertive create
                                value 50
                            exit
                        exit
                    exit
                    multi-homing all-active
                    lag 1
                    dot1q
                        q-tag-range 2
                    exit
                    no shutdown
                exit
            exit
        exit
        vpls 2 name "I-VPLS 2" customer 1 i-vpls create
            pbb
                backbone-vpls 100
                exit
            exit
            sap lag-1:2 create
            exit
            spoke-sdp 46:2 create
            exit
            no shutdown
        exit
        vpls 100 name "B-VPLS 100" customer 1 b-vpls create
            pbb
                use-es-bmac
            exit
        exit
    exit all
```

On PE-4, the source BMAC in the all-active MH ESI-34_2 is identical to the source BMAC on remote peer PE-3, but in the single-active MH ESI-24, the source BMAC must be different. Remote PEs might send traffic to the NDF PE based on the shared source BMAC, which is fine for all-active MH ESs, but not for single-active MH ESs.

The following service configuration on PE-3 includes an all-active virtual ES ESI-34_2 with preference 200, which makes PE-3 the DF for ESI-34_2 when no failover has occurred. After failover, PE-4 becomes DF, and it does not revert to PE-3 when available.

```
On PE-3:
configure
    service
```

```
        system
            bgp-evpn
                ethernet-segment "ESI-34_2" virtual create
                    esi 01:00:00:00:00:34:02:00:00:01
                    source-bmac-lsb 34-34 es-bmac-table-size 8
                    es-activation-timer 3
                    service-carving
                        mode manual
                        manual
                            preference non-revertive create
                                value 200
                            exit
                        exit
                    exit
                    multi-homing all-active
                    lag 1
                    dot1q
                        q-tag-range 2
                    exit
                    no shutdown
                exit
            exit
        exit
        vpls 2 name "I-VPLS 2" customer 1 i-vpls create
            pbb
                backbone-vpls 100
                exit
            exit
            sap lag-1:2 create
            exit
            no shutdown
        exit
        vpls 100 name "B-VPLS 100" customer 1 b-vpls create
            pbb
                use-es-bmac
            exit
        exit
    exit all
```

The following is the service configuration on MTU-6:

```
On MTU-6:
configure
    service
        sdp 62 mpls create
            far-end 192.0.2.2
            ldp
            no shutdown
        exit
        sdp 64 mpls create
            far-end 192.0.2.4
            ldp
            no shutdown
        exit
        vpls 2 name "VPLS 2" customer 1 create
            endpoint "x" create
            exit
            sap 1/2/c1/1:2 create
            exit
            spoke-sdp 62:2 endpoint "x" create
            exit
            spoke-sdp 64:2 endpoint "x" create
            exit
            no shutdown
        exit
    exit all
```

The following is the LAG configuration on MTU-9:

```
On MTU-9:
configure
    lag 1
        mode access
        encap-type dot1q
        port 1/1/c3/1
        port 1/1/c4/1
        lacp active administrative-key 32768
        no shutdown
    exit all
```

The configuration of VPLS 2 on MTU-9 is as follows:

```
On MTU-9:
configure
    service
        vpls 2 name "VPLS 2" customer 1 create
            sap lag-1:2 create
            exit
            sap 1/1/c1/1:2 create
            exit
            no shutdown
        exit all
```

For I-VPLS 2, PE-2 is the DF in ES ESI-24, as follows:

```
*A:PE-2# show service id 2 ethernet-segment
No sap entries

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
26:2                  ESI-24                           DF
===============================================================================
No vxlan instance entries
```

PE-3 is the DF in virtual ES ESI-34_2, as follows:

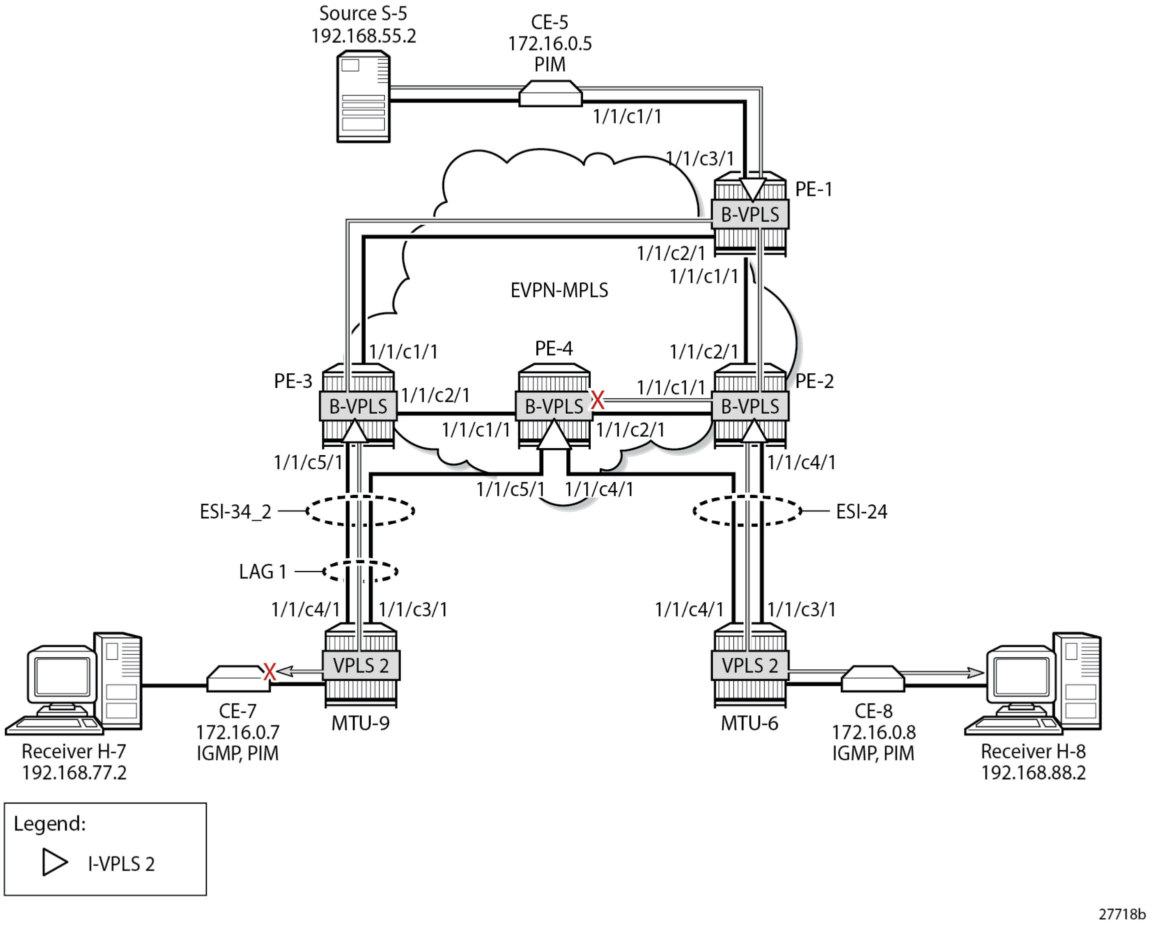
```
*A:PE-3# show service id 2 ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
lag-1:2               ESI-34_2                         DF
===============================================================================
No sdp entries
No vxlan instance entries
```

PE-4 is the Non-DF (NDF) for both ESI-24 and ESI-34_2, as follows:

```
*A:PE-4# show service id 2 ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
lag-1:2               ESI-34_2                         NDF
===============================================================================

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
46:2                  ESI-24                           NDF
===============================================================================
No vxlan instance entries
```

When H-8 sends an IGMP report to join multicast group 232.1.1.1 from source 192.168.55.2, CE-5 forwards the multicast stream after receiving the corresponding PIM join message. PE-1 forwards the multicast traffic on the P2MP mLDP tunnel to all EVPN-MPLS destinations: PE-2, PE-3, and PE-4. PE-2 is the DF for ESI-24 and forwards the traffic to MTU-6, which forwards it to CE-8, where it is sent to the attached receiver H-8 that joined the multicast group. PE-3 is the DF for ESI-34_2 and sends the multicast stream to MTU-9, which forwards it to CE-7, where it is dropped because no attached receiver has joined the multicast group. PE-4 is the NDF for both ESs, so it does not forward the traffic to MTU-6 or MTU-9. Figure 276: EVPN-MPLS with MH - PIM Snooping Disabled – Receiver H-8 Joined shows how this multicast is forwarded when PIM snooping is disabled.

Figure 276: EVPN-MPLS with MH - PIM Snooping Disabled – Receiver H-8 Joined



PBB-EVPN with MH – PIM Snooping for IPv4 Enabled

PIM snooping for IPv4 is enabled in I-VPLS 2 on all PEs with the following command:

```
On all PEs:
configure service vpls 2 pim-snooping
```

All PEs have three PIM snooping neighbors: CE-5, CE-7, and CE-8. The list of PIM snooping neighbors on PE-1 is as follows:

```
*A:PE-1# show service id 2 pim-snooping neighbor

===============================================================================
PIM Snooping Neighbors ipv4
===============================================================================
Port Id                 Nbr DR Prty     Up Time       Expiry Time     Hold Time
  Nbr Address
-------------------------------------------------------------------------------
SAP:1/1/c3/1:2          1               0d 00:00:59   0d 00:01:16     105
```

```
  172.16.0.5
b-EVPN-MPLS             1               0d 00:00:42   0d 00:01:33     105
  172.16.0.7
b-EVPN-MPLS             1               0d 00:01:08   0d 00:01:37     105
  172.16.0.8
-------------------------------------------------------------------------------
Neighbors : 3
===============================================================================
```

When H-7 and H-8 join the group 232.1.1.1 via source 192.168.55.2, the PIM join messages are snooped by the PEs and the MFIB is built. The MFIB on PE-1 contains one entry for group address 232.1.1.1 and source address 192.168.55.2 with four port IDs: the local SAP to CE-5 and the B-VPLS PBB-EVPN destinations, as follows:

```
*A:PE-1# show service id 2 mfib

===============================================================================
Multicast FIB, Service 2
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
192.168.55.2    232.1.1.1             sap:1/1/c3/1:2               Local    Fwd
                                      b-mpls:192.0.2.2:524282      100      Fwd
                                      b-mpls:192.0.2.3:524282      100      Fwd
                                      b-mpls:192.0.2.4:524282      100      Fwd
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

In a similar way, the other PEs that snooped PIM messages build their MFIBs. On PE-2, the following MFIB is shown when H-8 has joined the multicast group.

```
*A:PE-2# show service id 2 mfib

===============================================================================
Multicast FIB, Service 2
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
192.168.55.2    232.1.1.1             sdp:26:2                     Local    Fwd
                                      b-mpls:192.0.2.1:524282      100      Fwd
                                      b-mpls:192.0.2.3:524282      100      Fwd
                                      b-mpls:192.0.2.4:524282      100      Fwd
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

On PE-3, the following MFIB is present when H-7 has joined the multicast group:

```
*A:PE-3# show service id 2 mfib

===============================================================================
Multicast FIB, Service 2
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
192.168.55.2    232.1.1.1             sap:lag-1:2                  Local    Fwd
                                      b-mpls:192.0.2.1:524282      100      Fwd
                                      b-mpls:192.0.2.2:524282      100      Fwd
                                      b-mpls:192.0.2.4:524282      100      Fwd
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

Furthermore, data-driven PIM state synchronization between PEs in an all-active MH ES allows the NDF PE-4 to build its MFIB, even when the NDF does not forward multicast traffic to the receivers. When the NDF has the MFIB information, the failover is faster and the loss of traffic is limited. For data-driven PIM state synchronization, the source BMAC must be identical within the ES, so it only works for all-active MH in PBB-EVPN, not for single-active MH. The MFIB on PE-4 contains the SAP from the all-active MH ESI-34_2, but not the spoke-SDP from the single-active MH ESI-24, as follows:

```
*A:PE-4# show service id 2 mfib

===============================================================================
Multicast FIB, Service 2
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
192.168.55.2    232.1.1.1             sap:lag-1:2                  Local    Fwd
                                      b-mpls:192.0.2.1:524282      100      Fwd
                                      b-mpls:192.0.2.2:524282      100      Fwd
                                      b-mpls:192.0.2.3:524282      100      Fwd
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

The snooped PIM group information on PE-1 shows the SAP to CE-5 as incoming interface and the b-EVPN-MPLS interface as outgoing, as follows. The split-horizon mechanism prevents multicast traffic coming from the SAP to CE-5 from being returned.

```
*A:PE-1# show service id 2 pim-snooping group detail

===============================================================================
PIM Snooping Source Group ipv4
===============================================================================
Group Address      : 232.1.1.1
Source Address     : 192.168.55.2
Up Time            : 0d 00:02:19

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:41
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

RPF Neighbor       : 172.16.0.5
Incoming Intf      : SAP:1/1/c3/1:2
Outgoing Intf List : b-EVPN-MPLS, SAP:1/1/c3/1:2

Forwarded Packets  : 114411             Forwarded Octets   : 171616500
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

On PE-2, the incoming interface is the b-EVPN-MPLS interface and the outgoing interface is the spoke-SDP toward MTU-6, as follows:

```
*A:PE-2# show service id 2 pim-snooping group detail

===============================================================================
PIM Snooping Source Group ipv4
===============================================================================
Group Address      : 232.1.1.1
Source Address     : 192.168.55.2
Up Time            : 0d 00:00:49

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:59
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

RPF Neighbor       : 172.16.0.5
Incoming Intf      : b-EVPN-MPLS
Outgoing Intf List : b-EVPN-MPLS, SPOKE_SDP:26:2

Forwarded Packets  : 40909              Forwarded Octets   : 61936226
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

On PE-3, the incoming interface is the b-EVPN-MPLS interface and the outgoing interface is the SAP lag-1:2 toward MTU-9, as follows:

```
*A:PE-3# show service id 2 pim-snooping group detail

===============================================================================
PIM Snooping Source Group ipv4
===============================================================================
Group Address      : 232.1.1.1
Source Address     : 192.168.55.2
Up Time            : 0d 00:02:22

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:30
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

RPF Neighbor       : 172.16.0.5
Incoming Intf      : b-EVPN-MPLS
Outgoing Intf List : b-EVPN-MPLS, SAP:lag-1:2

Forwarded Packets  : 117006             Forwarded Octets   : 177147084
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

In case of all-active MH ES ESI-34_2, one of the PE -DF or NDF- in the ES forwards the PIM states to its remote peer and therefore, PE-4 has the same PIM snooping group information as PE-3, as follows:

```
*A:PE-4# show service id 2 pim-snooping group detail

===============================================================================
PIM Snooping Source Group ipv4
===============================================================================
Group Address      : 232.1.1.1
Source Address     : 192.168.55.2
Up Time            : 0d 00:02:23

Up JP State        : Joined             Up JP Expiry       : 0d 00:00:29
Up JP Rpt          : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

RPF Neighbor       : 172.16.0.5
Incoming Intf      : b-EVPN-MPLS
Outgoing Intf List : b-EVPN-MPLS, SAP:lag-1:2

Forwarded Packets  : 118048             Forwarded Octets   : 178724672
```

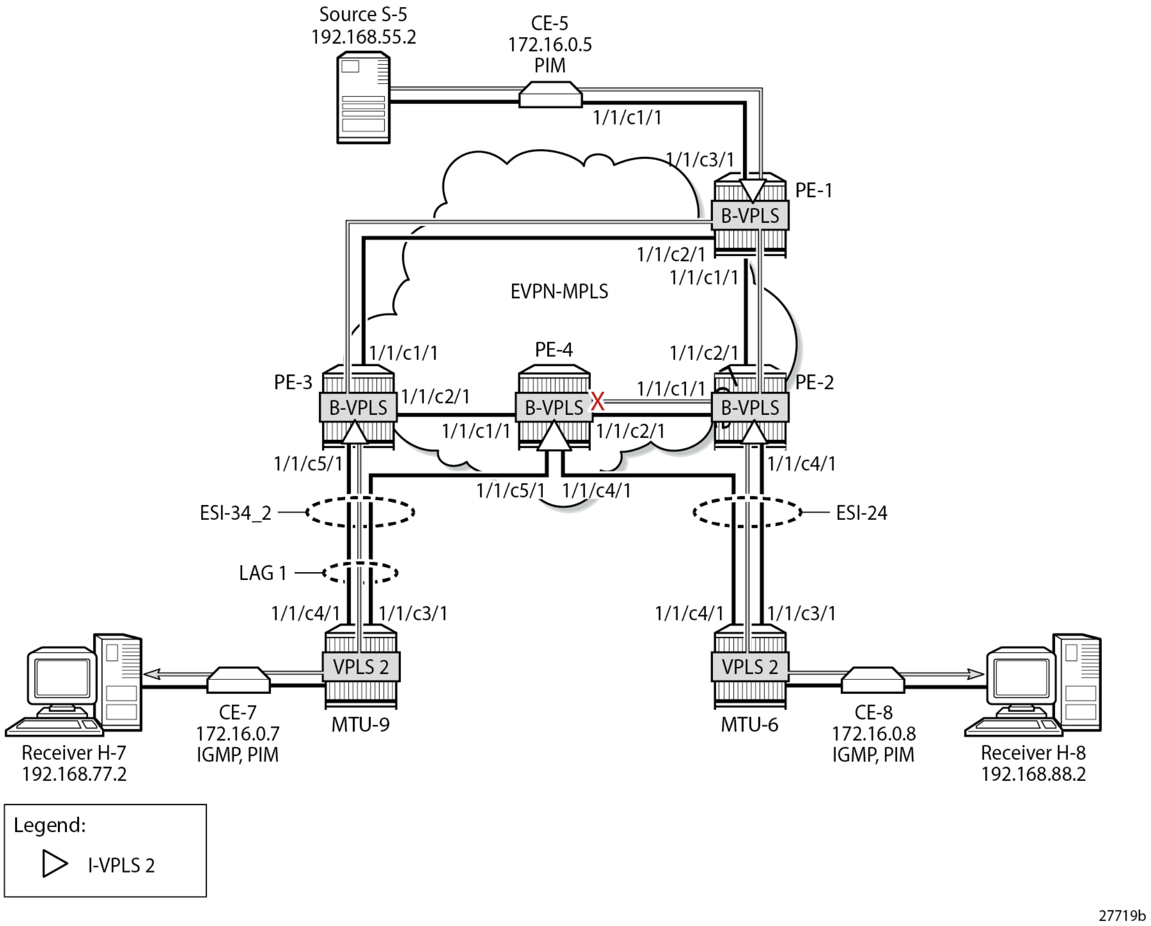
```
-------------------------------------------------------------------------------
Groups : 1
===============================================================================
```

With the PIM snooping group information available on the NDF, the traffic loss is limited when the NDF PE-4 becomes the DF after failover. Data-driven PIM state synchronization does not store PIM states in a database, so the DF election in the ES should be configured as non-revertive, to prevent that when the preferred DF is restored after a failover, the system would revert to a DF that is unaware of the PIM state.

PE-4 is also NDF in the single-active MH ES ESI-24, but it received no PIM state synchronization information from DF PE-2. Data-driven PIM state synchronization is not supported for single-active MH in PBB-EVPN services, because it is not allowed to have two PEs in a single-active MH ES using the same source BMAC, with the potential risk of traffic sent by remote PEs to the NDF PE (based on it sending to the shared source BMAC) being dropped. However, for a faster failover in single-active MH, multi-chassis synchronization (MCS) can be configured, as described in the next section.

Figure 277: EVPN-MPLS with MH and PIM Snooping – Receivers H-7 and H-8 Joined shows the multicast traffic flow when PIM snooping is enabled and both receivers H-7 and H-8 have joined the multicast group. All PEs receive the multicast traffic on the P2MP tunnel, but only DF PE-2 and DF PE-3 forward the multicast traffic to the MTUs, which forward the traffic to the CEs, where it is forwarded to the receivers.

Figure 277: EVPN-MPLS with MH and PIM Snooping – Receivers H-7 and H-8 Joined



PBB-EVPN with MH – PIM Snooping for IPv4 with MCS

MCS of the IPv4 PIM snooping state for SAPs and spoke-SDPs can optionally be configured in the case of MH. MCS reduces the failover time when data-driven PIM state synchronization is not supported; for example, for single-active MH in PBB-EVPN services. The synchronization information is stored in an MCS synchronization DB. MCS is configured on PE-2, identifying the peer (PE-4), with PIM snooping for spoke-SDPs as MCS client application and the list of spoke-SDPs, as follows:

```
On PE-2:
configure
    redundancy
        multi-chassis
            peer 192.0.2.4 create
                sync
                    pim-snooping spoke-sdps
                    sdp 26 create
                        range 2-2 sync-tag "syncSA"
                    exit
                    no shutdown
                exit
                no shutdown
            exit all
```

On PE-4, MCS is configured for peer PE-2, as follows:

```
On PE-4:
configure
    redundancy
        multi-chassis
            peer 192.0.2.2 create
                sync
                    pim-snooping spoke-sdps
                    sdp 46 create
                        range 2-2 sync-tag "syncSA"
                    exit
                    no shutdown
                exit
                no shutdown
            exit all
```

When H-8 has joined the multicast group, the MCS sync-database on PE-2 shows the PIM snooping entries on the spoke-SDP 26:2 of the single-active MH ESI-24, as follows:

```
*A:PE-2# tools dump redundancy multi-chassis sync-database detail

If no entries are present for an application, no detail will be displayed.

FLAGS LEGEND: ld - local delete; da - delete alarm; pd - pending global delete;
              oal - omcr alarmed; ost - omcr standby

Peer Ip 192.0.2.4


Application pim-snooping-sdp
Sdp-id             Client Key
 SyncTag                          DLen  Flags            timeStamp
  deleteReason code and description                   #ShRec
-------------------------------------------------------------------------------
26:2               Adj 172.16.0.8
```

```
 syncSA                          72    -- -- -- --- --- 08/24/2023 14:26:36
  0x0                                                 0
26:2               IfSG SG 192.168.55.2 232.1.1.1
 syncSA                          69    -- -- -- --- --- 08/24/2023 14:26:26
  0x0                                                 0

The following totals are for:
 peer ip ALL, port/lag/sdp ALL, sync-tag ALL, application ALL
Valid Entries:                   2
Locally Deleted Entries:         0
Locally Deleted Alarmed Entries: 0
Pending Global Delete Entries:   0
Omcr Alarmed Entries:            0
Omcr Standby Entries:            0
Associated Shared Records (ALL): 0
Associated Shared Records (LD):  0
```

On PE-4, the MCS sync-database is similar, but with SDP ID 46:2 instead of 26:2.

Even though PE-4 is the NDF for both ESs, the MFIB is populated with the spoke-SDP to MTU-6, as well as the B-VPLS PBB-EVPN destinations to the other PEs, as follows:

```
*A:PE-4# show service id 2 mfib

===============================================================================
Multicast FIB, Service 2
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
192.168.55.2    232.1.1.1             sdp:46:2                     Local    Fwd
                                      b-mpls:192.0.2.1:524282      100      Fwd
                                      b-mpls:192.0.2.2:524282      100      Fwd
                                      b-mpls:192.0.2.3:524282      100      Fwd
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

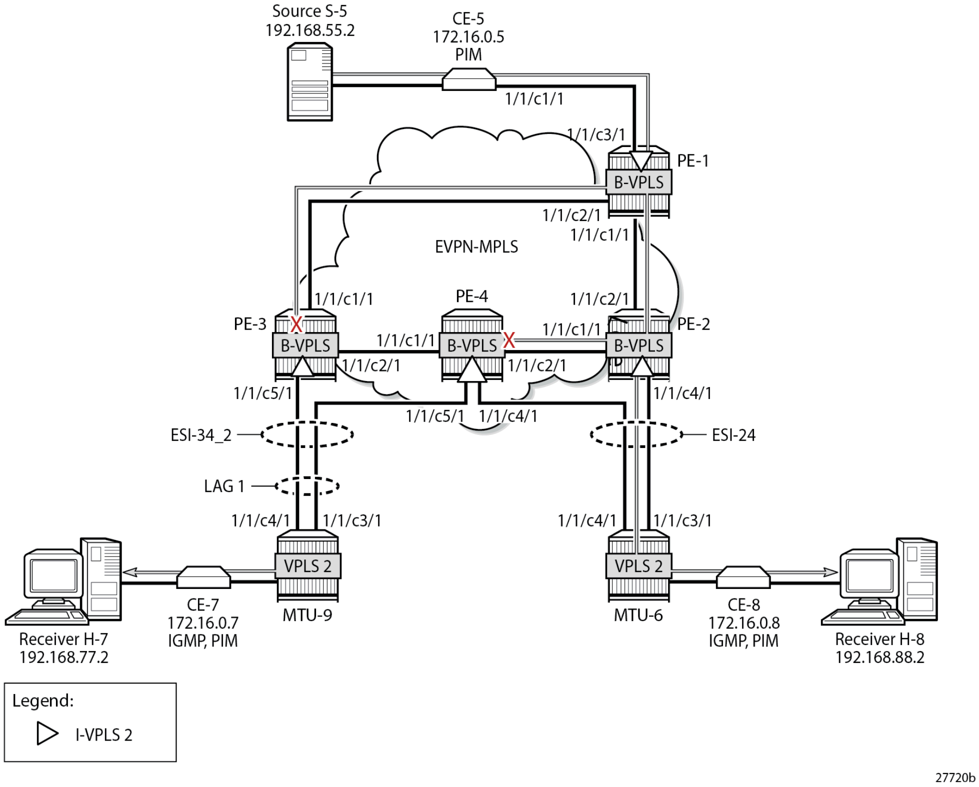
The following command on PE-4 shows that the incoming PIM interface is the B-VPLS EVPN-MPLS interface and the spoke-SDP is the outgoing interface. Again, the split-horizon mechanism prevents traffic received from the B-VPLS EVPN-MPLS interface from being forwarded on the B-VPLS EVPN-MPLS interface.

```
*A:PE-4# show service id 2 pim-snooping group detail

===============================================================================
PIM Snooping Source Group ipv4
===============================================================================
Group Address      : 232.1.1.1
Source Address     : 192.168.55.2
Up Time            : 0d 00:02:02

Up JP State        : Joined              Up JP Expiry       : 0d 00:01:19
Up JP Rpt          : Not Joined StarG    Up JP Rpt Override : 0d 00:00:00

RPF Neighbor       : 172.16.0.5
Incoming Intf      : b-EVPN-MPLS
Outgoing Intf List : b-EVPN-MPLS, SPOKE_SDP:46:2

Forwarded Packets  : 100350              Forwarded Octets   : 151929900
-------------------------------------------------------------------------------
Groups : 1
```

```
===============================================================================
```

However, PE-4 remains the NDF for both ESs and does not forward any traffic from the B-VPLS EVPN-MPLS interface to the spoke-SDP. Figure 278: PBB-EVPN with MH and PIM Snooping – Receiver H-8 Joined shows the multicast traffic flow when PIM snooping is enabled and receiver H-8 has joined.

Figure 278: PBB-EVPN with MH and PIM Snooping – Receiver H-8 Joined



Failover

Figure 277: EVPN-MPLS with MH and PIM Snooping – Receivers H-7 and H-8 Joined showed the multicast traffic flow when both H-7 and H-8 have joined the multicast group. PE-2 is the DF for ESI-24 and PE-4 is the DF for ESI-34_2. The following failures are introduced to force a failover from PE-2 to PE-4 and from PE-3 to PE-4. Data-driven PIM state synchronization is used for all-active MH; MCS is configured for fast failover in the single-active MH ES ESI-24.

On MTU-6, SDP 62 is disabled, as follows:

```
On MTU-6:
configure service sdp 62 shutdown
```

On MTU-9, port 1/1/c3/1 toward PE-3 is disabled, as follows:

```
On MTU-9:
configure port 1/1/c3/1 shutdown
```

Log 99 on PE-3 shows that the DF state in ESI-34_2 changed to false:

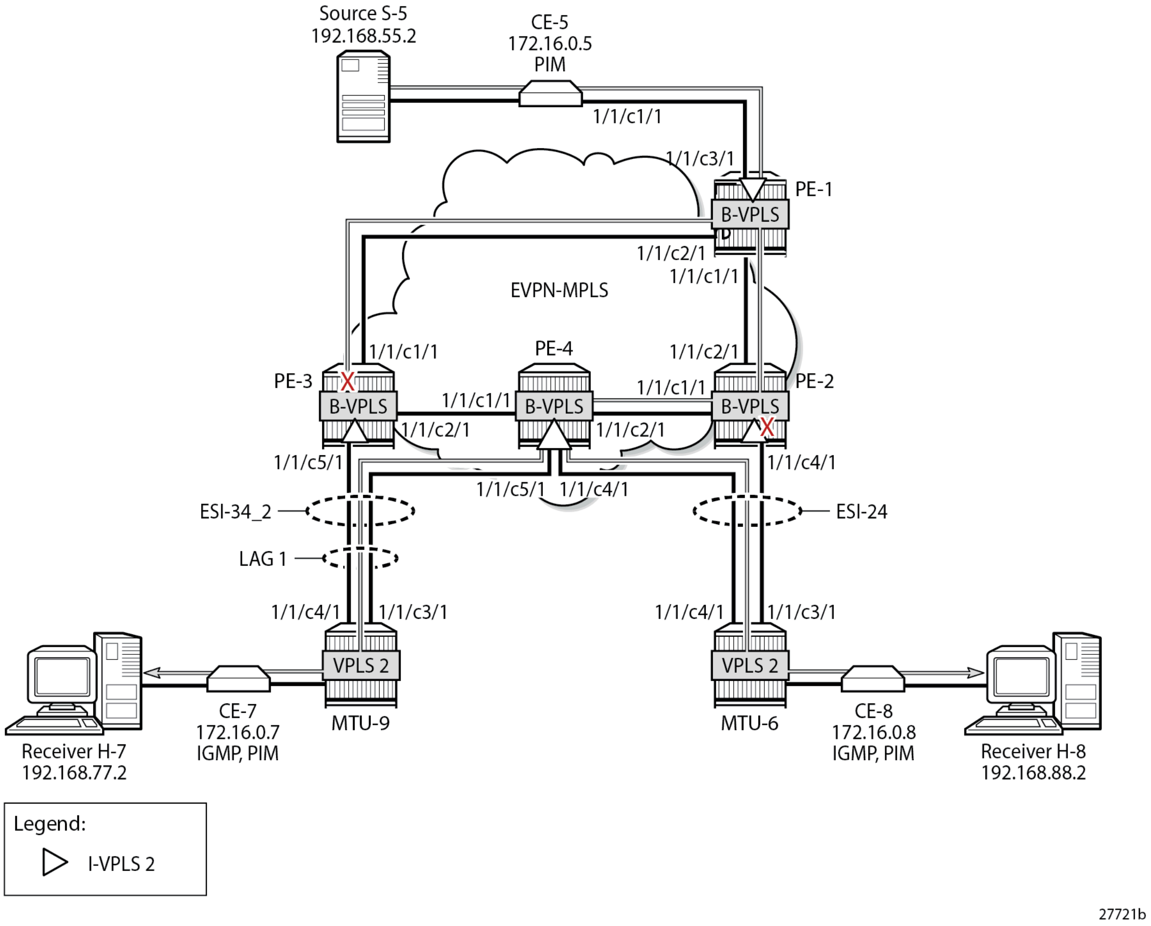
```
184 2023/08/24 14:29:36.634 UTC MINOR: SVCMGR #2095 Base
"Ethernet Segment:ESI-34_2, ISID:2, Designated Forwarding state changed to:false"
```

PE-4 becomes the DF for both ESs, as follows:

```
*A:PE-4# show service id 2 ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                           Status
-------------------------------------------------------------------------------
lag-1:2               ESI-34_2                          DF
===============================================================================

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                           Status
-------------------------------------------------------------------------------
46:2                  ESI-24                            DF
===============================================================================
No vxlan instance entries
```

Figure 279: EVPN-MPLS with MH and PIM Snooping – Multicast Flow after Failover shows the traffic flow after failover to the new DF, PE-4.

Figure 279: EVPN-MPLS with MH and PIM Snooping – Multicast Flow after Failover



PE-2 receives the multicast stream from PE-1 on port 1/1/c2/1 and forwards it to port 1/1/c1/1 to PE-4; it does not forward to port 1/1/c4/1 because SDP 26 is down, as follows:

```
*A:PE-2# show port 1/1/c1/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                         Ingress Packets                  Ingress Octets
Id                            Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c1/1                                  63                           11627
                                       16519                        25288008
===============================================================================

*A:PE-2# show port 1/1/c2/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                         Ingress Packets                  Ingress Octets
Id                            Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
```

```
1/1/c2/1                                16478                        25277265
                                           21                            2162
===============================================================================

*A:PE-2# show port 1/1/c3/1 statistics

*A:PE-2# show port 1/1/c4/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c4/1                                   16                            1703
                                           17                            1908
===============================================================================
```

PE-4 receives the multicast traffic on port 1/1/c2/1 and forwards it on port 1/1/c4/1 toward MTU-6, and on port 1/1/c5/1 to MTU-9, as follows:

```
*A:PE-4# show port 1/1/c1/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c1/1                                   19                            2068
                                           17                            1792
===============================================================================

*A:PE-4# show port 1/1/c2/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c2/1                                16519                        25288058
                                           62                           11487
===============================================================================

*A:PE-4# show port 1/1/c3/1 statistics

*A:PE-4# show port 1/1/c4/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c4/1                                   16                            1703
                                        16472                        25113619
===============================================================================

*A:PE-4# show port 1/1/c5/1 statistics

===============================================================================
Port Statistics on Slot 1
```

```
===============================================================================
Port                        Ingress Packets                  Ingress Octets
Id                           Egress Packets                   Egress Octets
-------------------------------------------------------------------------------
1/1/c5/1                                 20                            2560
                                      16477                        24752460
===============================================================================
```

MTU-6 forwards the traffic to CE-8, which forwards it to H-8. MTU-9 forwards the traffic to CE-7, which sends it to H-7. PE-3 drops the multicast traffic because LAG 1 is down because of the failure that was introduced at MTU-9 (port disabled).

Conclusion

PIM snooping reduces flooding of multicast traffic in L2 services and can be used in PBB-EVPN I-VPLSs in the same way as in I-VPLSs using B-VPLS without EVPN. PIM snooping can be used in all-active and single-active MH scenarios with data-driven state synchronization and MCS, respectively.

Preference-based and Non-revertive EVPN DF Election

This chapter provides information about Preference-based and Non-revertive EVPN DF Election.

Topics in this chapter include:

Applicability

This chapter was initially written based on SR OS Release 15.0.R3, but the CLI in the current edition corresponds to SR OS Release 21.2.R2. Preference-based and non-revertive EVPN Designated Forwarder (DF) election is supported in SR OS Release 15.0.R1, and later. This mechanism works for Ethernet Segments (ESs) and virtual ESs (vESs).

Overview

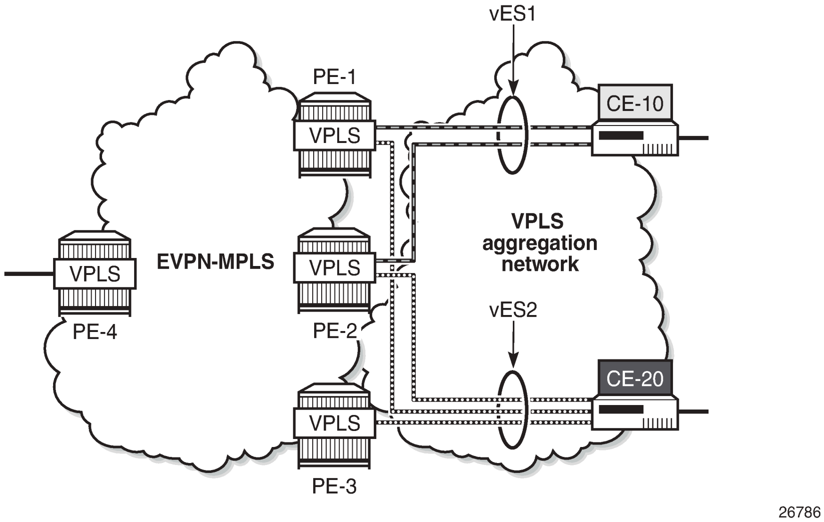
RFC 7432 defines the Designated Forwarder (DF) in (PBB-)EVPN networks as the PE that will forward the following packets to a multi-homed node:

- Broadcast, Unknown unicast, and Multicast (BUM) traffic in an all-active multi-homing Ethernet Segment (ES)
- BUM and unicast in a single-active multi-homing ES

For more information about vESs, see chapter Virtual Ethernet Segments.

Figure 280: Virtual Ethernet Segments shows a topology with two vESs.

Figure 280: Virtual Ethernet Segments

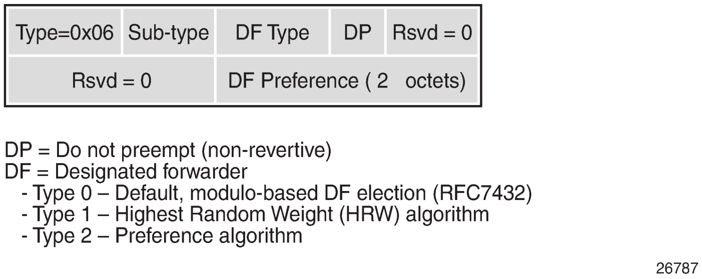


Taking the Ethernet VPN Identifier (EVI) or ISID and the number of PEs in the ES as input, the RFC 7432 service-carving algorithm elects the DF from the list of candidate PEs that advertise the ES identifier (ESI). While this algorithm provides an automated and fair DF distribution across services in the ES, it does not allow the operator to control what PE is the DF for which service. In addition, in case of a DF failure, when the former DF comes back up, a new DF switchover will cause unnecessary packet loss (this mode of operation is called revertive). SR OS implements *draft-ietf-bess-evpn-pref-df* to give more control to the operator on the DF election and avoid the revertive mode.

In SR OS, in addition to the automated service-carving, the DF election can also be controlled by configuring a preference manually. Also, it is possible to force an on-demand DF switchover without reconfiguring the PEs in the ES. Furthermore, the non-revertive option prevents an automatic switchover when a new active PE can preempt the existing DF PE. The non-revertive option avoids service impact when an ES comes back up.

Figure 281: BGP-EVPN extended community for DF election shows the BGP-EVPN extended community defined for DF election and the different values described in draft-ietf-bess-evpn-pref-df.

Figure 281: BGP-EVPN extended community for DF election



Type=0x06	Sub-type	DF Type	DP	Rsvd = 0
Rsvd = 0		DF Preference (2 octets)		

DP = Do not preempt (non-revertive)
DF = Designated forwarder
- Type 0 – Default, modulo-based DF election (RFC7432)
- Type 1 – Highest Random Weight (HRW) algorithm
- Type 2 – Preference algorithm

The "Do not preempt" (DP) bit is set to enable the non-revertive option. When preference-based service carving is configured in the ES, DF type 2 is advertised along with a 2-byte preference value, which is 32767 by default.

Service carving can be configured in auto mode or manual mode. The preference can only be configured in manual mode.

```
*A:PE-2>config>service>system>bgp-evpn>eth-seg>service-carving# mode ?
  - mode {auto|manual|off}

 <auto|manual|off>    : auto|manual|off
```

When manual mode is enabled, the following parameters can be configured to control which PE will be elected as DF:

```
*A:PE-2>config>service>system>bgp-evpn>eth-seg>service-carving# manual ?
  - manual

 [no] evi             - Configure EVI range (primary for non-preference based DF
                    election and lowest-preference for preference based DF election)
 [no] isid            - Configure ISID range (primary for non-preference based DF
                    election and lowest-preference for preference based DF election)
 [no] preference      + Configure DF preference election information
```

The EVI and ISID ranges configured in the service-carving context do not need to be consistent with any ranges configured for virtual ESs.

When preference is configured manually, a preference value can be provided along with the non-revertive option:

```
*A:PE-2>config>service>system>bgp-evpn>eth-seg>service-carving>manual# preference ?
  - no preference
  - preference [create] [non-revertive]

 <create>             : keyword
 <non-revertive>      : keyword

      value           - Configure DF preference value
```

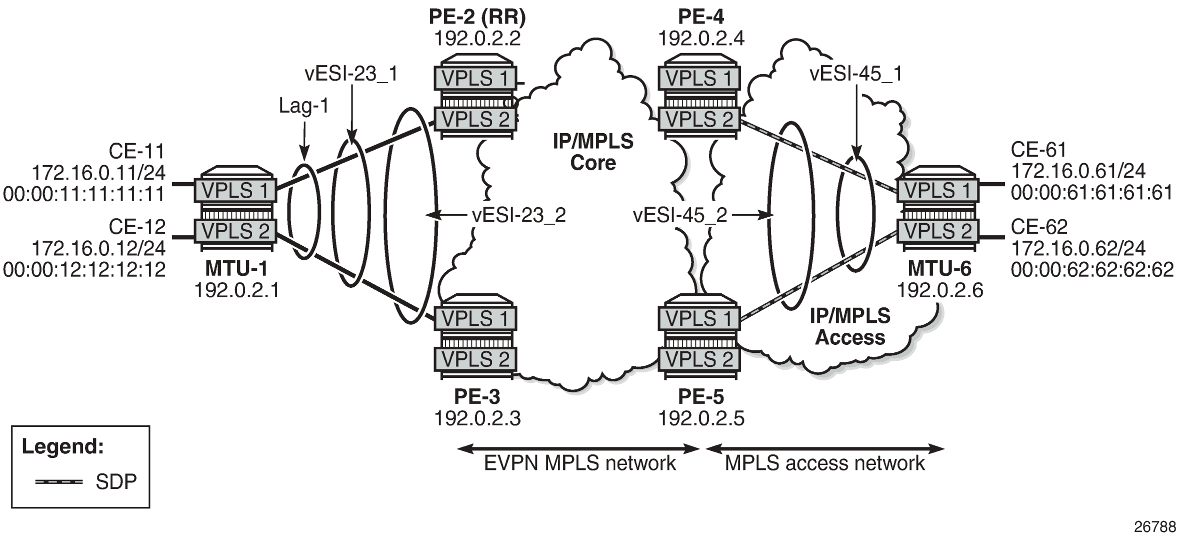
The preference-based EVPN DF election is as follows:

- By default, all SAPs and spoke-SDPs on the configured ES select the highest-preference PE as DF; however, when the EVI or ISID ranges are configured in the ES, the lowest-preference PE is selected.
- When the preference is equal, the DP bit is the tiebreaker: DP=1 wins over DP=0.
- For equal preference and DP, the PE IP address is the tiebreaker: the lowest IP address wins.

Configuration

Figure 282: Example topology with all-active and single-active vESs shows the example topology with six nodes. EVPN-MPLS is configured between the core PE nodes. All-active vESs are configured between PE-2 and PE-3 and single-active vESs are configured between PE-4 and PE-5.

Figure 282: Example topology with all-active and single-active vESs



The initial configuration includes:

- Cards, MDAs, ports
- LAG 1 between MTU-1, PE-2, PE-3
- Router interfaces
- IS-IS (alternatively, OSPF could be used)
- LDP

BGP is configured on the four core PEs with PE-2 as Route Reflector (RR). The BGP configuration on RR PE-2 is as follows:

```
# on RR PE-2:
configure
    router
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family evpn
                cluster 1.1.1.1
                peer-as 64500
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
```

VPLS 1 and VPLS 2 are configured on each node. The PEs have EVPN-MPLS enabled. The configuration on PE-2 is as follows:

```
# on PE-2:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap lag-1:1.1 create
                no shutdown
            exit
            no shutdown
        exit
        vpls 2 name "VPLS 2" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 2
                mpls bgp 1
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap lag-1:2.1 create
                no shutdown
            exit
            no shutdown
        exit
```

The configuration on the other PEs is similar; PE-4 and PE-5 have a spoke-SDP configured instead of a SAP. For an explanation of the configuration, see chapter EVPN for MPLS Tunnels.

Service carving: auto mode

On PE-2 and PE-3, the following all-active multi-homing vESs are configured:

```
# on PE-2, PE-3:
configure
```

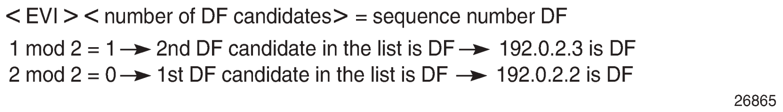
```
    service
        system
            bgp-evpn
                ethernet-segment "vESI-23_1" virtual create
                    esi 01:00:00:00:00:23:01:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 1
                    qinq
                        s-tag-range 1
                    exit
                    no shutdown
                exit
                ethernet-segment "vESI-23_2" virtual create
                    esi 01:00:00:00:00:23:02:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 1
                    qinq
                        s-tag-range 2
                    exit
                    no shutdown
                exit
```

The service carving mode is set to **auto**, so the DF election is based on a modulo function of the EVI and the number of DF candidates. In the vES "vESI-23_1", there are two DF candidates, PE-2 and PE-3, listed in that order because PE-2 has the lower system IP address, as follows:

```
*A:PE-3# show service system bgp-evpn ethernet-segment name "vESI-23_1" all
| match "EVI Information" post-lines 20
EVI Information
===============================================================================
EVI                 SvcId               Actv Timer Rem      DF
-------------------------------------------------------------------------------
1                   1                   0                   yes
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================

-------------------------------------------------------------------------------
DF Candidate list
-------------------------------------------------------------------------------
EVI                                     DF Address
-------------------------------------------------------------------------------
1                                       192.0.2.2
1                                       192.0.2.3
-------------------------------------------------------------------------------
Number of entries: 2
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
```

The first DF candidate from the list will be selected when the result of the modulo function equals 0; the second DF candidate when the result equals 1. The calculation is as follows:

Figure 283: Calculation


< EVI > < number of DF candidates> = sequence number DF
1 mod 2 = 1 → 2nd DF candidate in the list is DF → 192.0.2.3 is DF
2 mod 2 = 0 → 1st DF candidate in the list is DF → 192.0.2.2 is DF
26865

The following shows that PE-2 is not the DF for VPLS 1, but it is the DF for VPLS 2:

```
*A:PE-2# show service id 1 ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
lag-1:1.1             vESI-23_1                        NDF
===============================================================================
No sdp entries
No vxlan instance entries
```

```
*A:PE-2# show service id 2 ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
lag-1:2.1             vESI-23_2                        DF
===============================================================================
No sdp entries
No vxlan instance entries
```

Instead of the preceding show commands, the following tools commands can be used:

```
*A:PE-2# tools dump service system bgp-evpn ethernet-segment "vESI-23_1"  evi 1 df

[04/22/2021 15:20:08] Computed DF: 192.0.2.3 (Remote) (Boot Timer Expired: Yes)

*A:PE-2# tools dump service system bgp-evpn ethernet-segment "vESI-23_2"  evi 2 df

[04/22/2021 15:20:08] Computed DF: 192.0.2.2 (This Node) (Boot Timer Expired: Yes)
```

Service carving: preference-based manual mode

To have more control, the vES can be configured in manual mode. The following reconfigures the vES "vESI-23_1" in manual mode, preference-based and revertive with preference 32767 (default) on PE-2 and 5000 on PE-3, whereas vES "vESI-23_2" is preference-based and non-revertive with preference 15000 on PE-2 and 20000 on PE-3.

An EVI range is configured for ES "vESI-23_2", but not for ES "vESI-23_1". When no EVI range is configured, the highest preference wins; for configured EVI ranges, the lowest preference wins. When there are no failures, PE-2 will be the DF for "vESI-23_1" (highest preference) and for "vESI-23_2" (lowest preference for configured EVI 2).

To modify the service-carving mode from auto to manual, the ES must be disabled first (shutdown). The following is configured on PE-2:

```
# on PE-2:
configure
    service
        system
            bgp-evpn
                ethernet-segment "vESI-23_1" virtual create
                    shutdown
                    service-carving
                        mode manual
                        manual
                            preference create
                            exit
                        exit
                    exit
                    no shutdown
                exit
                ethernet-segment "vESI-23_2" virtual create
                    shutdown
                    service-carving
                        mode manual
                        manual
                            preference non-revertive create
                                value 15000
                            exit
                            evi 2
                        exit
                    exit
                    no shutdown
                exit
```

The keyword **non-revertive** is added for vES "vESI-23_2", but not for vES "vESI-23_1".

The following is configured on PE-3:

```
# on PE-3:
configure
    service
        system
            bgp-evpn
                ethernet-segment "vESI-23_1" virtual create
                    shutdown
                    service-carving
                        mode manual
                        manual
                            preference create
                                value 5000
                            exit
                        exit
                    exit
                    no shutdown
                exit
                ethernet-segment "vESI-23_2" virtual create
                    shutdown
                    service-carving
                        mode manual
                        manual
                            preference non-revertive create
                                value 20000
                            exit
                            evi 2
                        exit
                    exit
                    no shutdown
                exit
```

For the single-active multi-homing vESs on PE-4 and PE-5, the same preferences are configured manually. The ES configuration on PE-4 is as follows:

```
# on PE-4:
configure
    service
        system
            bgp-evpn
                ethernet-segment "vESI-45_1" virtual create
                    esi 01:00:00:00:00:45:01:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode manual
                        manual
                            preference create
                            exit
                        exit
                    exit
                    multi-homing single-active
                    sdp 46
                    vc-id-range 1
                    vc-id-range 500 to 501
                    no shutdown
                exit
                ethernet-segment "vESI-45_2" virtual create
                    esi 01:00:00:00:00:45:02:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode manual
                        manual
                            preference non-revertive create
                                value 15000
                            exit
                            evi 2
                        exit
                    exit
                    multi-homing single-active
                    sdp 46
                    vc-id-range 2
                    no shutdown
                exit
```

The ES configuration on PE-5 is as follows:

```
# on PE-5:
configure
    service
        system
            bgp-evpn
                ethernet-segment "vESI-45_1" virtual create
                    esi 01:00:00:00:00:45:01:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode manual
                        manual
                            preference create
                                value 5000
                            exit
                        exit
                    exit
                    multi-homing single-active
                    sdp 56
                    vc-id-range 1
                    vc-id-range 500 to 501
                    no shutdown
                exit
                ethernet-segment "vESI-45_2" virtual create
                    esi 01:00:00:00:00:45:02:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode manual
                        manual
                            preference non-revertive create
                                value 20000
                            exit
                            evi 2
                        exit
                    exit
                    multi-homing single-active
                    sdp 56
                    vc-id-range 2
                    no shutdown
                exit
```

The preference configuration must be consistent across the PEs in the ES (manual or auto), otherwise the system reverts to the modulo-based DF election.

With preference-based DF election configured with default preference value 32767 and revertive, PE-4 sends the following BGP-EVPN update to the RR PE-2. The **df-election** extended community shows the DP=0 (revertive) and DF preference 32767.

```
# on PE-4:
47 2021/04/22 15:21:19.329 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 71
    Flag: 0x90 Type: 14 Len: 34 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.4
        Type: EVPN-ETH-SEG Len: 23 RD: 192.0.2.4:0
               ESI: 01:00:00:00:00:45:01:00:00:01, IP-Len: 4 Orig-IP-Addr: 192.0.2.4
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        df-election::DF-Type:Preference/DP:0/DF-Preference:32767/AC:1
        target:00:00:00:00:45:01
"
```

The following command shows the information in the preceding BGP-EVPN Ethernet-segment route for "vESI-45_1" sent by PE-4 to the RR PE-2:

```
*A:PE-4# show router bgp routes evpn eth-seg hunt
===============================================================================
 BGP Router ID:192.0.2.4        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
```

```
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Eth-Seg Routes
===============================================================================
---snip---
-------------------------------------------------------------------------------
RIB Out Entries
-------------------------------------------------------------------------------
Network        : n/a
Nexthop        : 192.0.2.4
To             : 192.0.2.2
Res. Nexthop   : n/a
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : n/a
Connector      : None
Community      :
                 df-election::DF-Type:Preference/DP:0/DF-Preference:32767/AC:1
                 target:00:00:00:00:45:01
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.0.2.2
Origin         : IGP
AS-Path        : No As-Path
EVPN type      : ETH-SEG
ESI            : 01:00:00:00:00:45:01:00:00:01
Originator IP  : 192.0.2.4
Route Dist.    : 192.0.2.4:0
Route Tag      : 0
Neighbor-AS    : n/a
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
---snip---
```

The following command shows the DF preference election information for ES "vESI-45_1" with the preference mode revertive, the configured preference value on PE-4 (default 32767), and the operational preference value. No EVI ranges or ISID ranges are configured in this ES.

```
*A:PE-4# show service system bgp-evpn ethernet-segment name "vESI-45_1"

===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : vESI-45_1
Eth Seg Type            : Virtual
Admin State             : Enabled            Oper State         : Up
ESI                     : 01:00:00:00:00:45:01:00:00:01
Multi-homing            : singleActive       Oper Multi-homing  : singleActive
ES SHG Label            : 524276
Source BMAC LSB         : <none>
Sdp Id                  : 46
ES Activation Timer     : 3 secs
Oper Group              : (Not Specified)
Svc Carving             : manual             Oper Svc Carving   : manual
Cfg Range Type          : lowest-pref

-------------------------------------------------------------------------------
DF Pref Election Information
-------------------------------------------------------------------------------
Preference     Preference     Last Admin Change       Oper Pref      Do No
Mode           Value                                  Value          Preempt
```

```
-------------------------------------------------------------------------------
revertive      32767          04/22/2021 15:21:19      32767          Disabled
-------------------------------------------------------------------------------
EVI Ranges: <none>
ISID Ranges: <none>
===============================================================================
```

The following command shows the DF preference election information for ES "vESI-45_2" with the preference mode non-revertive, the configured preference value on PE-4 (15000), and the operational preference value. The only configured EVI range is from 2 to 2. No ISID ranges are configured. For the configured EVI or ISID values, the lowest preference wins, as shown by the **Cfg Range Type** : *lowest-pref* parameter.

```
*A:PE-4# show service system bgp-evpn ethernet-segment name "vESI-45_2"

===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : vESI-45_2
Eth Seg Type            : Virtual
Admin State             : Enabled            Oper State         : Up
ESI                     : 01:00:00:00:00:45:02:00:00:01
Multi-homing            : singleActive       Oper Multi-homing  : singleActive
ES SHG Label            : 524275
Source BMAC LSB         : <none>
Sdp Id                  : 46
ES Activation Timer     : 3 secs
Oper Group              : (Not Specified)
Svc Carving             : manual             Oper Svc Carving   : manual
Cfg Range Type          : lowest-pref

-------------------------------------------------------------------------------
DF Pref Election Information
-------------------------------------------------------------------------------
Preference     Preference     Last Admin Change        Oper Pref      Do No
Mode           Value                                   Value          Preempt
-------------------------------------------------------------------------------
non-revertive  15000          04/22/2021 15:21:19      15000          Enabled
-------------------------------------------------------------------------------

-------------------------------------------------------------------------------
EVI Ranges
-------------------------------------------------------------------------------
From                                    To
-------------------------------------------------------------------------------
2                                       2
-------------------------------------------------------------------------------
ISID Ranges: <none>
===============================================================================
```

It is important to note that a router will prune a remote PE from the DF candidate list for an ES if it does not receive the corresponding Auto Discovery (AD) per-EVI and AD per-ES routes for that PE. A remote PE will not be shown in the DF Candidate list if its AD per-ES route is withdrawn. This is only true for EVPN. In PBB-EVPN, there are no AD routes, therefore the DF Candidate list is built out of the ES routes only.

DF election: higher preference prevails for non-configured EVI ranges

The PEs run the DF election per PE per EVI, and the elected DF for a service will activate the SAP/Spoke-SDP when the es-activation-timer expires. PE-4 is the DF in "vESI-45_1" used in VPLS 1, as follows.

The EVI is not configured in ES "vESI-45_1", so the higher preference prevails. The ES "vESI-45_1" has (default) preference 32767 on PE-4 (DF) and preference 5000 on PE-5 (Non-Designated Forwarder (NDF)).

```
*A:PE-4# show service id 1 ethernet-segment
No sap entries

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
46:1                  vESI-45_1                        DF
===============================================================================
No vxlan instance entries
```

```
*A:PE-5# show service id 1 ethernet-segment
No sap entries

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
56:1                  vESI-45_1                        NDF
===============================================================================
No vxlan instance entries
```

The preference value can be modified on the fly on an active ES without the need to disable the ES (shutdown). This allows the user to force a new DF for the ES for maintenance operations on the former DF or other reasons.

DF election: lowest preference prevails for configured EVI ranges

ES "vESI-45_2" is configured with EVI 2, so the lowest preference prevails. The admin preference value is 15000 on PE-4 and 20000 on PE-5. Both PE-4 and PE-5 are DF candidates, but PE-4 has the lowest preference, so it will be the DF, as follows:

```
*A:PE-4# show service system bgp-evpn ethernet-segment name "vESI-45_2" all
| match "EVI Ranges" post-lines 30
EVI Ranges
-------------------------------------------------------------------------------
From                                    To
-------------------------------------------------------------------------------
2                                       2
-------------------------------------------------------------------------------
ISID Ranges: <none>
===============================================================================

===============================================================================
EVI Information
===============================================================================
EVI                 SvcId               Actv Timer Rem      DF
-------------------------------------------------------------------------------
2                   2                   0                   yes
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
-------------------------------------------------------------------------------
DF Candidate list
-------------------------------------------------------------------------------
EVI                                     DF Address
-------------------------------------------------------------------------------
2                                       192.0.2.4
2                                       192.0.2.5
-------------------------------------------------------------------------------
Number of entries: 2
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
```

DF election: DP prevails when preferences are equal

In the preceding example, PE-4 was the DF in ES "vESI-45_1" because of the higher preference. The ES configuration is modified on PE-5 as follows: the preference is set to the default, which is equal to the preference on PE-4, and the non-revertive (do not preempt - DP) option enabled. The **non-revertive** keyword can only be configured at creation time. An attempt to modify this behavior afterward results in the following error message:

```
*A:PE-5>config>service>system>bgp-evpn>eth-seg>service-carving>manual#
preference non-revertive create
MINOR: CLI revertive mode can be specified only at creation time.
```

The existing preference first needs to be removed, which can only be done when the ES is disabled (shutdown); if not, the following error is raised:

```
*A:PE-5>config>service>system>bgp-evpn>eth-seg>service-carving>manual#
 no preference
MINOR: SVCMGR #8074 Cannot delete preference - ethernet-segment not shut
```

The service carving in the ES is configured with default preference and non-revertive option, as follows:

```
# on PE-5:
configure
    service
        system
            bgp-evpn
                ethernet-segment "vESI-45_1" virtual create
                    ahutdown
                    service-carving
                        mode manual
                        manual
                            no preference
                            preference non-revertive create
                            exit
                        exit
                    exit
                    no shutdown
```

The ES configuration on PE-4 remains unchanged, so the behavior is revertive. PE-4 and PE-5 have the same preference (default 32767), but PE-5 is non-revertive and becomes the DF, as follows:

```
*A:PE-5# show service id 1 ethernet-segment
No sap entries
```

```
===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
56:1                  vESI-45_1                        DF
===============================================================================
No vxlan instance entries
```

DF election: lowest IP address prevails when preferences and DP are equal

The vES configuration on PE-4 is modified by enabling the non-revertive option (after deleting the existing preference configuration), as follows:

```
# on PE-4:
configure
    service
        system
            bgp-evpn
                ethernet-segment "vESI-45_1" virtual create
                    shutdown
                    service-carving
                        mode manual
                        manual
                            no preference
                            preference non-revertive create
                            exit
                        exit
                    exit
                    no shutdown
```

PE-4 and PE-5 have an equal preference (32767) and non-revertive behavior. The tiebreaker for the DF selection is the IP address. PE-4 has the lower IP address and becomes the DF, as follows:

```
*A:PE-4# show service id 1 ethernet-segment
No sap entries

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
46:1                  vESI-45_1                        DF
===============================================================================
No vxlan instance entries
```

Service-carving configuration must be consistent

When the service carving on one of the PEs in the ES is configured in auto mode while one of the other PEs in the ES is configured in manual mode, the system reverts to modulo-based auto mode. The configuration of ES "vESI-45_1" remains unchanged on PE-4, but is modified on PE-5, as follows:

```
*A:PE-5#
configure
    service
        system
```

```
            bgp-evpn
                ethernet-segment "vESI-45_1" virtual create
                    shutdown
                    service-carving
                        manual no preference
                        mode auto
                    exit
                    no shutdown
```

ES "vESI-45_1" will operate in auto mode on PE-4 and on PE-5. The following **show** command on PE-4 shows that the ES is configured in manual mode, but operates in auto mode:

```
*A:PE-4# show service system bgp-evpn ethernet-segment name "vESI-45_1"

===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : vESI-45_1
Eth Seg Type            : Virtual
Admin State             : Enabled            Oper State         : Up
ESI                     : 01:00:00:00:00:45:01:00:00:01
Multi-homing            : singleActive       Oper Multi-homing  : singleActive
ES SHG Label            : 524274
Source BMAC LSB         : <none>
Sdp Id                  : 46
ES Activation Timer     : 3 secs
Oper Group              : (Not Specified)
Svc Carving             : manual             Oper Svc Carving   : auto
Cfg Range Type          : lowest-pref

-------------------------------------------------------------------------------
DF Pref Election Information
-------------------------------------------------------------------------------
Preference     Preference     Last Admin Change        Oper Pref      Do No
Mode           Value                                   Value          Preempt
-------------------------------------------------------------------------------
non-revertive  32767          04/22/2021 15:34:52      32767          Enabled
-------------------------------------------------------------------------------
EVI Ranges: <none>
ISID Ranges: <none>
===============================================================================
```

The following command on PE-5 shows that the ES is configured in auto mode and operates in auto mode:

```
*A:PE-5# show service system bgp-evpn ethernet-segment name "vESI-45_1"

===============================================================================
Service Ethernet Segment
===============================================================================
Name                    : vESI-45_1
Eth Seg Type            : Virtual
Admin State             : Enabled            Oper State         : Up
ESI                     : 01:00:00:00:00:45:01:00:00:01
Multi-homing            : singleActive       Oper Multi-homing  : singleActive
ES SHG Label            : 524276
Source BMAC LSB         : <none>
Sdp Id                  : 56
ES Activation Timer     : 3 secs
Oper Group              : (Not Specified)
Svc Carving             : auto               Oper Svc Carving   : auto
Cfg Range Type          : primary
===============================================================================
```

For the remainder of the chapter, the vES configuration for "vESI-45_1" on PE-4 and PE-5 is restored to the initial settings: the behavior is revertive; PE-4 has the default preference, and PE-5 has preference 5000. When there are no failures, PE-4 is the DF, because it has a higher preference.

Revertive behavior

When SDP 64 fails on MTU-6, PE-4 becomes the NDF for ES "vESI-45_1" and PE-5 will be the DF instead, as follows. The failure is emulated by disabling the SDP on MTU-6.

```
# on MTU-6:
configure
    service
        sdp 64
            shutdown
```

When the PE is not a candidate DF because it cannot be used, the operational preference value equals 0, as follows:

```
*A:PE-4# show service system bgp-evpn ethernet-segment name "vESI-45_1"
                                            | match "DF Pref Election" post-lines 6
DF Pref Election Information
-------------------------------------------------------------------------------
Preference     Preference     Last Admin Change        Oper Pref      Do No
Mode           Value                                   Value          Preempt
-------------------------------------------------------------------------------
revertive      32767          04/22/2021 15:43:11      0              Disabled
-------------------------------------------------------------------------------
```

PE-5 is the only DF candidate in ES "vESI-45_1" for VPLS 1 :

```
*A:PE-4# show service system bgp-evpn ethernet-segment name "vESI-45_1" evi 1

===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
1             1             0                   no  04/22/2021 15:43:32
===============================================================================

===============================================================================
DF Candidates                           Time Added
-------------------------------------------------------------------------------
192.0.2.5                               04/22/2021 15:43:15
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

PE-5 is the DF in "vESI-45_1" for VPLS 1:

```
*A:PE-5# show service id 1 ethernet-segment
No sap entries

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                         Status
-------------------------------------------------------------------------------
```

```
56:1                  vESI-45_1                        DF
===============================================================================
No vxlan instance entries
```

The preference mode for this vES is revertive and the DF preference for PE-5 is 5000, as follows:

```
*A:PE-5# show service system bgp-evpn ethernet-segment name "vESI-45_1"
                                        | match "DF Pref Election" post-lines 6
DF Pref Election Information
-------------------------------------------------------------------------------
Preference     Preference     Last Admin Change        Oper Pref      Do No
Mode           Value                                   Value          Preempt
-------------------------------------------------------------------------------
revertive      5000           04/22/2021 15:43:19      5000           Disabled
-------------------------------------------------------------------------------
```

When the failure is restored, the system reverts and PE-4 will again be the DF for "vESI-45_1" in VPLS 1.

```
# on MTU-6:
configure
    service
        sdp 64
            no shutdown
```

```
*A:PE-4# show service id 1 ethernet-segment
No sap entries

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
46:1                  vESI-45_1                        DF
===============================================================================
No vxlan instance entries
```

Non-revertive behavior

When no failures have occurred, PE-4 is the DF for "vESI-45_2" because the lowest preference prevails for the configured EVI 2. The preference of PE-4 is 15000, which is lower than PE-5's preference of 20000.

```
*A:PE-4# show service id 2 ethernet-segment

No sap entries
===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
46:2                  vESI-45_2                        DF
===============================================================================
No vxlan instance entries
```

A failure is simulated as follows:

```
# on MTU-6:
configure
    service
```

```
        sdp 64
            shutdown
```

When SDP 64 on MTU-6 goes down, SDP 46 on PE-4 goes down which brings the vESs down on PE-4. PE-4 is no longer the DF for "vESI-45_2" and not even a DF candidate anymore. The operational preference value is 0.

```
*A:PE-4# show service system bgp-evpn ethernet-segment name "vESI-45_2"
                                       | match "DF Pref Election" post-lines 6
DF Pref Election Information
-------------------------------------------------------------------------------
Preference     Preference     Last Admin Change        Oper Pref      Do No
Mode           Value                                   Value          Preempt
-------------------------------------------------------------------------------
non-revertive  15000          04/22/2021 15:21:19      0              Disabled
-------------------------------------------------------------------------------
```

PE-5 becomes the DF for "vESI-45_2" in VPLS 2, as follows:

```
*A:PE-5# show service id 2 ethernet-segment
No sap entries

===============================================================================
SDP Ethernet-Segment Information
===============================================================================
SDP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
56:2                  vESI-45_2                        DF
===============================================================================
No vxlan instance entries
```

```
*A:PE-5# show service system bgp-evpn ethernet-segment name "vESI-45_2"
                                       | match "DF Pref Election" post-lines 6
DF Pref Election Information
-------------------------------------------------------------------------------
Preference     Preference     Last Admin Change        Oper Pref      Do No
Mode           Value                                   Value          Preempt
-------------------------------------------------------------------------------
non-revertive  20000          04/22/2021 15:21:27      20000          Enabled
-------------------------------------------------------------------------------
```

When the SDP is restored, the DF does not revert even though the list of DF candidates contains both PE-4 and PE-5. The preference mode is non-revertive; therefore, the DP bit has been set. PE-4 will not become the DF, as follows:

```
# on MTU-6:
configure
    service
        sdp 64
            no shutdown
```

```
*A:PE-4# show service system bgp-evpn ethernet-segment name "vESI-45_2" evi 2

===============================================================================
EVI DF and Candidate List
===============================================================================
EVI           SvcId         Actv Timer Rem      DF  DF Last Change
-------------------------------------------------------------------------------
2             2             0                   no  04/22/2021 15:46:25
```

```
===============================================================================

===============================================================================
DF Candidates                           Time Added
-------------------------------------------------------------------------------
192.0.2.4                               04/22/2021 15:47:32
192.0.2.5                               04/22/2021 15:46:02
-------------------------------------------------------------------------------
Number of entries: 2
===============================================================================
```

The operational preference value on NDF PE-4 equals the preference value on DF PE-5, as follows. In this example, EVI 2 is included in the configured EVI range, so the lowest preference wins. To avoid the system reverting to the lower preference of 15000, the operational preference is raised to the value of 20000, which equals the preference of the current DF PE-5.

```
*A:PE-4# show service system bgp-evpn ethernet-segment name "vESI-45_2"
                                       | match "DF Pref Election" post-lines 6
DF Pref Election Information
-------------------------------------------------------------------------------
Preference     Preference     Last Admin Change        Oper Pref      Do No
Mode           Value                                   Value          Preempt
-------------------------------------------------------------------------------
non-revertive  15000          04/22/2021 15:21:19      20000          Disabled
-------------------------------------------------------------------------------
```

PE-4 checks its own administrative preference and compares it with the one of the Highest-PE and Lowest-PE that have DP=1 in their ES routes.

- The Highest-PE is the PE with higher preference, using the DP bit (with DP=1 being better) and, after that, the lower PE-IP address as tie-breakers.
- The Lowest-PE is the PE with lower preference, using the DP bit (with DP=1 being better) and, after that, the lower PE-IP address as tie-breakers.

Depending on this comparison, PE-4 will send the ES route with a preference and DP that may be different from its administrative values.

- If PE-4's preference value is higher than the Highest-PE's, PE-4 will send the ES route with an 'in-use' operational preference equal to the Highest-PE's and DP=0.
- If PE-4's preference value is lower than the Lowest-PE's, PE-4 will send the ES route with an 'in-use' operational preference equal to the Lowest-PE's and DP=0.
- If PE-4's preference value is neither higher nor lower than the Highest-PE's or the Lowest-PE's respectively, PE-4 will send the ES route with its administrative [preference,DP]=[15000,1].

In this example, NDF PE-4 sends operational preference 20000 and DP=0, because its admin preference value was lower than the Lowest-PE's (PE-5), as follows:

```
# on PE-4:
195 2021/04/22 15:47:32.487 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 71
    Flag: 0x90 Type: 14 Len: 34 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.4
        Type: EVPN-ETH-SEG Len: 23 RD: 192.0.2.4:0
               ESI: 01:00:00:00:00:45:02:00:00:01, IP-Len: 4 Orig-IP-Addr: 192.0.2.4
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        df-election::DF-Type:Preference/DP:0/DF-Preference:20000/AC:1
        target:00:00:00:00:45:02
"
```

With equal operational preference, the current DF PE-5 sends DP=1, which is preferred over DP=0. The following output shows the BGP extended community of the ES routes for "vESI-45_2" in the RIB-In (received ES route from PE-5) and RIB-Out (sent ES route) on PE-4:

```
*A:PE-4# show router bgp routes evpn eth-seg hunt | match "target:00:00:00:00:45:02" pre-lines
 2
Community      :                                    # in RIB-In
                 df-election::DF-Type:Preference/DP:1/DF-Preference:20000/AC:1
                 target:00:00:00:00:45:02
Community      :                                    # in RIB-Out
                 df-election::DF-Type:Preference/DP:0/DF-Preference:20000/AC:1
                 target:00:00:00:00:45:02
```

Either of the following events cause PE-4 to re-advertise its admin preference 15000 and DP=1:

- DF PE-5 withdraws its ES route.
- The admin preference for ES "vESI-45_2" on DF PE-5 is modified by configuration to a value preferred over PE-4's admin preference; in this case, to a value lower than 15000.

The admin preference value can be modified on ES "vESI-45_2" on DF PE-5 on the fly, as follows:

```
# on PE-5:
configure
    service
        system
            bgp-evpn
                ethernet-segment "vESI-45_2" virtual create
                    service-carving
                        manual
                            preference non-revertive create
                                value 10000
                            exit
```

The preference value 10000 is lower than 15000 and, therefore, preferred when the lowest preference wins. PE-5 remains DF, but now there is no need to modify the preference of PE-4, because the system does not need to revert. Therefore, PE-4 can send the admin preference 15000 and configured DP=1, as follows:

```
# on PE-4:
198 2021/04/22 15:55:30.795 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 71
    Flag: 0x90 Type: 14 Len: 34 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 192.0.2.4
        Type: EVPN-ETH-SEG Len: 23 RD: 192.0.2.4:0
               ESI: 01:00:00:00:00:45:02:00:00:01, IP-Len: 4 Orig-IP-Addr: 192.0.2.4
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        df-election::DF-Type:Preference/DP:1/DF-Preference:15000/AC:1
        target:00:00:00:00:45:02
"
```

Conclusion

Preference-based DF election offers more control over the DF Election and applies to regular ESs and vESs, either in single-active or in all-active multi-homing mode, in VPLS, I-VPLS, or Epipe services. The DF election is by default revertive, but when preference mode is chosen, it can be configured as non-revertive to reduce service impact.

Proxy-ARP/ND MAC List for Dynamic Entries

This chapter provides information about Proxy-ARP/ND MAC List for Dynamic Entries.

Topics in this chapter include:

Applicability

This chapter was initially written based on SR OS Release 15.0.R4, but the CLI in the current edition is based on SR OS Release 21.2.R2. Proxy-Address Resolution Protocol/Neighbor Discovery (proxy-ARP/ND) MAC list for dynamic entries is supported in SR OS Release 15.0.R1, and later.

Overview

In some EVPN networks, the use of static proxy-ARP/ND entries is preferred to dynamically learned entries. For example, this is the case with some Internet eXchange Points (IXPs) that use EVPN and proxy-ARP/ND technologies. The MAC address in the static entry can be a MAC address from a list of n preregistered MAC addresses. The advantage is that—in case of a router or card failure—the hardware can be replaced, and no reconfiguration is required if the new MAC address is within a list of allowed MAC addresses.

In SR OS, these allow lists are called MAC lists. The associated proxy-ARP/ND entries will not be added upon configuration, but dynamically through a resolve procedure. This follows *draft-ietf-bess-evpn-proxy-arp-nd*.

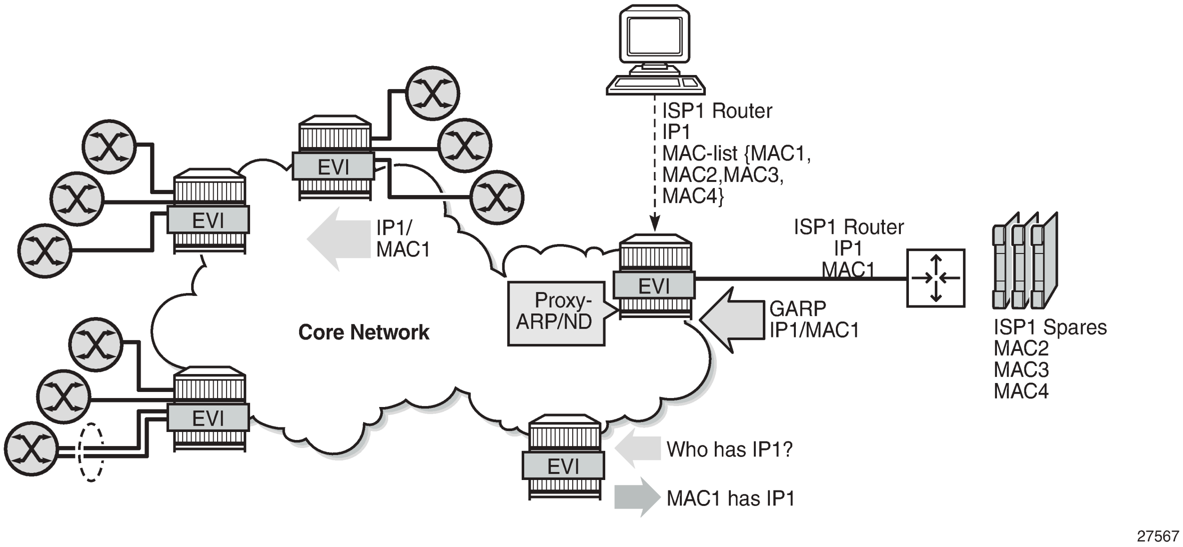
- When the dynamic proxy-ARP/ND IP address with its associated MAC list is configured, the system sends a resolve message to all its non-EVPN peers.
- The resolve message is an ARP request for IPv4, or a Neighbor Solicitation (NS) message for IPv6.
- The resolve message is sent at a configurable interval between 1 and 60 minutes; the default is 5 minutes.
- The system keeps sending resolve messages until a dynamic entry is created for the proxy-ARP/ND IP address. This entry is only created when two conditions are met:
  - An ARP/Gratuitous Address Resolution Protocol (GARP) or Neighbor Advertisement (NA) is received for the configured IP address.
  - The associated MAC address belongs to the MAC list configured for the IP address. If the MAC list is empty or not configured, the system will never create an entry for the IP address.

When the dynamic proxy-ARP/ND IP entry is created, the system advertises an EVPN-MAC update to its EVPN peers. The sticky bit will be set depending on how the corresponding MAC address is learned. If the

MAC address is learned on a SAP/SDP-binding with Auto-Learn MAC Protect (ALMP) enabled, the EVPN-MAC route will be advertised as static.

Figure 284: IXP with proxy-ARP/ND MAC list for dynamic entries shows an example of an IXP network that uses proxy-ARP/ND and a MAC list.

Figure 284: IXP with proxy-ARP/ND MAC list for dynamic entries



The ISP1 router with IP1 and MAC1 is connected to a PE in the core network that has proxy-ARP/ND enabled and a list of allowed MAC addresses. This MAC list contains four MAC addresses: MAC1 (for the hardware that is currently in use) and three MAC addresses for spares: MAC2, MAC3, and MAC4. The proxy-ARP/ND table will be populated as follows:

- The PE floods a resolve message for the configured IP address for proxy-ARP/ND to its non-EVPN peers.
- The ISP1 router that is connected to the network sends a GARP or ARP Reply message with IP1 and MAC1 that will be snooped by the PE.
- The PE checks whether IP1 is configured as a dynamic proxy-ARP/ND entry and MAC1 is in the MAC list assigned to proxy-ARP/ND entry IP1.
  - If true, the IP1/MAC1 entry is created in the proxy-ARP/ND table and advertised in EVPN.
  - If the GARP message contains MAC5, which is not in the MAC allow list, no proxy-ARP/ND entry is created, and IP/MAC is not advertised. If **no garp-flood-evpn** is configured, the GARP containing MAC5 will be discarded.

If after the proxy-ARP/ND creation, the corresponding MAC address is flushed from the Forwarding Database (FDB), the entry goes inactive. After the age-time, the inactive entry will age out and the resolve process will restart.

MAC lists are configured with the following command:

```
*A:PE-2>config>service>proxy-arp-nd# mac-list ?
  - mac-list <name> [create]
  - no mac-list <name>

 <name>               : [32 chars max]
 <create>             : keyword


 [no] mac             - Configure proxy ARP/ND MAC address information
```

The MAC list contains the allowed MAC addresses and can be associated in one or more services with a proxy-ARP/ND IP address. A MAC list is associated with dynamic proxy-ARP IP 1.1.1.1 with the following command:

```
*A:PE-2>config>service>vpls>proxy-arp# dynamic 1.1.1.1 create ?
  - dynamic <ip-address> [create]
  - no dynamic <ip-address>

 <ip-address>         : a.b.c.d
 <create>             : keyword


 [no] mac-list        - Configure MAC list
      resolve         - Configure address resolve time in minutes
```

The configuration for proxy-ND is similar:

```
*A:PE-2>config>service>vpls>proxy-nd# dynamic 2001:db8::99 create ?
  - dynamic <ipv6-address> [create]
  - no dynamic <ipv6-address>

 <ipv6-address>       : x:x:x:x:x:x:x:x   (eight 16-bit pieces)
                        x:x:x:x:x:x:d.d.d.d
                        x - [0..FFFF]H
                        d - [0..255]D
 <create>             : keyword


 [no] mac-list        - Configure MAC list
      resolve         - Configure address resolve time in minutes
```

- The MAC list can be associated with multiple configured dynamic IP addresses:
  - In different services
  - In the same service, for proxy-ARP and proxy-ND
- An empty MAC list can be configured and applied, but no proxy-ARP/ND entries will be created when the PE receives a GARP message containing a MAC address that is not in the allow list.
- MAC lists can be modified at any time: MAC addresses can be added or removed even when the MAC lists are associated with configured dynamic IP addresses. If the MAC list changes, all the IP addresses associated with that MAC list will delete the proxy entries and restart the resolve process.

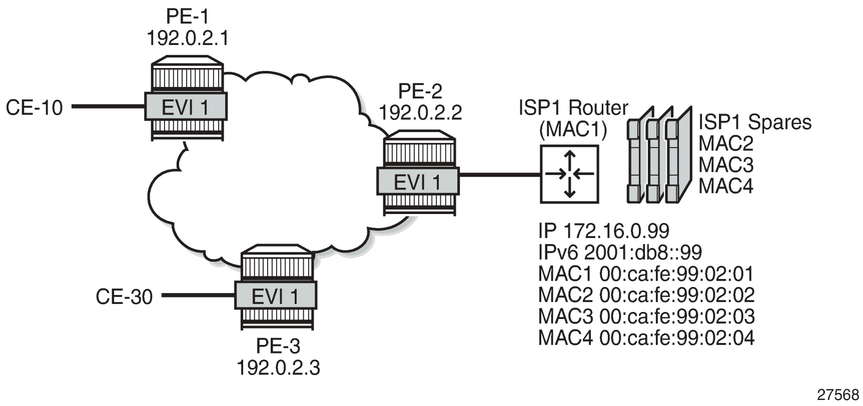
An existing dynamic proxy-ARP/ND entry IP1/MAC1 can be overridden when the system receives a GARP/ARP/NA for IP1 with another MAC address from the MAC list (IP1/MAC2). The system will first send a confirm message to check whether the old IP1/MAC1 is still reachable. Only when there is no answer, the entry IP1/MAC1 is replaced by IP1/MAC2. The existing dup-detect and confirm procedures are only applied for MAC address changes within the MAC list.

An existing dynamic proxy-ARP/ND entry IP1/MAC1 will be deleted when the system receives a GARP/ARP/NA IP1/MAC5 with a MAC address that is not contained in the MAC list. The GARP/ARP/NA message will be discarded and the resolve procedure is restarted.

Configuration

Figure 285: Example topology shows the example topology with three PEs. ISP router 1 is connected to PE-2. MAC1 is used; MAC2, MAC3, and MAC4 correspond to spares.

Figure 285: Example topology



The initial configuration includes:

- Cards, MDAs, ports
- Router interfaces
- IS-IS between the PEs (alternatively, OSPF can be used)
- LDP between the PEs

BGP is enabled between the PEs for address family EVPN. The BGP configuration on PE-2 is as follows:

```
# on PE-2:
configure
    router Base
        autonomous-system 64500
        bgp
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family evpn
                peer-as 64500
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.3
                exit
            exit
        exit
```

VPLS 1 is configured on PE-2 as follows. The configuration on the other PEs is similar.

```
# on PE-2:
configure
    service
        vpls 1 name "EVI-1" customer 1 create
```

```
            bgp
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    ingress-replication-bum-label
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/2/1:1 create
                no shutdown
            exit
            sap 1/2/1:3 create
                no shutdown
            exit
            no shutdown
        exit
```

MAC list

The following MAC lists are configured on PE-2: ISP1 is an empty list; ISP2 is a MAC list containing four MAC addresses.

```
# on PE-2:
configure
    service
        proxy-arp-nd
            mac-list "ISP1" create
            exit
            mac-list "ISP2" create
                mac 00:ca:fe:99:02:01
                mac 00:ca:fe:99:02:02
                mac 00:ca:fe:99:02:03
                mac 00:ca:fe:99:02:04
            exit
```

The following command shows the configured MAC lists on PE-2, with the number of MAC addresses and the number of associations. None of the MAC lists has been associated with a proxy-ARP/ND IP entry, so the number of associations is zero.

```
*A:PE-2# show service proxy-arp-nd mac-list

===============================================================================
MAC List Information
===============================================================================
MAC List Name                      Last Change         Num Macs    Num Assocs
-------------------------------------------------------------------------------
ISP1                               05/10/2021 14:25:45  0           0
ISP2                               05/10/2021 14:28:40  4           0
-------------------------------------------------------------------------------
Number of Entries: 2
===============================================================================
```

The following command shows the MAC addresses that are configured in MAC list ISP2. The timestamps show that all four MAC addresses were configured simultaneously, but MAC lists can be modified at any time.

```
*A:PE-2# show service proxy-arp-nd mac-list "ISP2"

===============================================================================
MAC List MAC Addr Information
===============================================================================
MAC Addr                                Last Change
-------------------------------------------------------------------------------
00:ca:fe:99:02:01                       05/10/2021 14:28:40
00:ca:fe:99:02:02                       05/10/2021 14:28:40
00:ca:fe:99:02:03                       05/10/2021 14:28:40
00:ca:fe:99:02:04                       05/10/2021 14:28:40
-------------------------------------------------------------------------------
Number of Entries: 4
===============================================================================
```

MAC list associated with proxy-ARP/ND in VPLS

MAC lists can be associated with one or more services. An empty MAC list—such as ISP1—can be associated, but it is impossible to associate a non-existing MAC list with a service. The following error is raised when attempting to associate the non-existing MAC list ISP3 with proxy-ARP IP 1.1.1.1 in VPLS 1 on PE-2:

```
*A:PE-2>config>service>vpls>proxy-arp>dynamic$ mac-list "ISP3"
MINOR: SVCMGR #8372 Cannot modify dynamic configured proxy arp entry - invalid mac-list
```

MAC list ISP2 is associated with proxy-ARP IP 172.16.0.99 and with proxy-ND IP 2001:db8::99 in VPLS 1 on PE-2, as follows:

```
# on PE-2:
configure
    service
        vpls "EVI-1"
            proxy-arp
                dynamic-arp-populate
                dynamic 172.16.0.99 create
                    mac-list ISP2
                    resolve 1
                exit
                no shutdown
            exit
            proxy-nd
                dynamic-nd-populate
                evpn-nd-advertise router
                dynamic 2001:db8::99 create
                    mac-list ISP2
                exit
                no shutdown
            exit
```

For proxy-ARP IP 172.16.0.99, the resolve interval is 1 minute, which is the minimum; for proxy-ND IP 2001:db::99, the resolve interval is the default of 5 minutes. In scaled environments, Nokia recommends using the default interval, or even configuring a longer interval. The proxy-ARP and proxy-ND tables can be populated with dynamic entries (**dynamic-arp-populate**/**dynamic-nd-populate**).

The following command shows all associations for MAC list ISP2: two associations are defined in VPLS 1: one for IP address 172.16.0.99 and another for IP address 2001:db8::99.

```
*A:PE-2# show service proxy-arp-nd mac-list "ISP2" associations

===============================================================================
MAC List Associations
===============================================================================
Service Id                    IP Addr
-------------------------------------------------------------------------------
1                             172.16.0.99
1                             2001:db8::99
-------------------------------------------------------------------------------
Number of Entries: 2
===============================================================================
```

Different dynamic proxy-ARP/ND entries

A distinction is made between regular dynamic entries and configured dynamic entries:

- No IP address needs to be configured for regular dynamic proxy-ARP/ND entries. What only needs to be configured, is the option **dynamic-arp-populate**/**dynamic-nd-populate**.
- IP address and MAC list need to be defined for configured proxy-ARP/ND entries.

Configured dynamic entries can override static and regular dynamic entries.

Regular dynamic proxy-ARP/ND entries can override configured dynamic entries.

EVPN entries cannot override configured dynamic entries, even though they can override regular dynamic entries.

Likewise, static entries can override regular dynamic entries, but they cannot override dynamic configured entries. The following error is raised when attempting to configure a static proxy-ARP entry for IP 172.16.0.99, which has already been configured as dynamic and associated with a MAC list.

```
*A:PE-2>config>service>vpls>proxy-arp# static 172.16.0.99 aa:bb:cc:99:02:02
MINOR: SVCMGR #8011 Cannot create static proxy arp entry - Dynamic configured entry exists
```

Debugging

Debugging for both proxy-ARP/ND IP entries is enabled on PE-2 as follows:

```
# on PE-2:
debug
    service
        id 1
            proxy-arp ip 172.16.0.99
            proxy-nd ip 2001:db8::99
        exit
    exit
exit
```

When the dynamic proxy-ARP IP 172.16.0.99 is configured with MAC list "ISP2", PE-2 floods a resolve message—in this case, an ARP request—to all its EVPN peers. Router ISP1 replies. PE-2 advertises an EVPN-MAC update to its EVPN peers PE-1 and PE-3. PE-2 adds a dynamic proxy-ARP entry for

172.16.0.99 with MAC address 00:ca:fe:99:02:01. Router ISP1 sends a GARP message. The following messages are logged:

```
49 2021/05/10 14:32:18.859 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.99 flood resolve"

50 2021/05/10 14:32:18.862 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.99 mac: 00:ca:fe:99:02:01 evpn advertise"

51 2021/05/10 14:32:18.862 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.99 type: Dyn mac: 00:ca:fe:99:02:01 Added"

54 2021/05/10 14:32:18.959 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.99 type: Dyn mac: 00:ca:fe:99:02:01 Gratuitous Update"
```

For proxy-ND, the following messages are logged:

```
55 2021/05/10 14:32:19.859 UTC MINOR: DEBUG #2001 Base proxy nd
"proxy nd:
svc: 1 ip: 2001:db8::99 flood resolve"

56 2021/05/10 14:32:19.862 UTC MINOR: DEBUG #2001 Base proxy nd
"proxy nd:
svc: 1 ip: 2001:db8::99 mac: 00:ca:fe:99:02:01 evpn advertise"

57 2021/05/10 14:32:19.862 UTC MINOR: DEBUG #2001 Base proxy nd
"proxy nd:
svc: 1 ip: 2001:db8::99 type: Dyn mac: 00:ca:fe:99:02:01 Added"

60 2021/05/10 14:32:19.959 UTC MINOR: DEBUG #2001 Base proxy nd
"proxy nd:
svc: 1 ip: 2001:db8::99 type: Dyn mac: 00:ca:fe:99:02:01 Gratuitous Update"
```

The following command shows the proxy-ARP details for VPLS 1 on PE-2. The only proxy-ARP entry is for IP address 172.16.0.99 with MAC address 00:ca:fe:99:02:01.

```
*A:PE-2# show service id 1 proxy-arp detail
-------------------------------------------------------------------------------
Proxy Arp
-------------------------------------------------------------------------------
Admin State       : enabled
Dyn Populate      : enabled
Age Time          : disabled            Send Refresh      : disabled
Table Size        : 250                 Total             : 1
Static Count      : 0                   EVPN Count        : 0
Dynamic Count     : 1                   Duplicate Count   : 0

Dup Detect
-------------------------------------------------------------------------------
Detect Window     : 3 mins              Num Moves         : 5
Hold down         : 9 mins
Anti Spoof MAC    : None

EVPN
-------------------------------------------------------------------------------
Garp Flood        : enabled             Req Flood         : enabled
Static Black Hole : disabled
EVPN Route Tag    : 0
```

```
-------------------------------------------------------------------------------

===============================================================================
VPLS Proxy Arp Entries
===============================================================================
IP Address          Mac Address         Type      Status    Last Update
-------------------------------------------------------------------------------
172.16.0.99         00:ca:fe:99:02:01   dyn       active    05/10/2021 14:32:19
-------------------------------------------------------------------------------
Number of entries : 1
===============================================================================
```

The following command shows the proxy-ND details for VPLS 1 on PE-2. The only proxy-ND entry if for IP address 2001:db8::99 with MAC address 00:ca:fe:99:02:01.

```
*A:PE-2# show service id 1 proxy-nd detail
-------------------------------------------------------------------------------
Proxy ND
-------------------------------------------------------------------------------
Admin State       : enabled
Dyn Populate      : enabled
Age Time          : disabled            Send Refresh      : disabled
Table Size        : 250                 Total             : 1
Static Count      : 0                   EVPN Count        : 0
Dynamic Count     : 1                   Duplicate Count   : 0

Dup Detect
-------------------------------------------------------------------------------
Detect Window     : 3 mins              Num Moves         : 5
Hold down         : 9 mins
Anti Spoof MAC    : None

EVPN
-------------------------------------------------------------------------------
Unknown NS Flood  : enabled             ND Advertise      : Router
Rtr Unsol NA Flood: enabled             Host Unsol NA Fld : enabled
EVPN Route Tag    : 0
-------------------------------------------------------------------------------

===============================================================================
VPLS Proxy ND Entries
===============================================================================
IP Address               Mac Address       Type Status Rtr/ Last Update
                                                       Host
-------------------------------------------------------------------------------
2001:db8::99             00:ca:fe:99:02:01 dyn  active Rtr  05/10/2021 14:32:20
-------------------------------------------------------------------------------
Number of entries : 1
===============================================================================
```

The proxy-ARP in VPLS 1 contains the following dynamic entry.

```
*A:PE-2# show service id 1 proxy-arp dynamic

===============================================================================
Proxy ARP Dyn Cfg Summary
===============================================================================
IP Addr                                 Mac List
-------------------------------------------------------------------------------
172.16.0.99                             ISP2
-------------------------------------------------------------------------------
Number of Entries: 1
```

```
===============================================================================
```

The following command shows the association for dynamic proxy-ARP IP address 172.16.0.99, with the configured resolve time in minutes and the remaining resolve time in seconds.

```
*A:PE-2# show service id 1 proxy-arp dynamic 172.16.0.99

===============================================================================
Proxy ARP Dyn Cfg Detail
===============================================================================
IP Addr        Mac List                           Resolve Time   Remaining
                                                  (mins)          Resolve Time
                                                                 (secs)
-------------------------------------------------------------------------------
172.16.0.99    ISP2                               1              0
-------------------------------------------------------------------------------
Number of Entries: 1
===============================================================================
```

The remaining resolve time is zero seconds because a dynamic proxy-ARP entry has been created and that suspends the resolve mechanism.

The proxy-ND in VPLS 1 contains the following dynamic entry.

```
*A:PE-2# show service id 1 proxy-nd dynamic

===============================================================================
Proxy ND Dyn Cfg Summary
===============================================================================
IP Addr                                 Mac List
-------------------------------------------------------------------------------
2001:db8::99                            ISP2
-------------------------------------------------------------------------------
Number of Entries: 1
===============================================================================
```

The following command shows the association for dynamic proxy-ND IP 2001:db8::99.

```
*A:PE-2# show service id 1 proxy-nd dynamic 2001:db8::99

===============================================================================
Proxy ND Dyn Cfg Detail
===============================================================================
IP Addr                                 Mac List
 Resolve Time(mins)                      Remaining Resolve Time(secs)
-------------------------------------------------------------------------------
2001:db8::99                            ISP2
5                                       0
-------------------------------------------------------------------------------
Number of Entries: 1
===============================================================================
```

Tools command to trigger resolve procedure

The following tools command can be used to force the system to send a resolve message to its non-EVPN peers. The **force** option will trigger the resolve process even for existing entries in the proxy-ARP/ND table.

```
*A:PE-2# tools perform service id 1 proxy-arp dynamic-resolve ?
```

```
  - dynamic-resolve all [force]
  - dynamic-resolve <ip-address> [force]

 <ip-address>         : a.b.c.d
 <all>                : keyword
 <force>              : keyword
```

```
*A:PE-2# tools perform service id 1 proxy-nd dynamic-resolve ?
  - dynamic-resolve all [force]
  - dynamic-resolve <ipv6-address> [force]

 <ipv6-address>       : x:x:x:x:x:x:x:x   (eight 16-bit pieces)
                        x:x:x:x:x:x:d.d.d.d
                        x - [0..FFFF]H
                        d - [0..255]D
 <all>                : keyword
 <force>              : keyword
```

Some examples:

```
*A:PE-2# tools perform service id 1 proxy-arp dynamic-resolve 172.16.0.99
*A:PE-2# tools perform service id 1 proxy-arp dynamic-resolve 172.16.0.99 force
*A:PE-2# tools perform service id 1 proxy-arp dynamic-resolve all
*A:PE-2# tools perform service id 1 proxy-arp dynamic-resolve all force

*A:PE-2# tools perform service id 1 proxy-nd dynamic-resolve 2001:db8::99
*A:PE-2# tools perform service id 1 proxy-nd dynamic-resolve 2001:db8::99 force
*A:PE-2# tools perform service id 1 proxy-nd dynamic-resolve all
*A:PE-2# tools perform service id 1 proxy-nd dynamic-resolve all force
```

Inactive proxy-ARP/ND entries

When the MAC address is flushed from the FDB, the proxy-ARP/ND entries become inactive.

```
*A:PE-2# clear service id 1 fdb mac 00:ca:fe:99:02:01
```

```
*A:PE-2# show service id 1 proxy-arp detail | match 172.16.0.99 pre-lines 6
                                                                post-lines 3

===============================================================================
VPLS Proxy Arp Entries
===============================================================================
IP Address          Mac Address         Type      Status    Last Update
-------------------------------------------------------------------------------
172.16.0.99         00:ca:fe:99:02:01   dyn       inActv    05/10/2021 14:34:40
-------------------------------------------------------------------------------
Number of entries : 1
===============================================================================
```

```
*A:PE-2# show service id 1 proxy-nd detail | match 2001:db8::99 pre-lines 7
                                                                post-lines 3

===============================================================================
VPLS Proxy ND Entries
===============================================================================
IP Address               Mac Address       Type Status Rtr/ Last Update
                                                       Host
```

```
-------------------------------------------------------------------------------
2001:db8::99             00:ca:fe:99:02:01 dyn  inActv Rtr  05/10/2021 14:34:40
-------------------------------------------------------------------------------
Number of entries : 1
===============================================================================
```

By default, aging is disabled, and the entries remain in the inactive status until the MAC address is learned again. However, if aging is enabled, the inactive proxy-ARP/ND entry will age out. After the entry is deleted, the system sends a resolve message. When the ISP1 router replies, the entry is created again in the proxy-ARP/ND table. The age time is configured in seconds with the following command:

```
*A:PE-2>config>service>vpls>proxy-arp# age-time ?
  - age-time <seconds>
  - no age-time

 <seconds>            : [60..86400]
```

```
# on PE-2:
configure
    service
        vpls "EVI-1"
            proxy-arp
                age-time 60
```

The following debug messages for proxy ARP IP 172.16.0.99 show that an EVPN-MAC withdraw message is sent (when the MAC address is flushed from the FDB) and—after time-out—the proxy-ARP entry is deleted. PE-2 sends a resolve message to all its non-EVPN peers. Router ISP1 replies and the proxy-ARP entry is created again; an EVPN-MAC update is sent to the EVPN peers. Similar debug messages occur for proxy-ND.

```
79 2021/05/10 14:34:47.107 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.99 mac: 00:ca:fe:99:02:01 evpn withdraw"

86 2021/05/10 14:36:18.359 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.99 type: Dyn mac: 00:ca:fe:99:02:01 Deleted"

88 2021/05/10 14:36:18.459 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.99 flood resolve"

89 2021/05/10 14:36:18.462 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.99 mac: 00:ca:fe:99:02:01 evpn advertise"

90 2021/05/10 14:36:18.462 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.99 type: Dyn mac: 00:ca:fe:99:02:01 Added"

95 2021/05/10 14:36:18.559 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.99 type: Dyn mac: 00:ca:fe:99:02:01 Gratuitous Update"
```

The following command shows that the entry is created again with active status.

```
*A:PE-2# show service id 1 proxy-arp detail | match 172.16.0.99 pre-lines 6
                                                                post-lines 3
```

```
===============================================================================
VPLS Proxy Arp Entries
===============================================================================
IP Address          Mac Address         Type      Status    Last Update
-------------------------------------------------------------------------------
172.16.0.99         00:ca:fe:99:02:01   dyn       active    05/10/2021 14:37:19
-------------------------------------------------------------------------------
Number of entries : 1
===============================================================================
```

MAC address replacement

When the system receives a GARP/ARP/NA for the same IP address, but with another MAC address from the MAC list, it will first send a confirm message to ensure that the old MAC address is not used anymore for the IP address. If the existing proxy-ARP/ND entry is IP1/MAC1 and a GARP/ARP/NA message is received for IP1/MAC4, the system sends an EVPN-MAC withdraw message for MAC1 and changes MAC1 to MAC4 for proxy-ARP/ND IP1, but the status is pending (pendng), as follows:

```
*A:PE-2# show service id 1 proxy-arp detail | match 172.16.0.99 pre-lines 6
                                                                post-lines 3

===============================================================================
VPLS Proxy Arp Entries
===============================================================================
IP Address          Mac Address         Type      Status    Last Update
-------------------------------------------------------------------------------
172.16.0.99         00:ca:fe:99:02:04   dyn       pendng    05/10/2021 14:37:37
-------------------------------------------------------------------------------
Number of entries : 1
===============================================================================
```

```
*A:PE-2# show service id 1 proxy-nd detail | match 2001:db8::99 pre-lines 7
                                                                post-lines 3

===============================================================================
VPLS Proxy ND Entries
===============================================================================
IP Address               Mac Address       Type Status Rtr/ Last Update
                                                       Host
-------------------------------------------------------------------------------
2001:db8::99             00:ca:fe:99:02:04 dyn  pendng Rtr  05/10/2021 14:37:36
-------------------------------------------------------------------------------
Number of entries : 1
===============================================================================
```

The system sends a confirm message (unicast ARP request) for the old entry IP1/MAC1 to ensure that there is no duplication. When there is no reply from MAC1, there is no duplication. An EVPN-MAC route is advertised for MAC4. The status of the proxy-ARP entry IP1/MAC4 changes to active. The following debug messages are logged for proxy-ARP 172.16.0.99:

```
113 2021/05/10 14:37:34.570 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.99 mac: 00:ca:fe:99:02:01 evpn withdraw"

114 2021/05/10 14:37:34.570 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.99 Mac Change: 00:ca:fe:99:02:01->00:ca:fe:99:02:04 "
```

```
121 2021/05/10 14:37:34.759 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.99 mac: 00:ca:fe:99:02:01 confirm"

124 2021/05/10 14:38:04.759 UTC MINOR: DEBUG #2001 Base proxy arp
"proxy arp:
svc: 1 ip: 172.16.0.99 mac: 00:ca:fe:99:02:04 evpn advertise"
```

The final status of the proxy-ARP IP 172.16.0.99 is active, as follows:

```
*A:PE-2# show service id 1 proxy-arp detail | match 172.16.0.99 pre-lines 6
                                                                post-lines 3
===============================================================================
VPLS Proxy Arp Entries
===============================================================================
IP Address          Mac Address         Type      Status    Last Update
-------------------------------------------------------------------------------
172.16.0.99         00:ca:fe:99:02:04   dyn       active    05/10/2021 14:37:37
-------------------------------------------------------------------------------
Number of entries : 1
===============================================================================
```

The mechanism is similar for proxy-ND.

The behavior is different when the system receives a GARP/ARP/NA for the IP address with a MAC address that is not contained in the MAC list. The GARP/ARP/NA message is discarded and the proxy-ARP/ND entry deleted. The resolve procedure gets restarted.

Modified MAC list

MAC lists can be modified at any time, as follows:

```
# on PE-2:
configure
    service
        proxy-arp-nd
            mac-list "ISP2" create
                mac 00:ca:fe:99:02:05
```

```
*A:PE-2# show service proxy-arp-nd mac-list "ISP2"

===============================================================================
MAC List MAC Addr Information
===============================================================================
MAC Addr                                Last Change
-------------------------------------------------------------------------------
00:ca:fe:99:02:01                       05/10/2021 14:28:40
00:ca:fe:99:02:02                       05/10/2021 14:28:40
00:ca:fe:99:02:03                       05/10/2021 14:28:40
00:ca:fe:99:02:04                       05/10/2021 14:28:40
00:ca:fe:99:02:05                       05/10/2021 14:39:29
-------------------------------------------------------------------------------
Number of Entries: 5
===============================================================================
```

The timestamps show when the different MAC addresses were added to the MAC list.

When the MAC list ISP2 is modified, proxy-ARP entry 172.16.0.99 and proxy-ND entry 2001:db8::99 will be deleted, an EVPN-MAC withdraw message will be sent, and the resolve procedure will be restarted. The following log messages occur for proxy-ND 2001:db8::99.

```
146 2021/05/10 14:39:29.205 UTC MINOR: DEBUG #2001 Base proxy nd
"proxy nd:
svc: 1 ip: 2001:db8::99 mac: 00:ca:fe:99:02:04 evpn withdraw"

147 2021/05/10 14:39:29.205 UTC MINOR: DEBUG #2001 Base proxy nd
"proxy nd:
svc: 1 ip: 2001:db8::99 type: Dyn mac: 00:ca:fe:99:02:04 Deleted"

151 2021/05/10 14:39:29.359 UTC MINOR: DEBUG #2001 Base proxy nd
"proxy nd:
svc: 1 ip: 2001:db8::99 flood resolve"

154 2021/05/10 14:39:29.362 UTC MINOR: DEBUG #2001 Base proxy nd
"proxy nd:
svc: 1 ip: 2001:db8::99 mac: 00:ca:fe:99:02:04 evpn advertise"

155 2021/05/10 14:39:29.362 UTC MINOR: DEBUG #2001 Base proxy nd
"proxy nd:
svc: 1 ip: 2001:db8::99 type: Dyn mac: 00:ca:fe:99:02:04 Added"

159 2021/05/10 14:39:29.459 UTC MINOR: DEBUG #2001 Base proxy nd
"proxy nd:
svc: 1 ip: 2001:db8::99 type: Dyn mac: 00:ca:fe:99:02:04 Gratuitous Update"
```

Conclusion

MAC lists can be associated with configured dynamic proxy-ARP/ND IP addresses. The actual proxy entries will only be created after a GARP/ARP/NA message is received for the IP address and one of the MAC addresses from the MAC list.

This tool complements the SR OS EVPN proxy-ARP/ND solution for providers present at IXPs.

Shortest Path Bridging for MAC

This chapter describes advanced shortest path bridging for MAC configurations.

Topics in this chapter include:

Applicability

This chapter was initially written for SR OS Release 11.0.R4, but the CLI in the current edition is based on SR OS Release 23.7.R2.

Overview

SPB enables a next generation control plane for Provider Backbone Bridges (PBB) and PBB-VPLS that adds the stability and efficiency of link state to unicast and multicast services (Epipes and I-VPLSs). In addition, SPBM provides resiliency, load balancing, and multicast optimization without the need for any other control plane in the B-VPLS (for example, there is no need for spanning tree, or G.8032, or Multiple MAC Registration Protocol (MMRP)).

SPBM exploits the complete knowledge of backbone addressing, which is a key consequence of the PBB hierarchy, by advertising and distributing the backbone MAC addresses (BMACs) through a link-state protocol, namely IS-IS. An immediate effect of this is that the old “flood-and-learn” can at last be turned off in the backbone and every B-VPLS node in the network will know what destination BMAC addresses are expected and valid. As a result of that, receiving an unknown unicast BMAC on a B-VPLS SAP or PW is indicative of an error, whereupon the frame is discarded (due to the Reverse Path Forwarding Check (RPFC) performed in SPBM) instead of flooded. Furthermore, SPBM allows condensing all the relevant information distribution (unicast and multicast) into a single control protocol: IS-IS.

SPBM can be easily enabled on the existing B-VPLS instances being used for multiplexing I-VPLS and Epipe services, providing the following benefits:

- Per-service flood containment (for I-VPLS services) without the need for an additional protocol such as MMRP,
- Loop avoidance in the B-VPLS domain without the need for MSTP or other technologies,
- No unknown BMAC flooding in the B-VPLS domain,
- No need for MAC notification mechanisms or vMEPs in the B-VPLS to update the B-VPLS forwarding databases (FDBs) (vMEPs can still be configured though for OAM purposes).

Some other characteristics of the SPB implementation in the SR OS are:

- The SR OS SPB implementation always uses Multi-Topology (MT) topology instance zero. However, up to four logical instances (that is, SPB instances in different B-VPLS services) are supported if different topologies are required for different services.
- Area addresses are not used and SPB is assumed to be a single area. SPB must be consistently configured on nodes in the system. SPB regions information and IS-IS hello logic that detect mismatched configuration are not supported. IS-IS area is always zero.
- SPB uses all-intermediate systems 09-00-2B-00-00-05 destination MAC to communicate.
- SPB source ID is always zero.
- SPB uses a separate instance of IS-IS from the base IP IS-IS. IS-IS for SPB is configured in the SPB context under the B-VPLS component. Up to four ISIS-SPB instances are supported, where the instance identifier can be any number between 1024 and 2047. The instance number is not in TLVs.
- Two Equal Cost Tree (ECT) algorithms (IEEE 802.1aq) per SPB instance are supported: low-path-id and high-path-id algorithms.
- SPB link state protocol data units (link state packets) contain BMACs, ISIDs (for multicast services) and link and metric information for an IS-IS database.
  - Epipe ISIDs are not distributed in SR OS SPB allowing high scalability of PBB Epipes.
  - I-VPLS ISIDs are distributed in SR OS SPB and the respective multicast group addresses (composed of PBB-OUI plus ISID) are automatically populated in a manner that provides automatic pruning of multicast to the subset of the multicast tree that supports an I-VPLS with a common ISID. This replaces the function of MMRP and is more efficient than MMRP.
- Multiple ISIS-SPB adjacencies between two nodes are not supported as per the IEEE 802.1aq standard specification. If multiple links between two nodes exist, LAG must be used.

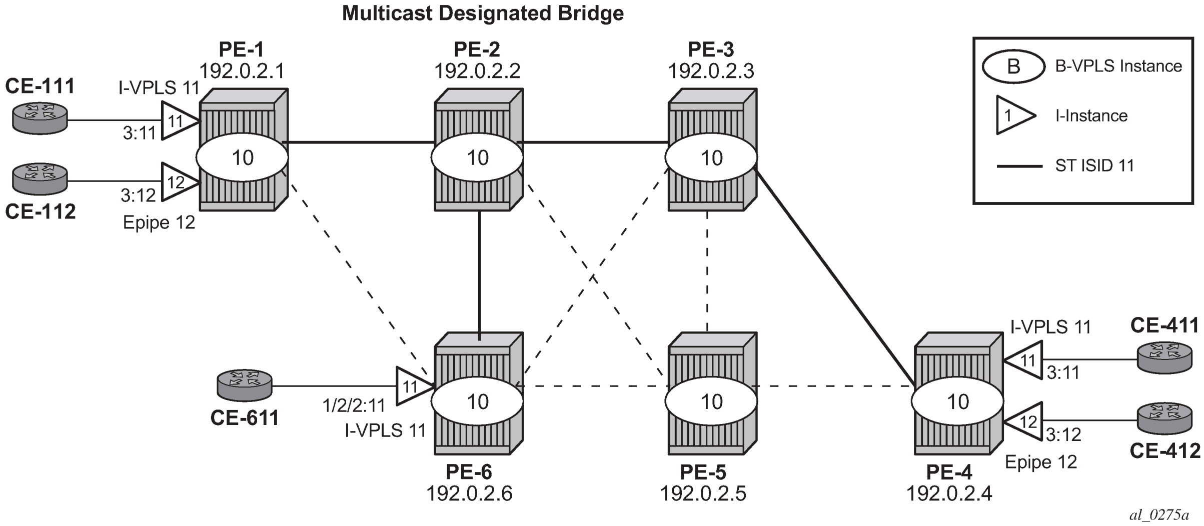
Configuration

This section describes the configuration of SPBM on SR OS as well as the available troubleshooting commands.

Basic SPBM configuration

Figure 286: Basic SPBM topology shows the topology used as an example of a basic SPBM configuration.

Figure 286: Basic SPBM topology



Assume the following protocols and objects are configured beforehand:

- The six PEs shown in Figure 286: Basic SPBM topology are running IS-IS for the global routing table with all the interfaces being level-2.
- LDP is used as the MPLS protocol to signal transport tunnel labels.
- LDP SDPs are configured among the six PEs, as shown in Figure 286: Basic SPBM topology (dashed lines and bold lines among PEs).

Once the network infrastructure is properly running, the actual service configuration can be carried out. In the example, B-VPLS-10 will provide backbone connectivity for the services I-VPLS-11 and Epipe-12.

The SPBM configuration is only relevant to the B-VPLS instance and can be added to an existing B-VPLS, assuming that such a B-VPLS does not contain any non-SPB-compatible configuration parameters. The following parameters are not supported in SPB-enabled B-VPLS instances:

- Mesh SDPs (only SAPs or spoke-SDPs are supported in SPB-enabled B-VPLS)
- Spanning tree protocol (STP)
- Split-horizon groups
- Non-conditional static-MAC addresses (configured under SAPs or spoke-SDPs, see the Static BMACs and static ISIDs configuration section)
- G.8032
- Propagate-mac-flush and send-flush-on-failure
- Maximum number of MAC addresses (max-nbr-mac-addr)
- Bridge Protocol Data Unit (BPDU) translation
- Layer 2 Protocol Termination (L2PT)
- MAC-pinning
- Operational groups
- MAC-move

- Any BGP, BGP auto-discovery (BGP-AD), or BGP virtual private LAN services (BGP-VPLS) parameters
- Endpoints
- Local/remote age
- MAC notification
- MAC protect
- Multiple MAC Registration Protocol (MMRP)
- Provider tunnel
- Temporary flooding

Assuming all the parameters mentioned are not configured in the B-VPLS (B-VPLS-10 in the example), SPBM can be enabled. The SPBM parameters are all configured in the **configure service vpls(b-vpls) spb** and **configure service vpls(b-vpls) spoke-sdp/sap spb** contexts:

```
*A:PE-1>config>service>vpls$ spb ?
  - no spb
  - spb [<isis-instance>] [fid <fid>] [create]

 <isis-instance>      : [1024..2047]
 <fid>                : [1..4095]


      level           + Configure SPB level information
 [no] lsp-lifetime    - Configure LSP lifetime
 [no] lsp-refresh-in* - Configure LSP refresh interval
 [no] overload        - Configure the local router so that it appears to be overloaded
 [no] overload-on-bo* - Configure the local router so that it appears to be overloaded at
                        boot up
 [no] shutdown        - Administratively enable or disable the operation of ISIS
      timers          + Configure ISIS timers
```

```
*A:PE-1>config>service>vpls>spb$ timers ?
  - timers

 [no] lsp-wait        - Configure ISIS LSP wait times
 [no] spf-wait        - Configure ISIS SPF wait times
```

```
*A:PE-1>config>service>vpls>spoke-sdp$ spb ?
  - no spb
  - spb [create]

 <create>             : keyword


      level           + Configure SPB level information
 [no] lsp-pacing-int* - Configure the interval for sending LSPs from the interface
 [no] retransmit-int* - Configure the minimum interval between LSP packets retransmission
                        for the given interface
 [no] shutdown        - Administratively Enable/disable the interface
```

```
*A:PE-1>config>service>vpls>spoke-sdp>spb$ level 1 ?
  - level <[1..1]>


 [no] hello-interval  - Configure hello-interval for this interface
 [no] hello-multipli* - Configure hello-multiplier for this level
```

```
[no] metric          - Configure IS-IS interface metric for IPv4 unicast
```

The parameters configured in the **spb** context refer to the SPB IS-IS and they should be configured following the same considerations as for the IS-IS base instance:

- spb [<isis-instance>] [fid <fid>] [create]
  - <isis-instance> identifies the SPB IS-IS process. Up to four different IS-IS SPB processes can be run in a system (range 1024 to 2047).
  - forwarding identifier <fid> identifies the standard SPBM B-VID which is signaled in IS-IS with each advertised BMAC. Each B-VPLS has a single configurable FID.
- spb lsp-lifetime <seconds> : [350..65535]
- spb lsp-refresh-interval <seconds> : [150..65535]
- spb overload [timeout <seconds>] : [60..1800]
- spb overload-on-boot [timeout <seconds>] : [60..1800]
- spb timers lsp-wait <lsp-wait> [<lsp-initial-wait> [<lsp-second-wait>]]
  - <lsp-wait> : [10..120000] in milliseconds
  - <lsp-initial-wait> : [10..100000] in milliseconds
  - <lsp-second-wait> : [10..100000] in milliseconds
- spb timers spf-wait <spf-wait> [<spf-initial-wait> [<second-wait>]]
  - <spf-wait> : [10..120000] in milliseconds
  - <spf-initial-wait> : [10..100000] in milliseconds
  - <second-wait> : [10..100000] in milliseconds
- spoke-sdp/sap spb lsp-pacing-interval <milli-seconds> : [0..65535]
- spoke-sdp/sap spb retransmit-interval <seconds> : [1..65535]
- spoke-sdp/sap spb level 1 hello-interval <seconds> : [1..20000]
- spoke-sdp/sap spb level 1 hello-multiplier <multiplier> : [2..100]

In the same way, lsp-wait (initial-wait) and spf-wait (initial-wait) can be tuned in the base router IS-IS instance to minimize the convergence time (to 0 and 10 respectively), the equivalent SPB IS-IS parameters should also be adjusted so that failover time is minimized at the service level.

The following parameters are specific to SPBM (note that only IS-IS level 1 is supported for SPB):

- **spb level 1 bridge-priority <bridge-priority> : [0..15]**

  This parameter influences the election of the multicast designated bridge through which all the Single Trees (STs) for the multicast traffic are established. The default value will be lowered on that node where the multicast designated bridge function is desired, normally because that node is the best connected node. In the example, PE-2 is the multicast designated bridge for B-VPLS-10 and therefore, PE-2 will be the root of the STs for the I-VPLS instances in that B-VPLS. Default value = 8.
- **spb level 1 ect-algorithm fid-range <fid-range> {low-path-id|high-path-id}**

  This command defines the ECT algorithm used and the FIDs assigned. Two algorithms are supported: low-path-id and high-path-id. They can provide the required path diversity for an efficient load balancing in the B-VPLS. Default = fid-range 1-4095 low-path-id
- **spb level 1 forwarding-tree-topology unicast {spf|st}**

This command configures the type of tree to be used for unicast traffic: shortest path tree or single tree. The multicast traffic (that encapsulated I-VPLS Broadcast, Unknown unicast, and Multicast (BUM) traffic always uses the ST path. Using SPF for unicast traffic can produce some packet re-ordering for unicast traffic compared to BUM traffic because different trees are used, therefore, when the B-VPLS transports I-VPLS traffic and the unicast and multicast trees do not follow the same path, it is recommended to use ST paths for unicast and multicast. Default value = spf.

- **spoke-sdp/sap spb level 1 metric <ipv4-metric> : [1..16777215]**

  This command configures the metric for each SPB interface (spoke-SDP or SAP). This value helps influence the SPF calculation in order to pick a certain path for the traffic to a remote system BMAC.When the SPB link metric advertised by two peers is different, the maximum value is chosen according to the RFC 6329. Default: no metric.

As an example, the following CLI output shows the relevant configuration of PE-1 and PE-2 (the multicast designated bridge). SPB has to be created and enabled at B-VPLS service level first and then created and enabled under every SAP or spoke-SDP in the B-VPLS. Non-SPB-enabled SAPs or spoke-SDPs can exist in the SPB B-VPLS only if conditional static-MACs are configured for them (see the Static BMACs and static ISIDs configuration section). As for regular B-VPLS services, the service MTU has to be changed from the default value (1500) to a number 18 bytes greater than the I-VPLS service MTU in order to allow for the PBB encapsulation.

```
# on PE-1:
configure
    service
        pbb
            source-bmac 00-00-5e-00-53-01
            mac-name "PE-1" 00-00-5e-00-53-01
            mac-name "PE-2" 00-00-5e-00-53-02
            mac-name "PE-3" 00-00-5e-00-53-03
            mac-name "PE-4" 00-00-5e-00-53-04
            mac-name "PE-5" 00-00-5e-00-53-05
            mac-name "PE-6" 00-00-5e-00-53-06
        exit
        vpls 10 name "B-VPLS-10" customer 1 b-vpls create
            service-mtu 2000
            stp
                shutdown
            exit
            spb 1024 fid 10 create
                overload-on-boot timeout 60
                timers
                    spf-wait 2000 spf-initial-wait 50000 spf-second-wait 100000
                    lsp-wait 8000 lsp-initial-wait 10 lsp-second-wait 1000
                exit
                no shutdown
            exit
            spoke-sdp 12:10 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 16:10 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            no shutdown
        exit
        vpls 11 name "I-VPLS-11" customer 1 i-vpls create
            pbb
                backbone-vpls 10 isid 11
                exit
            exit
            stp
                shutdown
            exit
            sap 1/1/c3/1:11 create
                no shutdown
            exit
            no shutdown
        exit
        epipe 12 name "Epipe-12" customer 1 create
            pbb
                tunnel 10 backbone-dest-mac "PE-4" isid 12
            exit
            sap 1/1/c3/1:12 create
                no shutdown
            exit
            no shutdown
        exit
```

As discussed, the **bridge-priority** influences the election of the multicast designated bridge. By making PE-2’s bridge-priority zero, it ensures that PE-2 becomes the root of all the STs for B-VPLS-10 as long as the priority for the rest of the PEs is larger than zero. In case of a tie, the PE owning the lowest system BMAC will be elected as multicast designated bridge. Figure 286: Basic SPBM topology shows the ST for I-VPLS-11 (see a thicker continuous line representing the ST). PE-2 is the root of the ST tree.

```
# on PE-2:
configure
    service
        pbb
            source-bmac 00-00-5e-00-53-02
            mac-name "PE-1" 00-00-5e-00-53-01
            mac-name "PE-2" 00-00-5e-00-53-02
            mac-name "PE-3" 00-00-5e-00-53-03
            mac-name "PE-4" 00-00-5e-00-53-04
            mac-name "PE-5" 00-00-5e-00-53-05
            mac-name "PE-6" 00-00-5e-00-53-06
        exit
        vpls 10 name "B-VPLS-10" customer 1 b-vpls create
            service-mtu 2000
            stp
                shutdown
            exit
            spb 1024 fid 10 create
                level 1
                    bridge-priority 0
                exit
                overload-on-boot timeout 60
                timers
                    spf-wait 2000 spf-initial-wait 50000 spf-second-wait 100000
                    lsp-wait 8000 lsp-initial-wait 10 lsp-second-wait 1000
                exit
                no shutdown
            exit
            spoke-sdp 21:10 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 23:10 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 25:10 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 26:10 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            no shutdown
        exit
```

The rest of the nodes is configured accordingly. SPB instance 1024 will set up shortest path first (SPF) trees for unicast traffic and a single tree (ST) per ISID with PE-2 as the root bridge (because it has the lowest bridge priority 0 configured) for BUM traffic. The ECT algorithm chosen for the B-VPLS FID (10) is the low-path-id (default).

Once SPBM is configured on all the six nodes, the six system BMAC addresses and the ISID 11 will be advertised by SPB IS-IS.

The following show commands can help understand the IS-IS configuration for SPB 1024 and the BMAC addresses populated by IS-IS:

- **show service id "B-VPLS-10" spb base** provides the SPB configuration and parameters for a particular SPB B-VPLS.

```

*A:PE-1# show service id "B-VPLS-10" spb base

===============================================================================
Service SPB Information
===============================================================================
Admin State        : Up                 Oper State         : Up
ISIS Instance      : 1024               FID                : 10
Bridge Priority    : 8                  Fwd Tree Top Ucast : spf
Fwd Tree Top Mcast : st
Bridge Id          : 80:00.00:00:5e:00:53:01
Mcast Desig Bridge : 00:00.00:00:5e:00:53:02

===============================================================================
Rtr Base ISIS Instance 1024 Interfaces
===============================================================================
Interface                        Level CircID  Oper      L1/L2 Metric     Type
                                               State
-------------------------------------------------------------------------------
sdp:12:10                        L1    65537   Up        10/-             p2p
sdp:16:10                        L1    65538   Up        10/-             p2p
-------------------------------------------------------------------------------
Interfaces : 2
===============================================================================
FID ranges using ECT Algorithm
-------------------------------------------------------------------------------
```

```
1-4095    low-path-id
===============================================================================
```

- **show service id "B-VPLS-10" spb fdb** provides the B-VPLS FDB that has been populated by IS-IS, for the unicast and multicast entries.

```
*A:PE-1# show service id "B-VPLS-10" spb fdb

==============================================================================
User service FDB information
==============================================================================
MAC Addr          UCast Source          State   MCast Source          State
------------------------------------------------------------------------------
00:00:5e:00:53:02 12:10                 ok      12:10                 ok
00:00:5e:00:53:03 12:10                 ok      12:10                 ok
00:00:5e:00:53:04 12:10                 ok      12:10                 ok
00:00:5e:00:53:05 12:10                 ok      12:10                 ok
00:00:5e:00:53:06 16:10                 ok      12:10                 ok
------------------------------------------------------------------------------
Entries found: 5
==============================================================================
```

The preceding output shows that the unicast (SPF) tree and the multicast (ST) tree differ with respect to PE-6.

The following commands help check the unicast and multicast topology for B-VPLS-10:

- **show service id "B-VPLS-10" spb routes** provides a detailed view of the unicast and multicast routes computed by SPF. As shown in the following command, the SPB unicast and multicast routes match on PE-2 because this node is the multicast designated bridge. Unicast and multicast routes will differ on most other nodes.

```

*A:PE-2# show service id "B-VPLS-10" spb routes

================================================================
MAC Route Table
================================================================
FID  MAC Addr                                     Ver.   Metric
      NextHop If                    SysID
----------------------------------------------------------------

Fwd Tree: unicast
----------------------------------------------------------------
10   00:00:5e:00:53:01                            1      10
      sdp:21:10                     PE-1
10   00:00:5e:00:53:03                            1      10
      sdp:23:10                     PE-3
10   00:00:5e:00:53:04                            1      20
      sdp:23:10                     PE-3
10   00:00:5e:00:53:05                            2      10
      sdp:25:10                     PE-5
10   00:00:5e:00:53:06                            2      10
      sdp:26:10                     PE-6

Fwd Tree: multicast
----------------------------------------------------------------
10   00:00:5e:00:53:01                            1      10
      sdp:21:10                     PE-1
10   00:00:5e:00:53:03                            1      10
      sdp:23:10                     PE-3
10   00:00:5e:00:53:04                            1      20
```

```
      sdp:23:10                     PE-3
10   00:00:5e:00:53:05                            2      10
      sdp:25:10                     PE-5
10   00:00:5e:00:53:06                            2      10
      sdp:26:10                     PE-6
----------------------------------------------------------------
No. of MAC Routes: 10
================================================================

================================================================
ISID Route Table
================================================================
FID  ISID                                         Ver.
      NextHop If                    SysID
----------------------------------------------------------------
10   11                                           1
      sdp:21:10                     PE-1
      sdp:23:10                     PE-3
      sdp:26:10                     PE-6
----------------------------------------------------------------
No. of ISID Routes: 1
================================================================
```

- **show service id "B-VPLS-10" spb mfib** and **show service id "B-VPLS-10" mfib** show information of the MFIB entries generated in the B-VPLS as well as the outgoing interface (OIF) associated with those MFIB entries.

```
*A:PE-2# show service id "B-VPLS-10" spb mfib

===============================================================================
User service MFIB information
===============================================================================
MAC Addr          ISID     Status
-------------------------------------------------------------------------------
01:1E:83:00:00:0B 11       Ok
-------------------------------------------------------------------------------
Entries found: 1
===============================================================================
```

```
*A:PE-2# show service id "B-VPLS-10" mfib

===============================================================================
Multicast FIB, Service 10
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
*               01:1e:83:00:00:0b     b-sdp:21:10                  Local    Fwd
                                      b-sdp:23:10                  Local    Fwd
                                      b-sdp:26:10                  Local    Fwd
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

SPB multicast trees (STs) are pruned for each particular I-VPLS ISID, based on the advertisement of I-VPLS ISIDs in SPB IS-IS by each individual PE. Multicast B-VPLS traffic not belonging to any particular I-VPLS follows the default tree. The default tree is an ST for the B-VPLS which is not pruned and therefore reaches all the PE nodes in the B-VPLS. For instance, Ethernet-CFM CCM messages sent from vMEPs configured on the SPB B-VPLS will use the default tree. The default tree does not consume MFIB entries and can be checked in each node through the use of the following command:

```
*A:PE-5# tools dump service id "B-VPLS-10" spb default-multicast-list
saps : { }
spoke-sdps : { 52:10 }
```

PE-5 is not part of the tree for I-VPLS-11. However, as with any SPB node part of B-VPLS-10, PE-5 is part of the default tree. Refer to Configuration of ISID policies in SPB B-VPLS to see more use cases for the default tree.

The following tools commands allow the operator to easily see the forwarding path (unicast and multicast) followed by the traffic to a remote node, with the aggregate metric from the source.

```
*A:PE-1# tools dump service id 10 spb fid 10 forwarding-path destination PE-4 forwarding-tree
 unicast

Hop  BridgeId                 Metric From Src
0    PE-1                     0
1    PE-2                     10
2    PE-3                     20
3    PE-4                     30
```

```
*A:PE-1# tools dump service id 10 spb fid 10 forwarding-path destination PE-4 forwarding-tree
 multicast

Hop  BridgeId                 Metric From Src
0    PE-1                     0
1    PE-2                     10
2    PE-3                     20
3    PE-4                     30
```

In large networks or networks where IP multicast, PBB, and PBB-SPB services coexist, the data plane MFIB entries is a hardware resource that should be periodically checked. The **tools dump service vpls-mfib-stats** command shows the total number of hardware MFIB entries (in this case, 40959 entries) and the entries being used by IP multicast or PBB (MMRP or SPB) (in this case, 16383 entries). The **tools dump service vpls-pbb-mfib-stats** shows the breakdown between MFIB entries populated by MMRP, SPB, or by EVPN, and the individual limits, system-wide, and per service:

```
*A:PE-2# tools dump service vpls-mfib-stats
Service Manager VPLS MFIB info at 10/16/2023 09:42:57:

Statistics last cleared at 10/16/2023 09:35:40

            Statistic                        |    Count
---------------------------------------------+-------------
                         HW limit SG entries |      40959
                          Current SG entries |          1
                    Limit Non PBB SG entries |      16383
                  Current Non PBB SG entries |          0
                                SG limit hit |          0
---snip---
```

```
*A:PE-2# tools dump service vpls-pbb-mfib-stats detail

Service Manager VPLS PBB MFIB statistics at 10/16/2023 09:42:57:

Usage per Service
```

```
   ServiceId    MFIB User      Count
  ------------+--------------+-------
   10           spb               1
  ------------+--------------+-------
                       Total       1

MMRP
  Current Usage     :     0
  System Limit      :  8191 Full, 40959 ESOnly
  Per Service Limit :  2048 Full,  8192 ESOnly

SPB
  Current Usage     :     1
  System Limit      :  8191
  Per Service Limit :  8191

Evpn
  Current Usage     :     0
  System Limit      : 40959
  Per Service Limit :  8191
```

Finally, the following debug commands can help monitor the SPB IS-IS process and the protocol PDU exchanges:

- debug service id <svcId> spb
- debug service id <svcId> spb adjacency
- debug service id <svcId> spb interface
- debug service id <svcId> spb l2db
- debug service id <svcId> spb lsdb
- debug service id <svcId> spb packet <detail>
- debug service id <svcId> spb spf

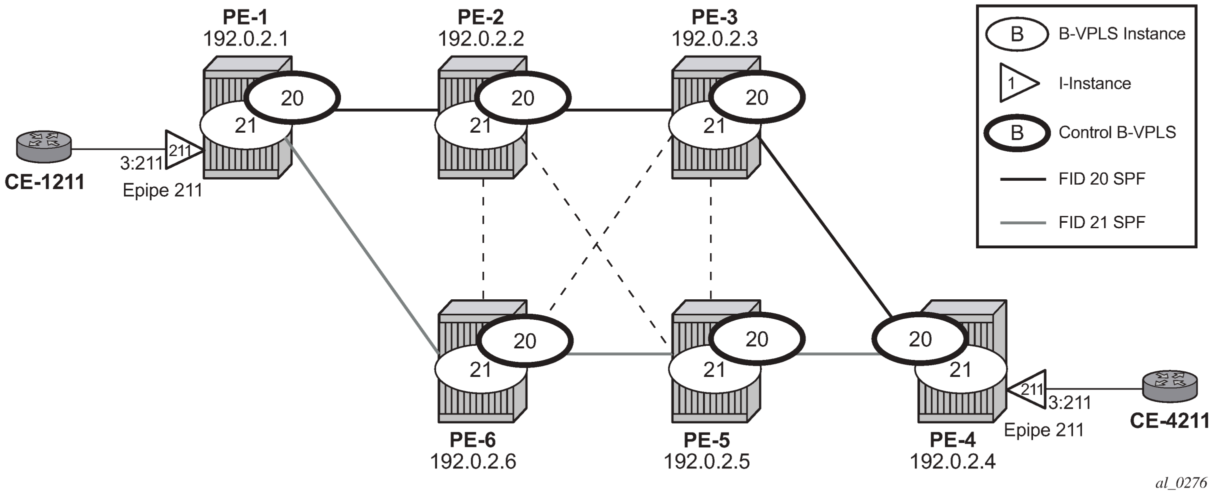
Control and user B-VPLS configuration

The SR OS implementation of SPB allows a single SPB IS-IS instance to control the paths and FDBs of many B-VPLS instances. This is done by using the control B-VPLS, user B-VPLS, and fate-sharing concepts.

The control B-VPLS will be SPB-enabled and configured with all the related SPB IS-IS parameters. Although the control B-VPLS might or might not have I-VPLSs or Epipes directly attached, it must be configured on all the nodes where SPB forwarding is expected to be active. SPB uses the logical instance and a forwarding ID (FID) to identify SPB locally on the node. That FID must be consistently configured on all the nodes where the B-VPLS exists. User B-VPLS are other instances of B-VPLS that are usually configured to separate the traffic for manageability reasons, QoS, or ECT different treatment.

Figure 287: Control and user B-VPLS example topology illustrates the control B-VPLS "control B-VPLS-20" and user B-VPLS "user B-VPLS-21" concept. In this example, there is only one user B-VPLS, but there can be several user B-VPLSs sharing fate with the same control B-VPLS. The control B-VPLS and the user B-VPLS must share the same topology and both B-VPLSs must share exactly the same interfaces. The user B-VPLS, which is linked to the control B-VPLS by its FID, follows—that is, inherits the state of—the control B-VPLS, but may use a different ECT path in case of equal metric paths, like in this example: FID 20, that is, the control B-VPLS, follows the low-path-id ECT, whereas FID 21, for example, the user B-VPLS, follows the high-path-id ECT.

Figure 287: Control and user B-VPLS example topology



The configurations of B-VPLSs 20 and 21, on PE-1 and PE-2, are as follows. The **spbm-control-vpls 20 fid 21** command in user B-VPLS-21 associates FID 21 to the user B-VPLS and links the B-VPLS to its control B-VPLS-20.

```
# on PE-1:
configure
    service
        vpls 20 name "control B-VPLS-20" customer 1 b-vpls create
            service-mtu 2000
            stp
                shutdown
            exit
            spb 1025 fid 20 create
                level 1
                    ect-algorithm fid-range 21-4095 high-path-id
                exit
                no shutdown
            exit
            spoke-sdp 12:20 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 16:20 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            no shutdown
        exit
        vpls 21 name "user B-VPLS-21" customer 1 b-vpls create
            service-mtu 2000
            stp
                shutdown
            exit
            spbm-control-vpls 20 fid 21
            spoke-sdp 12:21 create
                no shutdown
            exit
            spoke-sdp 16:21 create
                no shutdown
            exit
            no shutdown
        exit
        epipe 211 name "Epipe-211" customer 1 create
            pbb
                tunnel 21 backbone-dest-mac "PE-4" isid 211
            exit
            sap 1/1/c3/1:211 create
                no shutdown
            exit
            no shutdown
        exit
```

```
# on PE-2:
configure
    service
        vpls 20 name "control B-VPLS-20" customer 1 b-vpls create
            service-mtu 2000
            stp
                shutdown
            exit
            spb 1025 fid 20 create
                level 1
                    ect-algorithm fid-range 21-4095 high-path-id
                exit
                no shutdown
            exit
            spoke-sdp 21:20 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 23:20 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 25:20 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 26:20 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            no shutdown
        exit
        vpls 21 name "user B-VPLS-21" customer 1 b-vpls create
            service-mtu 2000
            stp
                shutdown
            exit
            spbm-control-vpls 20 fid 21
            spoke-sdp 21:21 create
```

```
                no shutdown
            exit
            spoke-sdp 23:21 create
                no shutdown
            exit
            spoke-sdp 25:21 create
                no shutdown
            exit
            spoke-sdp 26:21 create
                no shutdown
            exit
            no shutdown
        exit
```

If there is a mismatch between the topology of a user B-VPLS and its control B-VPLS, only the user B-VPLS links and nodes that are in common with the control B-VPLS will function.

User B-VPLS instances supporting only unicast services (PBB-Epipes) may share the FID with the other B-VPLS (control or user). This is a configuration shortcut that reduces the LSP advertisement size for B-VPLS services but results in the same separation for forwarding between the B-VPLS services. In the case of PBB-Epipes, only BMACs are advertised per FID, but BMACs are populated per B-VPLS in the FIB. If I-VPLS services are to be supported on a B-VPLS, that B-VPLS must have an independent FID.

Although user B-VPLS-21 does not have any SPB setting (other than the **spbm-control-vpls**), the spoke-SDPs use the same SDPs as the parent "control B-VPLS-20". The **show service id <user b-vpls> spb fate-sharing** command shows the control spoke-SDP/SAPs that control the user spoke-SDP/SAPs.

```
*A:PE-1# show service id "user B-VPLS-21" spb fate-sharing

===============================================================================
User service fate-shared sap/sdp-bind information
===============================================================================
Control   Control Sap/             FID       User      User Sap/
SvcId     SdpBind                            SvcId     SdpBind
-------------------------------------------------------------------------------
20        12:20                    21        21        12:21
20        16:20                    21        21        16:21
===============================================================================
```

SPBM access resiliency configuration

The following example shows how to configure an I-VPLS or Epipe attached to an SPB-enabled B-VPLS when access resiliency is used.

Multi-Chassis LAG (MC-LAG) is the only resiliency mechanism supported for PBB-Epipes. The MC-LAG active node will advertise the MC-LAG BMAC (or SAP BMAC) in SPB IS-IS. In case of failure, when the standby node takes over, it will advertise the MC-LAG SAP BMAC. Without SPB, the MC-LAG solution for PBB-Epipe required the use of MAC notification and periodic MAC notification. SPB provides a faster and more efficient solution without the need for any extra MAC notification mechanism. In the example described in this section, Epipe 31 uses MC-LAG access resiliency to get connected to the B-VPLS-30 on nodes PE-2 and PE-6.

As far as I-VPLS access resiliency is concerned, the same mechanisms supported for regular B-VPLS are supported for SPB-enabled B-VPLS, except for G.8032. A very important aspect of the I-VPLS resiliency is a proper MAC flush propagation when there is a failure at the I-VPLS access links.

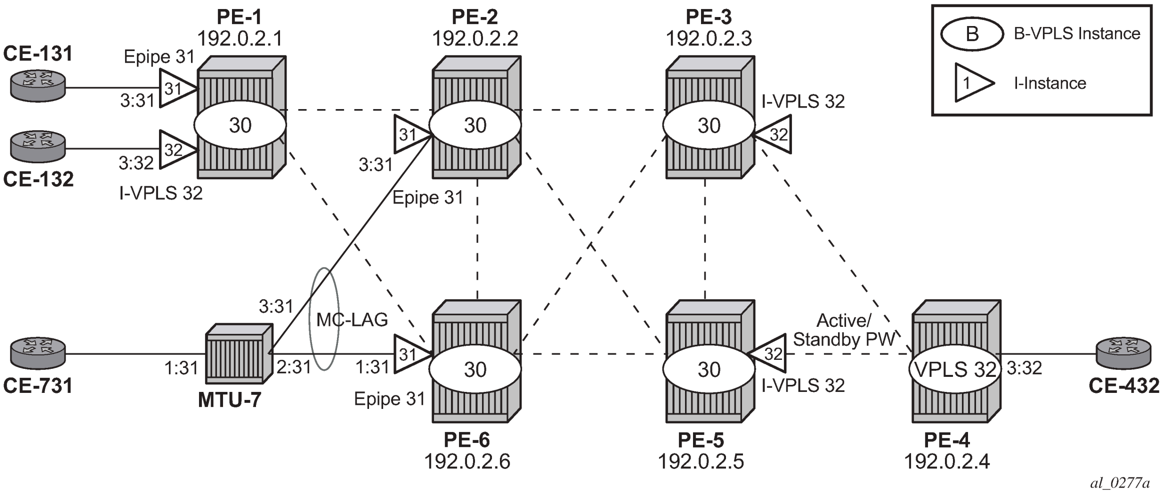
If the SPB-enabled B-VPLS uses B-SAPs for its connectivity to the backbone, there is no MAC flush propagation (because there is no TLDP). In this case, if MC-LAG is used and there is an MC-LAG switchover, the new active chassis will keep using the same source BMAC, such as the SAP BMAC, and it will advertise it in the B-VPLS domain so that the remote FDBs can be properly updated. No MAC flush is required in this case.

When the B-VPLS uses spoke-SDPs for its backbone connectivity, the traditional LDP MAC flush propagation mechanisms and commands can be used as follows:

- **send-flush-on-failure** works as expected when SPB is used at the B-VPLS. When configured, a flush-all-from-me event is triggered upon a SAP or spoke-SDP failure in the I-VPLS.
- **send-bvpls-flush** works as expected when SPB is used at the B-VPLS. Two variants are configurable: all-from-me/all-but-mine. Any I-VPLS SAP or spoke-SDP failure is propagated to the I-VPLS on the peers to flush their respective customer MAC addresses (CMACs). It works only in conjunction with send-flush-on-failure configuration on I-VPLS. The associated ISID list is passed along with the LDP MAC flush message, which is flushed or retained according to the **all-from-me**/**all-but-me** flag.
- **send-flush-on-bvpls-failure** works as expected when SPB is used at the B-VPLS. A local B-VPLS failure is propagated to the I-VPLS, which then triggers a LDP MAC flush if it has any spoke SDP on it.
- **propagate-mac-flush-from-bvpls** does not work when SPB is used at the B-VPLS (because failures within the B-VPLS are handled by SPB) and its configuration is blocked.

In the example described later in this section, I-VPLS-32 uses active/standby spoke-SDP resiliency to get connected to the B-VPLS-30 on nodes PE-3 and PE-5.

Figure 288: Access resiliency example topology



As an example of MC-LAG connectivity, the Epipe-31 configuration is shown. Just like for regular PBB-VPLS, a SAP BMAC is used as source BMAC for the Epipe traffic from PE-2 or PE-6 to PE-1. A SAP BMAC is a virtual BMAC formed from the configured source BMAC plus the MC-LAG LACP-key (if configured this way) and owned by the MC-LAG active chassis.

The following CLI output shows the configuration of MC-LAG as well as the generation of the SAP BMAC. Once it is properly configured and the MC-LAG and Epipe are up and running, SPB IS-IS will distribute the SAP BMAC throughout the B-VPLS, as it does for the system BMACs and OAM vMEP MACs. In this example, PE-2 is the MC-LAG active node, therefore the SAP BMAC for Epipe 31 is generated from PE-2.

```
# on PE-2:
configure
    lag 1 name "lag-1"
        mode access
        encap-type dot1q
        port 1/1/c3/1
        lacp active administrative-key 32768
        no shutdown
    exit
    redundancy
        multi-chassis
            peer 192.0.2.6 create
                mc-lag
                    lag 1 lacp-key 1 system-id 00:00:00:00:02:06 system-priority 65535
                                                          source-bmac-lsb use-lacp-key
                    no shutdown
                exit
                no shutdown
            exit
        exit
```

```
# on PE-2:
configure
    service
        vpls 30 name "B-VPLS-30" customer 1 b-vpls create
            service-mtu 2000
            pbb
                use-sap-bmac
            exit
            stp
                shutdown
            exit
            spb 1026 fid 30 create
                level 1
                    bridge-priority 0
                exit
                no shutdown
            exit
            spoke-sdp 21:30 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 23:30 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 25:30 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 26:30 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            no shutdown
        exit
        epipe 31 name "Epipe-31" customer 1 create
            pbb
                tunnel 30 backbone-dest-mac "PE-1" isid 31
            exit
            sap lag-1:31 create
                no shutdown
            exit
            no shutdown
        exit
```

```
*A:PE-6# show service id "B-VPLS-30" spb fdb

==============================================================================
User service FDB information
==============================================================================
MAC Addr          UCast Source          State   MCast Source          State
------------------------------------------------------------------------------
00:00:5e:00:00:01 62:30                 ok      62:30                 ok
00:00:5e:00:53:01 61:30                 ok      62:30                 ok
00:00:5e:00:53:02 62:30                 ok      62:30                 ok
00:00:5e:00:53:03 63:30                 ok      62:30                 ok
00:00:5e:00:53:05 65:30                 ok      62:30                 ok
------------------------------------------------------------------------------
Entries found: 5
==============================================================================
```

The VPLS configuration on PE-4 and PE-3 is as follows.

```
# on PE-4:
configure
    service
        vpls 32 name "VPLS-32" customer 1 create     # Ordinary VPLS, no I-VPLS
            endpoint "CORE" create
                no suppress-standby-signaling
            exit
            stp
                shutdown
            exit
            sap 1/1/c3/1:32 create
                no shutdown
            exit
            spoke-sdp 43:32 endpoint "CORE" create
                stp
                    shutdown
                exit
                precedence primary
                no shutdown
            exit
            spoke-sdp 45:32 endpoint "CORE" create
                stp
                    shutdown
                exit
                no shutdown
            exit
            no shutdown
        exit
```

```
# on PE-3:
```

```
configure
    service
        vpls 30 name "B-VPLS-30" customer 1 b-vpls create
            service-mtu 2000
            stp
                shutdown
            exit
            spb 1026 fid 30 create
                no shutdown
            exit
            spoke-sdp 32:30 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 35:30 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 36:30 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            no shutdown
        exit
        vpls 32 name "I-VPLS-32" customer 1 i-vpls create
            send-flush-on-failure
            pbb
                backbone-vpls 30
                exit
                send-bvpls-flush all-from-me
            exit
            spoke-sdp 34:32 create
                no shutdown
            exit
            no shutdown
        exit
```

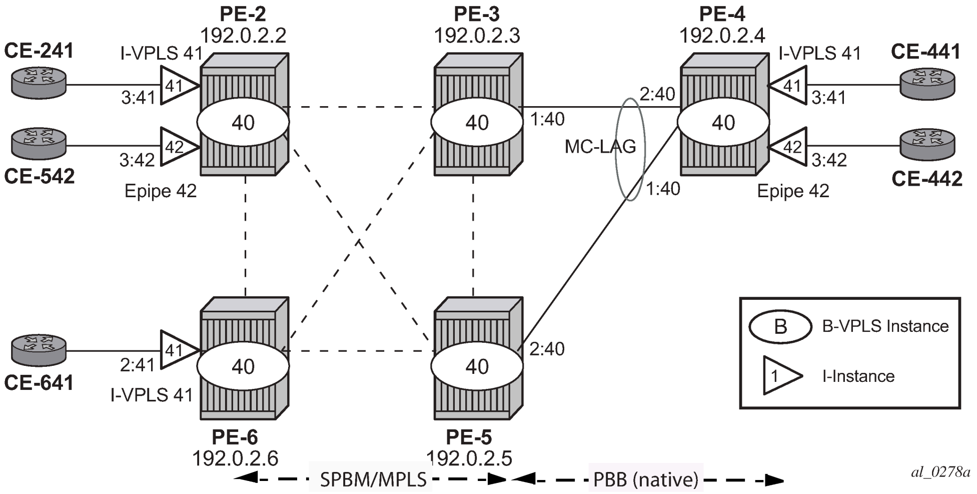
As discussed, **send-flush-on-failure** and **send-bvpls-flush all-from-me** are configured in the I-VPLS. When the active spoke-SDP goes down on PE-3, a flush-all-from-me message will be propagated through the backbone and will flush the corresponding CMACs associated to I-VPLS-32 in node PE-1. MAC flush-all-from-me messages are automatically propagated in the core up to the remote I-VPLS-32 on node PE-1 (there is no need for any propagate-mac-flush in the intermediate nodes). The **send-flush-on-bvpls-failure** command works as expected. The command **propagate-mac-flush-from-bvpls** is never used when the B-VPLS is SPB-enabled (the command is not allowed).

Static BMACs and static ISIDs configuration

SR OS supports the interworking between SPB-enabled B-VPLS and non-SPB B-VPLS instances. SPB networks can be connected to non-SPB capable nodes, for example third party vendor PBB switches or 7210 SAS nodes. This is possible through the use of conditional static BMACs and static ISIDs on the nodes doing the interworking function. Conditional static BMACs and static ISIDs can be associated to non-SPB B-VPLS SAPs or spoke-SDPs.

The following example shows an SPB-enabled B-VPLS (B-VPLS-40) on nodes PE-2, PE-6, PE-3, and PE-5. Node PE-4 supports PBB, but not SPB and it is connected by a MC-LAG to nodes PE-3 and PE-5. Services I-VPLS-41 and Epipe-42 have endpoints on node PE-4. In this example, nodes PE-3 and PE-5 are acting as interworking nodes. They will be configured with the BMAC of PE-4 so that the MC-LAG active node advertises the non-SPB capable node BMAC into SPB IS-IS. The BMAC will be configured as a conditional static BMAC so that an SPB node, such as PE-3 or PE-5, will only advertise PE-4’s BMAC if its connection to PE-4 is active. Besides the conditional static BMAC, nodes PE-3 and PE-5 should advertise the I-VPLS ISIDs defined in PE-4. Epipe ISIDs are not advertised in SPB IS-IS, therefore, it is not necessary to create a static ISID for Epipe-42.

Figure 289: Access resiliency example topology



The commands to configure conditional static BMACs and static ISIDs are as follows.

```
*A:PE-3>config>service>vpls$ static-mac mac ?
  - mac <ieee-address> [create] black-hole
  - mac <ieee-address> [create] sap <sap-id> monitor {fwd-status}
  - no mac <ieee-address>
  - mac <ieee-address> [create] spoke-sdp <sdp-id:vc-id> monitor {fwd-status}
---snip---
```

```
*A:PE-3>config>service>vpls>sap$ static-isid range ?
  - no range <range-id>
  - range <range-id> isid <isid-value> [to <isid-value>] [create]

 <range-id>           : [1..8191]
 <isid-value>         : [1..16777215]
 <create>             : keyword
```

The **monitor fwd-status** attribute identifies this to be a conditional MAC and is mandatory for static BMAC addresses. This parameter instructs SR OS to advertise the BMAC only if the corresponding SAP or spoke-SDP is in forwarding state.

The configuration of the conditional static BMAC and static ISID is as follows. The values for **spf-wait** are the default ones.

```
# on PE-3:
configure
```

```
    service
        vpls 40 name "B-VPLS-40" customer 1 b-vpls create
            service-mtu 2000
            stp
                shutdown
            exit
            spb 1027 fid 40 create
                timers
                    spf-wait 10000 spf-initial-wait 10 spf-second-wait 1000 # default
                exit
                no shutdown
            exit
            sap lag-1:40 create
                static-isid
                    range 1 create isid 41
                exit
            exit
            spoke-sdp 32:40 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 35:40 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 36:40 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            static-mac
                mac 00:00:5e:00:53:04 create sap lag-1:40 monitor fwd-status
            exit
            no shutdown
        exit
```

```
# on PE-5:
configure
    service
        vpls 40 name "B-VPLS-40" customer 1 b-vpls create
            service-mtu 2000
            stp
                shutdown
            exit
            spb 1027 fid 40 create
                timers
                    spf-wait 10000 spf-initial-wait 10 spf-second-wait 1000 # default
                exit
                no shutdown
            exit
            sap lag-1:40 create
                static-isid
                    range 1 create isid 41
                exit
            exit
            spoke-sdp 52:40 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 53:40 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            spoke-sdp 56:40 create
                spb create
                    no shutdown
                exit
                no shutdown
            exit
            static-mac
                mac 00:00:5e:00:53:04 create sap lag-1:40 monitor fwd-status
            exit
            no shutdown
        exit
```

The configuration of the conditional static BMAC is different from the legacy **static-mac** command, configured within the SAP/SDP-binding context. The latter static-MAC is not conditional and it is always added to the FDB. The conditional static BMAC is added to the FDB based on the SAP/SDP-binding state (the conditional static BMAC is tagged in the FDB as *CStatic*, for Conditional Static).

```
*A:PE-3# show lag 1

===============================================================================
Lag Data
===============================================================================
Lag-id         Adm     Opr     Weighted Threshold Up-Count MC Act/Stdby
    name
-------------------------------------------------------------------------------
1              up      up      No       0         1        active
    lag-1
===============================================================================
```

```
*A:PE-3# show service id "B-VPLS-40" fdb pbb

=======================================================================
Forwarding Database, b-Vpls Service 40
=======================================================================
MAC               Source-Identifier     iVplsMACs  Epipes     Type/Age
 Transport:Tnl-Id
-----------------------------------------------------------------------
00:00:5e:00:53:02 sdp:32:40             0          0          Spb
00:00:5e:00:53:04 sap:lag-1:40          0          0          CStatic
00:00:5e:00:53:05 sdp:35:40             0          0          Spb
00:00:5e:00:53:06 sdp:36:40             0          0          Spb
=======================================================================
```

On PE-5, LAG 1 is in standby, as follows:

```
*A:PE-5# show lag 1

===============================================================================
Lag Data
===============================================================================
Lag-id         Adm     Opr     Weighted Threshold Up-Count MC Act/Stdby
    name
```

```
-------------------------------------------------------------------------------
1              up      down    No       0         0        standby
    lag-1
===============================================================================
```

SAP LAG 1 in B-VPLS-40 is not forwarding any traffic. The FDB for B-VPLS-40 on PE-5 does not contain any conditional static MAC addresses, even though MAC 00:00:5e:00:53:04 is configured on SAP LAG 1. In the FDB for B-VPLS-40 on PE-5, this MAC address is assigned to SDP 53:40 (type SPB), as follows:

```
*A:PE-5# show service id "B-VPLS-40" fdb pbb

=======================================================================
Forwarding Database, b-Vpls Service 40
=======================================================================
MAC               Source-Identifier     iVplsMACs  Epipes     Type/Age
 Transport:Tnl-Id
-----------------------------------------------------------------------
00:00:5e:00:53:02 sdp:52:40             0          0          Spb
00:00:5e:00:53:03 sdp:53:40             0          0          Spb
00:00:5e:00:53:04 sdp:53:40             0          0          Spb
00:00:5e:00:53:06 sdp:56:40             0          0          Spb
=======================================================================
```

The **static-isid** command identifies a set of ISIDs for I-VPLS services that are external to SPBM. These ISIDs are advertised as supported locally on this node unless altered by an ISID policy. Although the preceding example shows the use of the static ISID associated to a MC-LAG SAP, regular SAPs or spoke-SDPs are also supported. ISIDs declared in this way become part of the ISID multicast and consume MFIBs. Multiple SPBM static-ISID ranges are allowed under a SAP or spoke-SDP. ISIDs are advertised as if they were attached to the local BMAC. Only remote I-VPLS ISIDs need to be defined. In the MFIB, the backbone group MAC addresses are then associated with the active SAP or spoke-SDP.

Once the conditional static BMAC for PE-4 and the static ISID 41 (for I-VPLS-41) are configured as described, the advertised BMAC and ISID can be checked in the remote SPB nodes:

```
*A:PE-6# show service id "B-VPLS-40" spb fdb

==============================================================================
User service FDB information
==============================================================================
MAC Addr          UCast Source          State   MCast Source          State
------------------------------------------------------------------------------
00:00:5e:00:53:02 62:40                 ok      62:40                 ok
00:00:5e:00:53:03 63:40                 ok      62:40                 ok
00:00:5e:00:53:04 63:40                 ok      62:40                 ok
00:00:5e:00:53:05 65:40                 ok      62:40                 ok
------------------------------------------------------------------------------
Entries found: 4
==============================================================================
```

```
*A:PE-6# show service id "B-VPLS-40" spb mfib

===============================================================================
User service MFIB information
===============================================================================
MAC Addr          ISID     Status
-------------------------------------------------------------------------------
01:1E:83:00:00:29 41       Ok
-------------------------------------------------------------------------------
Entries found: 1
```

```
===============================================================================
```

```
*A:PE-6# show service id "B-VPLS-40" mfib

===============================================================================
Multicast FIB, Service 40
===============================================================================
Source Address  Group Address         Port Id                      Svc Id   Fwd
                                                                            Blk
-------------------------------------------------------------------------------
*               01:1e:83:00:00:29     b-sdp:62:40                  Local    Fwd
-------------------------------------------------------------------------------
Number of entries: 1
===============================================================================
```

The group address terminates in hex 29, which corresponds to ISID 41.

The configured static ISIDs can be displayed with the following command (a range 41-100 has been added to the SAP lag-1:40 to demonstrate this output):

```
# on PE-5:
configure
    service
        vpls "B-VPLS-40"
            sap lag-1:40 create
                static-isid
                    range 1 create isid 41 to 100
                exit
            exit
```

```
*A:PE-5# show service id "B-VPLS-40" sap lag-1:40 static-isids

===================================
Static Isid Entries
===================================
Entry          Range
-----------------------------------
1              41-100
===================================
```

Configuration of ISID policies in SPB B-VPLS

ISID policies are an optional aspect of SPBM which allow additional control of the advertisement of ISIDs and creation of MFIB entries for I-VPLS (Epipe services do not trigger ISID advertisements or the creation of MFIB entries). By default, if no ISID policies are used, SPBM automatically advertises and populates MFIB entries for I-VPLS and static ISIDs. ISID policies can be used on any SPB-enabled node with locally defined I-VPLS instances or static ISIDs. The ISID policy parameters are as follows:

```
*A:PE-3>config>service>vpls>isid-policy# entry ?
  - entry <range-entry-id> [create]
  - no entry <range-entry-id>

 <range-entry-id>     : [1..8191]
 <create>             : keyword


 [no] advertise-local - Configure local advertisement of the range
 [no] range           - Configure ISID range for the entry
 [no] use-def-mcast   - Use default multicast tree to propogate ISID range
```

Where:

- **advertise-local** defines whether the local ISIDs (I-VPLS ISIDs linked to the B-VPLS) or static ISIDs contained in the configured range are advertised in SPBM.
- **use-def-mcast** controls whether the ISIDs contained in the range use MFIB entries (if **no use-def-mcast** is used) or just the default tree which does not use any MFIB entry.

The ISID policy becomes active as soon as it is defined, as opposed to other policies in SR OS, which require the policy itself to be applied within the configuration.

The typical use of ISID policies is to reduce the number of ISIDs being advertised and to save MFIB space (in deployments where MFIB space is shared with MMRP and IP multicast). The use of ISID policies is recommended for I-VPLS where most of the traffic is unicast or for I-VPLS where the ISID endpoints are present in all the backbone edge bridges (BEBs) of the SPB network. In both cases, advertising ISIDs or consuming MFIB entries for those I-VPLSs has little value because no multicast (first case) or the default tree (second case) are as efficient as using MFIB entries.

The following configuration example will use the example topology in Figure 289: Access resiliency example topology. In this case, the objective of the ISID policy will be to use the default tree for all the I-VPLS services with ISIDs between 41 and 100, excluding the range 80-90. The following example shows the policy configuration in the SPB nodes PE-2, PE-3, PE-5, and PE-6:

```
# on PE-2, PE-3, PE-5, PE-6:
configure
    service
        vpls "B-VPLS-40"
            isid-policy
                entry 10 create
                    range 80 to 90
                exit
                entry 20 create
                    use-def-mcast
                    no advertise-local
                    range 41 to 79
                exit
                entry 30 create
                    use-def-mcast
                    no advertise-local
                    range 91 to 100
                exit
            exit
```

The **no advertise-local** option can only be configured if the **use-def-mcast** option is also configured.

```
*A:PE-3>config>service>vpls>isid-policy>entry$ no advertise-local
MINOR: SVCMGR #7855 Cannot set AdvLocal for entry - advertise-local or use-def-mcast option
 must be specified
```

Overlapping ISID values can be configured as long as the actions are consistent for the same ISID. Conflicting actions are shown in the CLI.

```
*A:PE-3>config>service>vpls>isid-policy>entry# use-def-mcast
MINOR: SVCMGR #7854 Cannot set UseDefMctree for entry - Conflicting Actions with Entry-10
```

The ISID policy configured for B-VPLS-40 in all the four nodes makes the SPB network to use the default tree for ISIDs 41-79 and 91-100 and not advertise those ISIDs in SPB ISIS even if the ISID is locally defined (as in the case for ISIDs 41-100 in PE-3). As discussed in Basic SPBM configuration , the default tree path can be checked from each node by using the **tools dump service id 40 spb default-multicast-list** command.

Due to entry 10 in the policy, ISIDs 80-90 will be advertised by PE-3 (active MC-LAG node). However, nodes PE-2 and PE-6 will not create any MFIB entry for those ISIDs until the corresponding I-VPLS ISIDs are locally created (or configured through static-ISIDs). The following command executed on PE-2 proves that ISIDs 80-90 are indeed being advertised by PE-3:

```
*A:PE-2# show service id "B-VPLS-40" spb database detail

===============================================================================
Rtr Base ISIS Instance 1027 Database (detail)
===============================================================================

Displaying Level 1 database
-------------------------------------------------------------------------------
---snip---

-------------------------------------------------------------------------------
LSP ID    : PE-3.00-00                                  Level     : L1
---snip---

TLVs :
---snip---
  MT Capability :
    TLV Len        : 56
    MT ID          : 0
    SPBM Service ID:
    Sub TLV Len    : 52
      BMac Addr              : 00:00:5e:00:53:03
      Base VID               : 40
      ISIDs                  :
        80       Flags:TR
        81       Flags:TR
        82       Flags:TR
        83       Flags:TR
        84       Flags:TR
        85       Flags:TR
        86       Flags:TR
        87       Flags:TR
        88       Flags:TR
        89       Flags:TR
        90       Flags:TR
  TE IS Nbrs   :
---snip---
```

The **mfib** parameter in the **show service id 40 sap static-isids mfib** command can help understand the state of the MFIB entries added (or not) by the configured static ISID. The following possible states can be shown:

- If the static ISID is configured and programmed in the MFIB, the status is shown as:
  - ok
- If the static ISID is not configured and not programmed in the MFIB, the reasons can be (order of priority):
  - useDefMCTree - ISID policy is applied on the service for the ISID.

  - sysMFibLimit - system MFIB limit has been exceeded
  - addPending - adding pending due to processing delays
- If the static ISID is not configured, but present in the MFIB:
  - delPending - cleanup pending due to processing delays.

The following output shows the status of the static ISIDs:

```
*A:PE-5# show service id "B-VPLS-40" sap lag-1:40 static-isids mfib

===================================
ISID Detail
===================================
ISID           Status
-----------------------------------
41             useDefMCTree
42             useDefMCTree
---snip---
79             useDefMCTree
80             ok
81             ok
82             ok
83             ok
84             ok
85             ok
86             ok
87             ok
88             ok
89             ok
90             ok
91             useDefMCTree
---snip---
100            useDefMCTree
===================================
```

Conclusion

SR OS supports an efficient SPBM implementation in the context of a B-VPLS, where system BMACs, vMEP OAM BMACs, and SAP BMACs are advertised in SPB IS-IS. SPBM provides a simple solution where no other control plane protocol is required in the B-VPLS to take care of the resiliency, load-balancing, and multicast optimization. The SPBM implementation in the SR OS provides scale optimization through the use of control and user B-VPLSs, allows the interworking between SPB networks and PBB networks, as well as the optimization of the MFIB resources and advertisement of ISIDs through the use of ISID policies.

Static VXLAN Termination in Epipe Services

This chapter provides information about Static VXLAN Termination in Epipe Services.

Topics in this chapter include:

Applicability

This chapter was initially written for SR OS Release 15.0.R6, but the CLI in the current edition is based on SR OS Release 21.5.R1. Static VXLAN termination for Epipe services is supported in SR OS Release 15.0.R1, and later.

Overview

Static Virtual eXtensible Local Area Network (VXLAN) termination on non-system IP addresses of the PEs is supported in VPLS services, as described in chapter VXLAN Forwarding Path Extension, and in Epipe services, as described in this chapter. Whereas VPLSs using VXLAN require BGP-EVPN control plane in the current release, Epipe services using VXLAN do not. This implies that only the configured values are used because no auto-discovery of the remote Termination Endpoints (TEPs) can be done without BGP-EVPN.

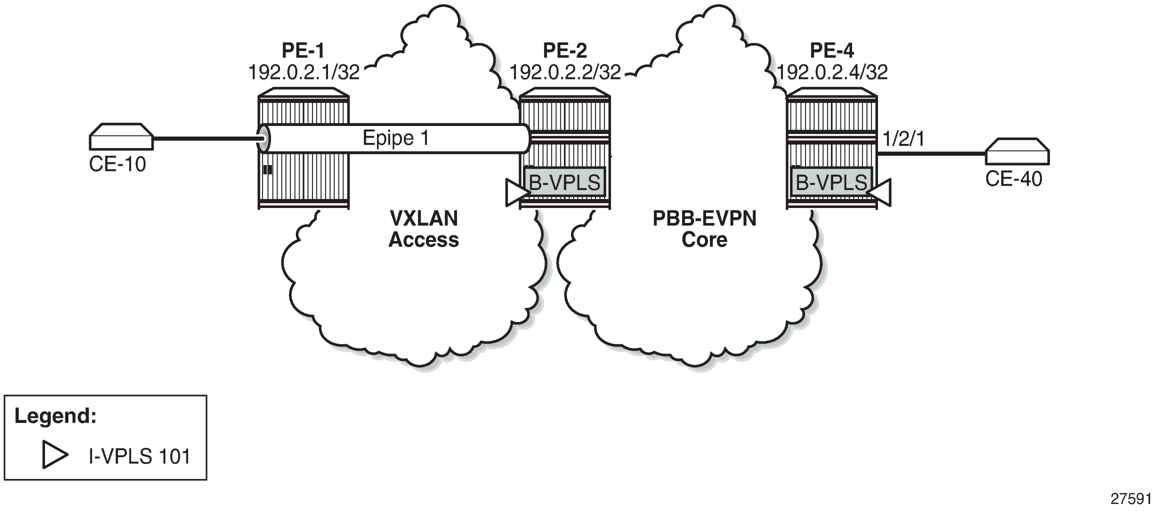
This chapter describes the configuration and use of static VXLAN as an access tunneling mechanism to a PBB-EVPN network. This is a design deployed in some service provider networks where the aggregation network is a non-MPLS IP network.

Static VXLAN termination for Epipe services can be applied on system IP addresses or non-system IP addresses.

Static VXLAN termination on system IP addresses

Figure 290: Static VXLAN termination on system IP addresses shows an example topology with three PEs and two CEs. Epipe 1 is configured on PE-1 and PE-2. PE-2 and PE-4 are part of a PBB-EVPN network. On PE-2, a port cross-connect (PXC) is configured to connect the SAP in Epipe 1 and the SAP in I-VPLS 101. CE-10 and CE-40 can send traffic to each other.

Figure 290: Static VXLAN termination on system IP addresses



On PE-1, Epipe 1 is configured with egress VXLAN VNI 1, egress VXLAN Termination Endpoint (VTEP) 192.0.2.2, oper-group op-grp-1, and a SAP toward CE-10, as follows:

```
# on PE-1:
configure
    service
        oper-group "op-grp-1" create
        exit
        epipe 1 name "Epipe 1" customer 1 create
            vxlan instance 1 vni 1 create
                egr-vtep 192.0.2.2
                    oper-group "op-grp-1"
                exit
            exit
            sap 1/2/1:1.* create
                no shutdown
            exit
            no shutdown
        exit
```

where:

- The configured VXLAN Virtual Network Identifier (VNI) is used by the system as follows:
  - As the egress VNI when sending VXLAN packets for the Epipe service
  - As the source VNI that identifies the VXLAN packet to be part of the Epipe
  - Unique in the system, so it can only be configured in one service, either VPLS or Epipe

  The configuration of the VXLAN VNI in an Epipe is similar to the configuration of the VXLAN VNI in a VPLS, except that in a VPLS, the VNI is only used as the source VNI, because the egress VNI is learned from BGP-EVPN. However, in Epipe services with static VXLAN, the egress VNI is also the configured VNI.
- The egress VTEP is the system IP address of the remote PE. The system will add the configured egress VTEP IP address as the remote VTEP when encapsulating the frames into VXLAN packets.

Only the egress VTEP is configured, not the source VTEP. The PE receiving VXLAN packets will not check the source VTEP.

- The egress VTEP IP address must be in the Routing Table Manager (RTM). An oper-group is associated with the egress VTEP IP address, so that when the egress VTEP disappears from the base route table, the oper-group is brought operationally down, which propagates the failure to other objects that have this oper-group associated. The status of the oper-group and the service will be as follows:
  - When the egress VTEP disappears from the RTM, the VXLAN binding goes operationally down and the oper-group associated with the egress VTEP goes operationally down.
  - When the Epipe SAP goes down, the service goes down too.
  - When the VXLAN binding goes down, the service remains up as long as the access SAP is up.
  - When the service is admin shutdown, the VXLAN binding and the oper-group associated with the egress VTEP are both brought operationally down.
- Only SAPs can be associated with the Epipe; no spoke-SDPs are supported in SR OS Release 21.5.R1, as follows. Regular SAPs and PXC SAPs are supported.

```
*A:PE-1>config>service>epipe# spoke-sdp 11:1 create
MINOR: SVCMGR #1957 SDP binding not supported - service has vxlan vtep configured
```

Frame encapsulation and forwarding

Incoming traffic in the PEs is treated as follows:

- For frames received from the SAPs, a SAP lookup identifies all frames matching the configured SAP (on PE-1, SAP 1/2/1:1.*). The matching frames will be encapsulated into VXLAN IPv4 packets with the following fields:
  - Source VTEP = system IP address
  - Destination VTEP = configured address in **egr-vtep**
  - VNI = configured VXLAN VNI
  - Source and destination UDP ports will be populated as per the existing VXLAN implementation VPLS services, with the source UDP port populated with the result of a hash on the ingress packets.
- For VXLAN frames received from the VXLAN network, a VNI lookup is done for packets with IP DA = system IP address. Frames with the configured VNI 1 are assigned to Epipe 1. The VXLAN encapsulation is removed and the frames are forwarded to the SAP.

Per-service hashing is not supported in Epipe-VXLAN services; only regular hashing and spraying in LAG/ECMP is supported as in any Epipe.

Static VXLAN termination on IPv6 or non-system IPv4 addresses

The non-system IPv4 or IPv6 VXLAN termination on Epipe services is configured in the same way as for VPLS services and described in the VXLAN Forwarding Path Extension chapter, using the FPE function for additional processing. The following steps are required for configuring the FPE for VXLAN termination:

1. Create FPE.
2. Associate FPE with VXLAN termination.

3. Configure the loopback router interface subnet for VXLAN termination and its advertisement into the routing protocol. The subnet can be IPv4 or IPv6.
4. Configure the loopback address for VXLAN termination.
5. Add the service configuration.

Configuration

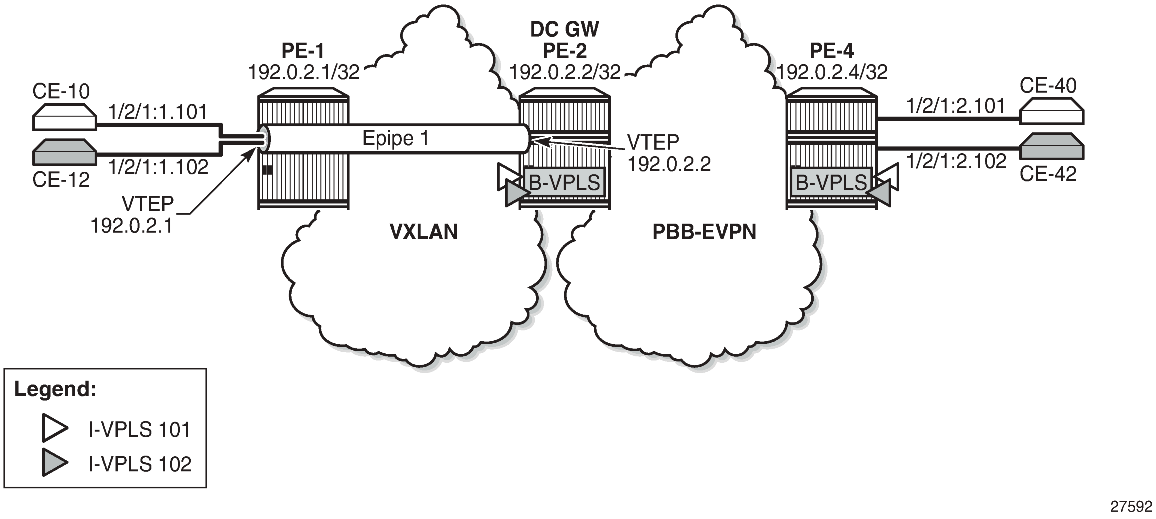
In this section, static VXLAN termination for Epipe services is configured for the following cases:

- VXLAN termination on system IP addresses
- VXLAN termination on non-system IPv4 addresses
- VXLAN termination on IPv6 addresses
- Static VXLAN used as access network for PBB-EVPN core: all-active multi-homing

Static VXLAN termination on system IP addresses

Figure 291: Example topology for static VXLAN termination on system IP addresses shows the example topology for static VXLAN termination on system IP addresses. The initial configuration of the PEs includes the cards, MDAs, ports, router interfaces, and IGP. BGP is not required on PE-1; on PE-2 and PE-4, BGP is configured for address family EVPN.

Figure 291: Example topology for static VXLAN termination on system IP addresses



On PE-1, Epipe 1 is configured with egress VXLAN VNI 1, egress VTEP 192.0.2.2, oper-group op-grp-1, and a SAP toward CE-10, as follows. This configuration was explained in the text under Figure 290: Static VXLAN termination on system IP addresses.

```
# on PE-1:
configure
    service
```

```
        oper-group "op-grp-1" create
        exit
        epipe 1 name "Epipe 1" customer 1 create
            vxlan instance 1 vni 1 create
                egr-vtep 192.0.2.2
                    oper-group "op-grp-1"
                exit
            exit
            sap 1/2/1:1.* create
                no shutdown
            exit
            no shutdown
        exit
```

On PE-2, BGP is configured for address family EVPN, as follows:

```
# on PE-2:
configure
    router
        autonomous-system 64500
        bgp
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family evpn
                peer-as 64500
                neighbor 192.0.2.4
                exit
            exit
```

There is a PXC configured on port 1/2/1 that will connect SAP pxc-21.a:1.* in Epipe 1, SAP pxc-21.b:1.101 in I-VPLS 101, and SAP pxc-21.b:1.102 in I-VPLS 102. The PXC is configured on PE-2 as follows. See chapter *Port Cross-Connect (PXC)* for more information.

```
# on PE-2:
configure
    port-xc
        pxc 21 create
            port 1/2/1
            no shutdown
        exit
    exit
    port pxc-21.a
        ethernet
            encap-type qinq
        exit
        no shutdown
    exit
    port pxc-21.b
        ethernet
            encap-type qinq
        exit
        no shutdown
    exit
    port 1/2/1
        no shutdown
    exit
```

The service configuration on PE-2 includes Epipe 1, B-VPLS 100, and I-VPLSs 101-102, as follows:

```
# on PE-2:
configure
    service
        oper-group "op-grp-1" create
        exit
        epipe 1 name "Epipe 1" customer 1 create
            vxlan instance 1 vni 1 create
                egr-vtep 192.0.2.1
                    oper-group "op-grp-1"
                exit
            exit
            sap pxc-21.a:1.* create
                no shutdown
            exit
            no shutdown
        exit
        vpls 100 name "B-VPLS 100" customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:00:00:00:00:02
            exit
            bgp
            exit
            bgp-evpn
                evi 100
                mpls bgp 1
                    ingress-replication-bum-label
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
        vpls 101 name "I-VPLS 101" customer 1 i-vpls create
            pbb
                backbone-vpls 100
                exit
            exit
            sap pxc-21.b:1.101 create
                no shutdown
            exit
            no shutdown
        exit
        vpls 102 name "I-VPLS 102" customer 1 i-vpls create
            pbb
                backbone-vpls 100
                exit
            exit
            sap pxc-21.b:1.102 create
                no shutdown
            exit
            no shutdown
        exit
```

The service configuration on PE-4 is similar for the B-VPLS and the I-VPLSs, but Epipe 1 is not configured on PE-4.

The following command shows the VXLAN information for Epipe 1 on PE-1. By default, the source VTEP is the system IP address 192.0.2.1.

```
*A:PE-1# show service id 1 vxlan
===============================================================================
Vxlan Src Vtep IP: N/A
===============================================================================

===============================================================================
Vxlan Instance
===============================================================================
VXLAN Instance                VNI                      Oper-flags
-------------------------------------------------------------------------------
1                             1                        none
-------------------------------------------------------------------------------
Number of Entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

```
*A:PE-1# show service id 1 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address                            Egress VNI          Oper      Vxlan
                                                            State     Type
-------------------------------------------------------------------------------
192.0.2.2                               1                   Up        static
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

The following command shows the oper-group information on PE-1 with the list of egress VTEP members.

```
*A:PE-1# show service oper-group "op-grp-1" detail

===============================================================================
Service Oper Group Information
===============================================================================
Oper Group       : op-grp-1
Creation Origin  : manual                          Oper Status: up
Hold DownTime    : 0 secs                          Hold UpTime: 4 secs
Members          : 1                               Monitoring : 0
===============================================================================

===============================================================================
Member Egr-Vtep for OperGroup: op-grp-1
===============================================================================
Svc Id              VNI                 VTEP Address
-------------------------------------------------------------------------------
1                   1                   192.0.2.2
-------------------------------------------------------------------------------
Egr-Vtep Entries found: 1
===============================================================================
```

The oper-group with member egress VTEP 192.0.2.2 cannot be monitored on a SAP in the same Epipe. The following error is raised when attempting to configure the same oper-group for the SAP in Epipe 1 on PE-1:

```
*A:PE-1>config>service>epipe>sap# oper-group "op-grp-1"
MINOR: SVCMGR #6221 Oper-group can not have monitor and member in the same service
```

The following ports on PE-2 are disabled to make the destination VTEP unreachable from PE-1:

```
# on PE-2:
configure
    port 1/1/1
        shutdown
    exit
    port 1/1/2
        shutdown
    exit
```

When the destination VTEP disappears from the RTM, the oper-group op-grp-1 goes down and the VXLAN binding in Epipe 1 goes down, while the Epipe service remains up, as follows:

```
*A:PE-1# show service oper-group "op-grp-1"

===============================================================================
Service Oper Group Information
===============================================================================
Oper Group       : op-grp-1
Creation Origin  : manual                          Oper Status: down
Hold DownTime    : 0 secs                          Hold UpTime: 4 secs
Members          : 1                               Monitoring : 0
===============================================================================
```

```
*A:PE-1# show service id 1 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address                            Egress VNI          Oper      Vxlan
                                                            State     Type
-------------------------------------------------------------------------------
192.0.2.2                               1                   Down      static
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

```
*A:PE-1#*A:PE-1# show service id 1 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : Epipe
---snip---

Admin State       : Up                  Oper State        : Up
---snip---
```

```
-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/2/1:1.*                            qinq         1578    1578    Up   Up
===============================================================================
```

The output is similar on PE-2. The ports are re-enabled on PE-2, which will cause the VXLAN binding and the oper-group to be operationally up again:

```
# on PE-2:
configure
    port 1/1/1
        no shutdown
    exit
    port 1/1/2
        no shutdown
    exit
```

The preceding example proved that the Epipe service remains up when the VXLAN binding goes down. The following example shows that the Epipe service goes down when the SAP goes down. On PE-1, port 1/2/1 is disabled, as follows:

```
# on PE-1:
configure
    port 1/2/1
        shutdown
```

The following command shows that SAP 1/2/1:1.* and Epipe 1 are down on PE-1:

```
*A:PE-1# show service id 1 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : Epipe
---snip---

Admin State       : Up                  Oper State        : Down
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/2/1:1.*                            qinq         1578    1578    Up   Down
===============================================================================
```

The port is re-enabled and SAP 1/2/1:1 and service Epipe 1 will be up again.

```
# on PE-1:
configure
    port 1/2/1
        no shutdown
```

When the service is disabled (admin shutdown), the SAP goes down, the VXLAN binding goes down, and the oper-group goes down, as follows:

```
# on PE-1:
configure
    service
        epipe "Epipe 1"
            shutdown
```

```
*A:PE-1# show service id 1 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 1                   Vpn Id            : 0
Service Type      : Epipe
---snip---

Admin State       : Down                Oper State        : Down
---snip---

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sap:1/2/1:1.*                            qinq         1578    1578    Up   Down
===============================================================================
```

```
*A:PE-1# show service id 1 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address                            Egress VNI          Oper      Vxlan
                                                            State     Type
-------------------------------------------------------------------------------
192.0.2.2                               1                   Down      static
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================
```

```
*A:PE-1# show service oper-group

===============================================================================
Service Oper Group Information
===============================================================================
Name                             Oper   Creation Hold   Hold   Members Monitor
                                 Status Origin   UpTime DnTime
                                                 (secs) (secs)
-------------------------------------------------------------------------------
op-grp-1                         down   manual   4      0      1       0
-------------------------------------------------------------------------------
Entries found: 1
===============================================================================
```

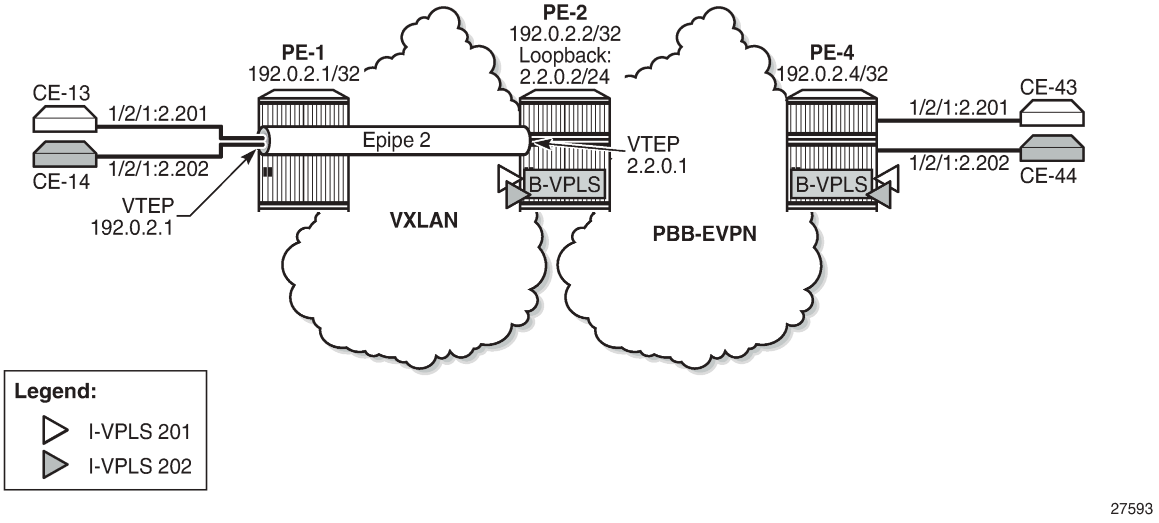
Static VXLAN termination on non-system IPv4 addresses

Non-system IP VXLAN termination is provisioned as follows:

1. Create FPE
2. Associate FPE with VXLAN termination
3. Configure router loopback interface
4. Configure non-system VXLAN termination VTEP addresses
5. Add the service configuration

Figure 292: Example topology for static VXLAN termination on non-system IPv4 addresses shows the example topology with PE-1 and PE-2 in a VXLAN network. The non-system loopback address on PE-2 will be used for VXLAN termination, whereas the system IP address will be used on PE-1.

Figure 292: Example topology for static VXLAN termination on non-system IPv4 addresses



Create FPE

FPE uses the back-to-back PXC, either a PXC port or a LAG-based PXC. The following PXC is created on PE-2:

```
# on PE-2:
configure
    port-xc
        pxc 1 create
            port 1/2/5
            no shutdown
        exit
```

The PXC sub-ports and ports are enabled as follows:

```
# on PE-2:
configure
```

```
    port pxc-1.a
        ethernet
            encap-type dot1q
        exit
        no shutdown
    exit
    port pxc-1.b
        ethernet
            encap-type dot1q
        exit
        no shutdown
    exit
    port 1/2/5
        no shutdown
    exit
```

```
*A:PE-2# show port pxc 1

===============================================================================
Ports on Port Cross Connect 1
===============================================================================
Port          Admin Link Port    Cfg  Oper LAG/ Port Port Port   C/QS/S/XFP/
Id            State      State   MTU  MTU  Bndl Mode Encp Type   MDIMDX
-------------------------------------------------------------------------------
pxc-1.a       Up    Yes  Up      1574 1574    - hybr dotq xgige
pxc-1.b       Up    Yes  Up      1574 1574    - hybr dotq xgige
===============================================================================
```

The following FPE uses the PXC:

```
# on PE-2:
configure
    fwd-path-ext
        fpe 1 create
            path pxc 1
        exit
```

The following shows that FPE 1 uses PXC 1 and has no VXLAN termination associated:

```
*A:PE-2# show fwd-path-ext fpe 1

===============================================================================
FPE Id: 1
===============================================================================
Description        : (Not Specified)
Path               : pxc 1
Pw Port            : Disabled                          Oper     : down
Sub Mgmt Extension : Disabled                          Oper     : N/A
Vxlan Termination  : Disabled                          Oper     : down
Segment-Routing V6 : Disabled
===============================================================================
```

Associate FPE with VXLAN termination

The following command associates FPE 1 with VXLAN termination:

```
# on PE-2:
configure
    fwd-path-ext
```

```
        sdp-id-range from 10000 to 10127
        fpe 1 create
            path pxc 1
            vxlan-termination
        exit
```

When attempting to associate the FPE with VXLAN termination without configuring a range of SDP IDs for FPE, the following error is raised:

```
*A:PE-2>config>fwd-path-ext>fpe# vxlan-termination
MINOR: FPE #1021 sdp-id-range is not configured
```

The following shows the range of SDP IDs for FPE and the list of configured FPEs; see the VXLAN Forwarding Path Extension chapter for more information about the use of SDP IDs. The application for FPE 1 is VXLAN termination.

```
*A:PE-2# show fwd-path-ext

===============================================================================
FPE Info
===============================================================================
FPE Id                   Path                          Application
                         pxc/xc-a, xc-b
-------------------------------------------------------------------------------
1                        pxc 1                         vxlan-term
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
SDP-Id Range: 10000 - 10127
===============================================================================
```

After the FPEs are associated with VXLAN termination, the system creates two internal router interfaces per FPE, one per PXC sub-port, as follows:

```
*A:PE-2# show router interface

===============================================================================
Interface Table (Router: Base)
===============================================================================
Interface-Name                   Adm       Opr(v4/v6)  Mode    Port/SapId
   IP-Address                                                  PfxState
-------------------------------------------------------------------------------
_tmnx_fpe_1.a                    Up        Up/Up       Network pxc-1.a:1
   fe80::100/64                                                PREFERRED
_tmnx_fpe_1.b                    Up        Up/Up       Network pxc-1.b:1
   fe80::101/64                                                PREFERRED
---snip---
```

The configuration of the internal interfaces can be verified as follows:

```
*A:PE-2>config>router# interface "_tmnx_fpe_1.a"
*A:PE-2>config>router>if# info
----------------------------------------------
            port pxc-1.a:1
            mac 00:00:00:00:00:01
            ipv6
                link-local-address fe80::100 dad-disable
                neighbor fe80::101 00:00:00:00:00:02
            exit
            no shutdown
```

```
----------------------------------------------
```

```
*A:PE-2>config>router# interface "_tmnx_fpe_1.b"
*A:PE-2>config>router>if# info
----------------------------------------------
            port pxc-1.b:1
            mac 00:00:00:00:00:02
            ipv6
                link-local-address fe80::101 dad-disable
                neighbor fe80::100 00:00:00:00:00:01
            exit
            no shutdown
----------------------------------------------
```

Configure router loopback interface

The following loopback interface is configured in PE-2 and added to the IS-IS context. The IPv6 address is not required yet.

```
# on PE-2:
configure
    router Base
        interface "loopback1"
            address 2.2.0.2/24
            loopback
            ipv6
                address 220::2/120
            exit
        exit
        isis 0
            interface "loopback1"
                no shutdown
            exit
        exit
```

A subnet must be assigned to the loopback interface, but not a /32 or /128 subnet mask, because the system cannot terminate VXLAN on a local interface address. In the preceding example, all addresses in the subnet 2.2.0.0/24 can be used for VXLAN tunnel termination, except for 2.2.0.2. The subnet will be advertised by the IGP. The subnet can be as small as /31 or /127.

Configure non-system VTEP addresses

On PE-2, non-system IP address 2.2.0.1 in the subnet of the loopback address 2.2.0.2/24 is configured as VTEP, as follows. Up to three non-system VTEP addresses can be configured to terminate VXLAN tunnels and their corresponding FPEs.

```
# on PE-2:
configure
    service
        system
            vxlan
                tunnel-termination 2.2.0.1 fpe 1 create
            exit
        exit
```

No non-system VTEP addresses need to be configured on PE-1.

When the non-system VTEP address is configured, an internal loopback interface _tmnx_vli_vxlan_1_131075 with VTEP address 2.2.0.1/32 is auto-created that can respond to ICMP requests.

```
*A:PE-2# show router interface

===============================================================================
Interface Table (Router: Base)
===============================================================================
Interface-Name                   Adm       Opr(v4/v6)  Mode    Port/SapId
   IP-Address                                                  PfxState
-------------------------------------------------------------------------------
_tmnx_fpe_1.a                    Up        Up/Up       Network pxc-1.a:1
   fe80::100/64                                                PREFERRED
_tmnx_fpe_1.b                    Up        Up/Up       Network pxc-1.b:1
   fe80::101/64                                                PREFERRED
_tmnx_vli_vxlan_1_131075         Up        Up/Up       Network loopback
   2.2.0.1/32                                                  n/a
   fe80::13:ffff:fe00:0/64                                     PREFERRED
---snip---
```

The system does not verify if there is a local base router loopback interface with a subnet corresponding to the VTEP address. If a tunnel termination address is configured and the FPE is up, the system will start terminating VXLAN traffic and responding ICMP for that address, regardless of the presence of a loopback in the base router. It is also possible that a non-loopback interface has an IP address in the configured subnet.

Configure the services

Epipe 2 is configured on PE-1 as follows. By default, the system IP address will be used as source VTEP of the VXLAN-encapsulated frames. The non-system IP address 2.2.0.1 is used as egress VTEP.

```
# on PE-1:
configure
    service
        epipe 2 name "Epipe 2" customer 1 create
            vxlan instance 1 vni 2 create
                egr-vtep 2.2.0.1
                exit
            exit
            sap 1/2/1:2.* create
                no shutdown
            exit
            no shutdown
        exit
```

The configuration of Epipe 2 on PE-2 defines the non-system IP address 2.2.0.1 as source VTEP, as follows. The egress VTEP is 192.0.2.1, the system IP address of PE-1. The configuration of the B-VPLS is the same as in the preceding example; the configuration of the I-VPLSs 201 and 202 is similar to the configuration of I-VPLS 101 in the preceding example.

```
# on PE-2:
configure
    service
        epipe 2 name "Epipe 2" customer 1 create
            vxlan-src-vtep 2.2.0.1
            vxlan instance 1 vni 2 create
                egr-vtep 192.0.2.1
                exit
            exit
            sap pxc-21.a:2.* create
                no shutdown
            exit
            no shutdown
        exit
```

The following **show** command on PE-1 shows that no VXLAN source VTEP IP address is configured:

```
*A:PE-1# show service id 2 vxlan
===============================================================================
Vxlan Src Vtep IP: N/A
===============================================================================

===============================================================================
Vxlan Instance
===============================================================================
VXLAN Instance                VNI                      Oper-flags
-------------------------------------------------------------------------------
1                             2                        none
-------------------------------------------------------------------------------
Number of Entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

The following shows that the egress VTEP is 2.2.0.1, which is a non-system VTEP on PE-2. The VXLAN tunnel is operationally up.

```
*A:PE-1# show service id 2 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address                            Egress VNI          Oper      Vxlan
                                                            State     Type
-------------------------------------------------------------------------------
2.2.0.1                                 2                   Up        static
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

The same commands on PE-2 show that source VTEP IP address 2.2.0.1 is configured and the egress VTEP is 192.0.2.1, which is the system IP address of PE-1, as follows:

```
*A:PE-2# show service id 2 vxlan
===============================================================================
Vxlan Src Vtep IP: 2.2.0.1
===============================================================================

===============================================================================
Vxlan Instance
===============================================================================
VXLAN Instance                VNI                      Oper-flags
-------------------------------------------------------------------------------
1                             2                        none
-------------------------------------------------------------------------------
```

```
Number of Entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

```
*A:PE-2# show service id 2 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address                            Egress VNI          Oper      Vxlan
                                                            State     Type
-------------------------------------------------------------------------------
192.0.2.1                               2                   Up        static
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

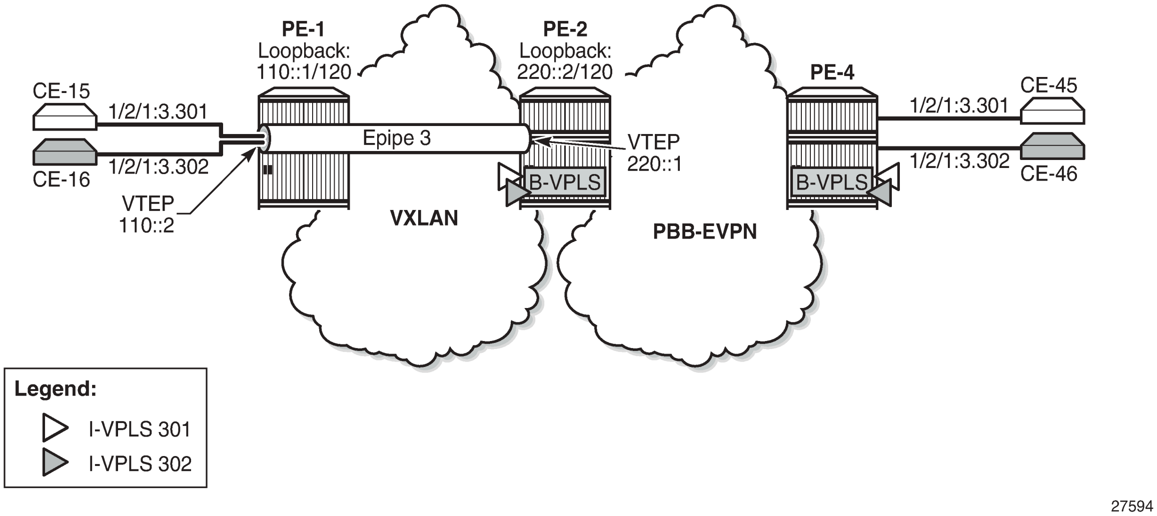
Static VXLAN termination on IPv6 addresses

IPv6 VXLAN termination is provisioned as follows:

1. Create FPE
2. Associate FPE with VXLAN termination
3. Configure router loopback interface
4. Configure non-system VXLAN termination VTEP addresses
5. Add the service configuration

Figure 293: Example topology for static VXLAN termination on IPv6 addresses shows the example topology with PE-1 and PE-2 in a VXLAN network. The loopback addresses on PE-1 and PE-2 will be used for IPv6 VXLAN termination. The existing PXC 1 on PE-2 is reused for FPE; only an IPv6 VTEP address needs to be added.

Figure 293: Example topology for static VXLAN termination on IPv6 addresses



For IPv6 routing, the following option is configured for IS-IS on all nodes:

```
# on all PEsL
configure
    router Base
        isis 0
            ipv6-routing native
```

Create FPE

The following PXC is created on PE-1; PXC 1 will be used for FPE:

```
# on PE-1:
configure
    port-xc
        pxc 1 create
            port 1/2/5
            no shutdown
        exit
```

The PXC sub-ports and ports are enabled as follows:

```
# on PE-1:
configure
    port pxc-1.a
        ethernet
            encap-type dot1q
        exit
        no shutdown
    exit
    port pxc-1.b
        ethernet
            encap-type dot1q
        exit
        no shutdown
    exit
    port 1/2/5
        no shutdown
    exit
```

```
*A:PE-1# show port pxc 1

===============================================================================
Ports on Port Cross Connect 1
===============================================================================
Port          Admin Link Port    Cfg  Oper LAG/ Port Port Port   C/QS/S/XFP/
Id            State      State   MTU  MTU  Bndl Mode Encp Type   MDIMDX
-------------------------------------------------------------------------------
pxc-1.a       Up    Yes  Up      1574 1574    - hybr dotq xgige
pxc-1.b       Up    Yes  Up      1574 1574    - hybr dotq xgige
===============================================================================
```

FPE 1 uses PXC 1:

```
# on PE-1:
configure
    fwd-path-ext
        fpe 1 create
            path pxc 1
        exit
```

The following shows that FPE 1 uses PXC 1 and has no VXLAN termination associated:

```
*A:PE-1# show fwd-path-ext fpe 1

===============================================================================
FPE Id: 1
===============================================================================
Description        : (Not Specified)
Path               : pxc 1
Pw Port            : Disabled                          Oper     : down
Sub Mgmt Extension : Disabled                          Oper     : N/A
Vxlan Termination  : Disabled                          Oper     : down
Segment-Routing V6 : Disabled
===============================================================================
```

Associate FPE with VXLAN termination

The following command associates FPE 1 with VXLAN termination:

```
# on PE-1:
configure
    fwd-path-ext
        sdp-id-range from 10000 to 10127
        fpe 1 create
            path pxc 1
            vxlan-termination
        exit
```

The following shows the range of SDP IDs for FPE and the list of configured FPEs. The application for FPE 1 is VXLAN termination.

```
*A:PE-1# show fwd-path-ext
```

```
===============================================================================
FPE Info
===============================================================================
FPE Id                   Path                          Application
                         pxc/xc-a, xc-b
-------------------------------------------------------------------------------
1                        pxc 1                         vxlan-term
-------------------------------------------------------------------------------
Number of entries : 1
-------------------------------------------------------------------------------
SDP-Id Range: 10000 - 10127
===============================================================================
```

After the FPEs are associated with VXLAN termination, the system creates two internal router interfaces per FPE, one per PXC sub-port, as follows:

```
*A:PE-1# show router interface

===============================================================================
Interface Table (Router: Base)
===============================================================================
Interface-Name                   Adm       Opr(v4/v6)  Mode    Port/SapId
   IP-Address                                                  PfxState
-------------------------------------------------------------------------------
_tmnx_fpe_1.a                    Up        Up/Up       Network pxc-1.a:1
   fe80::100/64                                                PREFERRED
_tmnx_fpe_1.b                    Up        Up/Up       Network pxc-1.b:1
   fe80::101/64                                                PREFERRED
---snip---
```

Configure router loopback interface

The following loopback interface is configured in PE-1 and added to the IS-IS context:

```
# on PE-1:
configure
    router Base
        interface "loopback1"
            address 1.1.0.1/24
            loopback
            ipv6
                address 110::1/120
            exit
        exit
        isis 0
            interface "loopback1"
                no shutdown
            exit
        exit
```

All IPv6 addresses in the 110::/120 subnet can be used for VXLAN tunnel termination, except for 110::1.

Configure non-system VTEP addresses

On PE-1, IPv6 address 110::2 in the subnet of the loopback address 110::1/120 is configured as VTEP, as follows:

```
# on PE-1:
configure
    service
        system
            vxlan
                tunnel-termination 110::2 fpe 1 create
            exit
        exit
```

On PE-2, IPv6 address 220::1 in the subnet of the loopback address 220::2/120 is configured as VTEP, as follows:

```
# on PE-2:
configure
    service
        system
            vxlan
                tunnel-termination 220::1 fpe 1 create
            exit
        exit
```

When the IPv6 VTEP address is configured on PE-1, an internal loopback interface _tmnx_vli_vxlan_1_131075 is created, as follows.

```
*A:PE-1# show router interface

===============================================================================
Interface Table (Router: Base)
===============================================================================
Interface-Name                   Adm       Opr(v4/v6)  Mode    Port/SapId
   IP-Address                                                  PfxState
-------------------------------------------------------------------------------
_tmnx_fpe_1.a                    Up        Up/Up       Network pxc-1.a:1
   fe80::100/64                                                PREFERRED
_tmnx_fpe_1.b                    Up        Up/Up       Network pxc-1.b:1
   fe80::101/64                                                PREFERRED
_tmnx_vli_vxlan_1_131075         Up        Down/Up     Network loopback
   110::2/128                                                  PREFERRED
   fe80::f:ffff:fe00:0/64                                      PREFERRED
---snip---
```

The following IPv6 route table on PE-1 contains an internal static route for source VTEP 110::2/128 using the FPE internal interface _tmnx_fpe_1.a:

```
*A:PE-1# show router route-table ipv6

===============================================================================
IPv6 Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
110::/120                                     Local   Local     01h34m32s  0
       loopback1                                                    0
```

```
110::2/128                                    Remote  Static    00h33m20s  5
       fe80::101-"_tmnx_fpe_1.a"                                    1
220::/120                                     Remote  ISIS      00h15m03s   15
       fe80::616:1ff:fe01:2-"int-PE-1-PE-2"                         10
---snip---
```

The following IPv6 route table on PE-2 shows that an internal static route is configured for the source VTEP 220::1/128 using the FPE internal interface _tmnx_fpe_1.a:

```
*A:PE-2# show router route-table ipv6

===============================================================================
IPv6 Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
110::/120                                     Remote  ISIS      00h00m46s   15
       fe80::10:1ff:fe01:1-"int-PE-2-PE-1"                          10
220::/120                                     Local   Local     00h05m08s  0
       loopback1                                                    0
220::1/128                                    Remote  Static    00h00m24s  5
       fe80::101-"_tmnx_fpe_1.a"                                    1
---snip---
```

Configure the services

Epipe 3 is configured on PE-1 with **vxlan-src-vtep 110::2**, which is the VTEP address configured in the preceding step (VXLAN tunnel termination). The egress VTEP is 220::1, which is the VXLAN termination configured on PE-2.

```
# on PE-1:
configure
    service
        epipe 3 name "Epipe 3" customer 1 create
            vxlan-src-vtep 110::2
            vxlan instance 1 vni 3 create
                egr-vtep 220::1
                exit
            exit
            sap 1/2/1:3.* create
                no shutdown
            exit
            no shutdown
        exit
```

Epipe 3 on PE-2 has VXLAN source VTEP 220::1 and egress VTEP 110::2.

```
# on PE-2:
configure
    service
        epipe 3 name "Epipe 3" customer 1 create
            vxlan-src-vtep 220::1
            vxlan instance 1 vni 3 create
                egr-vtep 110::2
                exit
            exit
            sap pxc-21.a:3.* create
                no shutdown
            exit
            no shutdown
        exit
```

The configuration of the B-VPLS is the same as in the preceding example. The configuration of I-VPLS 302 is similar.

```
# on PE-2:
configure
    service
        vpls 301 name "I-VPLS 301" customer 1 i-vpls create
            pbb
                backbone-vpls 100
                exit
            exit
            sap pxc-21.b:3.301 create
                no shutdown
            exit
            no shutdown
        exit
```

The following **show** commands on PE-1 show that the VXLAN source VTEP IP address is 110::2 and the egress VTEP is 220::1. The VXLAN tunnel is operationally up.

```
*A:PE-1# show service id 3 vxlan
===============================================================================
Vxlan Src Vtep IP: 110::2
===============================================================================

===============================================================================
Vxlan Instance
===============================================================================
VXLAN Instance                VNI                      Oper-flags
-------------------------------------------------------------------------------
1                             3                        none
-------------------------------------------------------------------------------
Number of Entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

```
*A:PE-1# show service id 3 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address                            Egress VNI          Oper      Vxlan
                                                            State     Type
-------------------------------------------------------------------------------
220::1                                  3                   Up        static
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

The same commands on PE-2 show VXLAN source VTEP 220::1 and egress VTEP 110::2, as follows:

```
*A:PE-2# show service id 3 vxlan
===============================================================================
Vxlan Src Vtep IP: 220::1
```

```
===============================================================================

===============================================================================
Vxlan Instance
===============================================================================
VXLAN Instance                VNI                      Oper-flags
-------------------------------------------------------------------------------
1                             3                        none
-------------------------------------------------------------------------------
Number of Entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

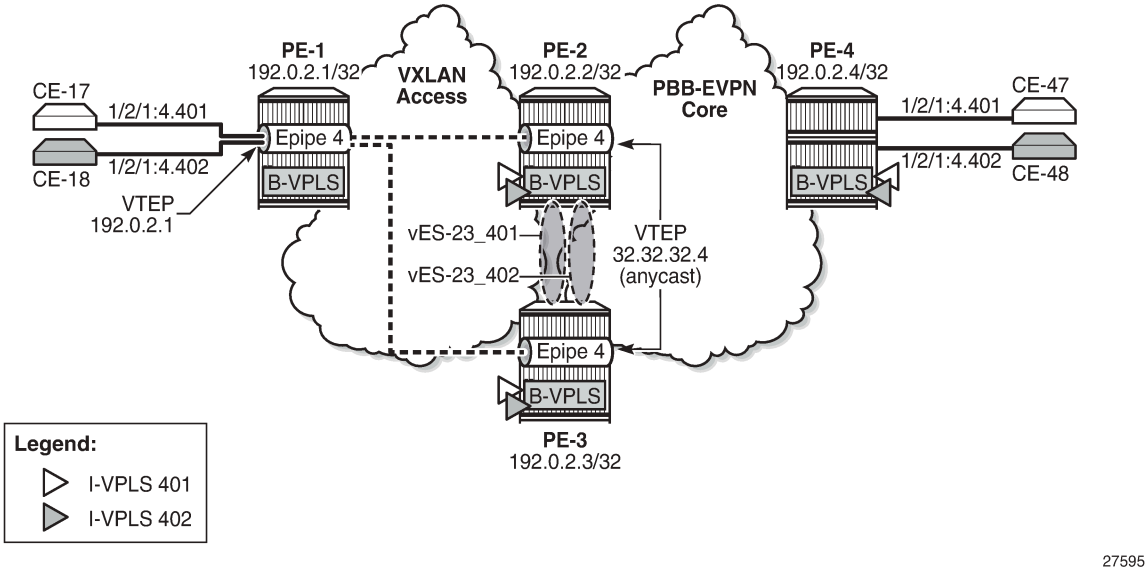
```
*A:PE-2# show service id 3 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address                            Egress VNI          Oper      Vxlan
                                                            State     Type
-------------------------------------------------------------------------------
110::2                                  3                   Up        static
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

Static VXLAN used as access network for PBB-EVPN core: all-active multi-homing and anycast VTEPs

Figure 294: Example topology for static VXLAN termination using anycast shows the example topology with PE-1, PE-2, and PE-3 in the VXLAN access network. Epipe 4 is configured on PE-1, PE-2, and PE-3. On PE-1, the system IP address 192.0.2.1 is used as source VTEP, while (anycast) IP address 23.23.23.4 is used as source VTEP on PE-2 and PE-3.

In the PBB-EVPN core network, all-active multi-homing virtual Ethernet segments vES-23_401 and vES-23_402 are configured on PE-2 and PE-3.

Figure 294: Example topology for static VXLAN termination using anycast



VXLAN access network

On PE-2 and PE-3, PXC ports are configured: PXC 2 will be used as FPE, whereas PXC-3 and PXC-4 will be used to make a LAG for the PXC between Epipe and I-VPLS services. The configuration of the PXC ports is as follows:

```
# on PE-2, PE-3:
configure
    port-xc
        pxc 2 create
            port 1/2/6
            no shutdown
        exit
        pxc 3 create
            port 1/2/7
            no shutdown
        exit
        pxc 4 create
            port 1/2/8
            no shutdown
        exit
```

The PXC sub-ports for FPE have dot1q encapsulation whereas the PXC sub-ports for port cross-connect have qinq encapsulation. The sub-ports and ports are enabled, as follows:

```
# on PE-2, PE-3:
configure
    port pxc-2.a
        ethernet
            encap-type dot1q
        exit
        no shutdown
    exit
    port pxc-2.b
        ethernet
            encap-type dot1q
        exit
        no shutdown
    exit
    port pxc-3.a
        ethernet
            encap-type qinq
        exit
        no shutdown
    exit
    port pxc-3.b
        ethernet
            encap-type qinq
        exit
        no shutdown
    exit
    port pxc-4.a
        ethernet
            encap-type qinq
        exit
        no shutdown
    exit
    port pxc-4.b
        ethernet
            encap-type qinq
        exit
        no shutdown
    exit all
    port 1/2/6
        no shutdown
    exit
    port 1/2/7
        no shutdown
    exit
    port 1/2/8
        no shutdown
    exit
```

On PE-2 and PE-3, FPE 2 is configured as follows:

```
# on PE-2, PE-3:
configure
    fwd-path-ext
        fpe 2 create
            path pxc 2
        exit
```

FPE 2 is associated with VXLAN termination and two internal interfaces will be auto-created: _tmnx_fpe_2.a and _tmnx_fpe_2.b.

```
# on PE-2, PE-3:
configure
    fwd-path-ext
        sdp-id-range from 10000 to 10127
        fpe 2 create
            path pxc 2
            vxlan-termination
        exit
```

```
*A:PE-2# show router interface

===============================================================================
Interface Table (Router: Base)
===============================================================================
Interface-Name                   Adm       Opr(v4/v6)  Mode    Port/SapId
   IP-Address                                                  PfxState
-------------------------------------------------------------------------------
_tmnx_fpe_1.a                    Up        Up/Up       Network pxc-1.a:1
   fe80::100/64                                                PREFERRED
_tmnx_fpe_1.b                    Up        Up/Up       Network pxc-1.b:1
   fe80::101/64                                                PREFERRED
_tmnx_fpe_2.a                    Up        Up/Up       Network pxc-2.a:1
   fe80::200/64                                                PREFERRED
_tmnx_fpe_2.b                    Up        Up/Up       Network pxc-2.b:1
   fe80::201/64                                                PREFERRED
---snip---
```

A router loopback interface with IP address 23.23.23.2/24 is created on PE-2, and on PE-3 with IP address 23.23.23.3/24:

```
# on PE-2:
configure
    router Base
        interface "loopback2"
            address 23.23.23.2/24
            loopback
            no shutdown
        exit
        isis 0
            interface "loopback2"
                no shutdown
            exit
        exit
```

On PE-2 and PE-3, the VTEP 23.23.23.4 is configured for FPE 2, as follows:

```
# on PE-2, PE-3:
configure
    service
        system
            vxlan
                tunnel-termination 23.23.23.4 fpe 2 create
            exit
```

The following command shows an additional VTEP 23.23.23.4 to the existing router interface _tmnx_vli_vxlan_1_131075 on PE-2:

```
*A:PE-2# show router interface "_tmnx_vli_vxlan_1_131075"

===============================================================================
Interface Table (Router: Base)
===============================================================================
Interface-Name                   Adm       Opr(v4/v6)  Mode    Port/SapId
   IP-Address                                                  PfxState
-------------------------------------------------------------------------------
_tmnx_vli_vxlan_1_131075         Up        Up/Up       Network loopback
   2.2.0.1/32                                                  n/a
   220::1/128                                                  PREFERRED
   23.23.23.4/32                                               n/a
   fe80::13:ffff:fe00:0/64                                     PREFERRED
-------------------------------------------------------------------------------
Interfaces : 1
===============================================================================
```

On PE-2 and PE-3, the VXLAN Epipe 4 uses LAG 4 (composed of pxc-3.b and pxc-4.b) to extend the VXLAN toward the I-VPLSs 401 and 402. The I-VPLS SAPs use LAG 3 (composed of pxc-3.a and pxc-4.a). The PXC LAGs provide higher bandwidth and better resiliency. The LAGs are configured as follows on both PE-2 and PE-3:

```
# on PE-2, PE-3:
configure
    lag 3
        mode hybrid
        encap-type qinq
        port pxc-3.a
        port pxc-4.a
        no shutdown
    exit
    lag 4
        mode hybrid
        encap-type qinq
        port pxc-3.b
        port pxc-4.b
        no shutdown
    exit
```

Epipe 4 is configured on PE-1, PE-2, and PE-3. On PE-1, no FPE is required because the system IP address is used as VTEP. Epipe 4 is configured on PE-1 with egress VTEP 23.23.23.4, as follows:

```
# on PE-1:
configure
    service
        epipe 4 name "Epipe 4" customer 1 create
            vxlan instance 1 vni 4 create
                egr-vtep 23.23.23.4
                exit
            exit
            sap 1/2/1:4.* create
                no shutdown
            exit
            no shutdown
        exit
```

Epipe 4 is configured on PE-2 and PE-3 with source VTEP 23.23.23.4 and egress VTEP 192.0.2.1, as follows. The SAP uses LAG 4, which is composed of PXC sub-ports pxc-3.b and pxc-4.b.

```
# on PE-2, PE-3:
configure
    service
        epipe 4 name "Epipe 4" customer 1 create
            vxlan-src-vtep 23.23.23.4
            vxlan instance 1 vni 4 create
                egr-vtep 192.0.2.1
                exit
            exit
            sap lag-4:4.* create
                no shutdown
            exit
            no shutdown
        exit
```

The following command on PE-1 shows that the egress VTEP in Epipe 4 equals 23.23.23.4.

```
*A:PE-1# show service id 4 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address                            Egress VNI          Oper      Vxlan
                                                            State     Type
-------------------------------------------------------------------------------
23.23.23.4                              4                   Up        static
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

The following commands for Epipe 4 on PE-2 show a source VTEP equal to 23.23.23.4 and an egress VTEP equal to the system address of PE-1 (192.0.2.1), as follows:

```
*A:PE-2# show service id 4 vxlan
===============================================================================
Vxlan Src Vtep IP: 23.23.23.4
===============================================================================

===============================================================================
Vxlan Instance
===============================================================================
VXLAN Instance                VNI                      Oper-flags
-------------------------------------------------------------------------------
1                             4                        none
-------------------------------------------------------------------------------
Number of Entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

```
*A:PE-2# show service id 4 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
VTEP Address                            Egress VNI          Oper      Vxlan
                                                            State     Type
-------------------------------------------------------------------------------
192.0.2.1                               4                   Up        static
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================
---snip---
```

The output on PE-3 is identical: source VTEP 23.23.23.4 and egress VTEP 192.0.2.1.

The following route table on PE-1 shows that the best route toward 23.23.23.4 is via PE-2:

```
*A:PE-1# show router route-table 23.23.23.4

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
23.23.23.0/24                                 Remote  ISIS      00h04m13s  15
       192.168.12.2                                                 10
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

PBB-EVPN core network

Two all-active multi-homing virtual ESs are configured on PE-2 and PE-3. The preference for the DF election is configured manually, with opposite preference values for the vESs so that DF load balancing is achieved. While vES-23_401 has preference 5000 on PE-2 and preference 10000 on PE-3, vES-23_402 has preference 10000 on PE-2 and preference 5000 on PE-3. When no event has occurred that caused a DF switchover, PE-2 is DF for vES-23_402 and PE-3 is DF for vES-23_401. Both vESs use LAG 3, which is composed of pxc-3.a and pxc-4.a. For vES-23_401, the qinq encapsulation must match S-tag 4 and C-tag 401; for vES-23_402, the S-tag must be 4 and the C-tag 402. On PE-2, the vESs are configured as follows.

```
# on PE-2:
configure
    service
        system
            bgp-evpn
                ethernet-segment "vES-23_401" virtual create
                    esi 01:00:00:00:23:04:01:00:00:01
                    source-bmac-lsb 23-41 es-bmac-table-size 8
                    service-carving
                        mode manual
                        manual
                            preference non-revertive create
                                value 5000
                            exit
                        exit
                    exit
                    multi-homing all-active
                    lag 3
                    qinq
                        s-tag 4 c-tag-range 401
                    exit
                    no shutdown
                exit
                ethernet-segment "vES-23_402" virtual create
                    esi 01:00:00:00:23:04:02:00:00:01
                    source-bmac-lsb 23-42 es-bmac-table-size 8
                    service-carving
                        mode manual
                        manual
                            preference non-revertive create
                                value 10000
                            exit
                        exit
                    exit
                    multi-homing all-active
                    lag 3
                    qinq
                        s-tag 4 c-tag-range 402
                    exit
                    no shutdown
                exit
```

The B-VPLS 100 is configured to use the ES-BMAC. On PE-2, the B-VPLS is configured as follows.

```
# on PE-2:
configure
    service
        vpls 100 name "B-VPLS 100" customer 1 b-vpls create
            service-mtu 2000
            pbb
                source-bmac 00:00:00:00:00:02
                use-es-bmac
            exit
            bgp
            exit
            bgp-evpn
                evi 100
                mpls bgp 1
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            no shutdown
```

On PE-4, the following configuration sets ECMP to a value of 2 in the **bgp-evpn mpls** context of the B-VPLS, so that aliasing is possible.

```
# on PE-4:
configure
    service
        vpls "B-VPLS 100"
            bgp-evpn
                mpls
                    ecmp 2
```

On PE-2 and PE-3, the I-VPLSs are configured with SAP LAG 3, which is composed of pxc-3.a and pxc-4.a, as follows. The qinq encapsulation 4.401 in I-VPLS 401 matches the condition in vES-23_401, whereas qinq 4.402 in I-VPLS 402 matches vES-23_402.

```
# on PE-2, PE-3:
configure
    service
        vpls 401 name "I-VPLS 401" customer 1 i-vpls create
            pbb
                backbone-vpls 100
                exit
            exit
            sap lag-3:4.401 create
                no shutdown
            exit
            no shutdown
        exit
        vpls 402 name "I-VPLS 402" customer 1 i-vpls create
            pbb
                backbone-vpls 100
                exit
            exit
            sap lag-3:4.402 create
                no shutdown
            exit
            no shutdown
        exit
```

With the preceding configuration, PBB-EVPN all-active multi-homing and the anycast VTEP at the access VXLAN network can be combined for an efficient and fully redundant network. PE-4 can alias the known unicast traffic to PE-2 and PE-3 on a per-flow basis, whereas if ECMP (and shared queuing) is enabled on PE-1, traffic can also be load-balanced to PE-2 and PE-3. BUM traffic sent from PE-4 will be forwarded by the corresponding DF for the ES.

See chapter EVPN for PBB over MPLS (PBB-EVPN) for more information about PBB-EVPN and all-active multi-homing.

Verification

The following command shows that PE-2 is NDF in vES-23_401 in I-VPLS 401:

```
*A:PE-2# show service id 401 ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
lag-3:4.401           vES-23_401                       NDF
===============================================================================
No sdp entries
No vxlan instance entries
```

For I-VPLS 402, PE-2 is DF, as follows:

```
*A:PE-2# show service id 402 ethernet-segment

===============================================================================
SAP Ethernet-Segment Information
===============================================================================
SAP                   Eth-Seg                          Status
-------------------------------------------------------------------------------
lag-3:4.402           vES-23_402                       DF
===============================================================================
No sdp entries
No vxlan instance entries
```

For PE-3, the reverse is true: PE-3 is DF in vES-23_401 for I-VPLS 401 and NDF in vES-23_402 for I-VPLS 402.

Within B-VPLS 100, the BMAC addresses are advertised via BGP-EVPN. On PE-2, the following FDB for B-VPLS 100 contains the BMAC addresses of PE-3 and PE-4, which are advertised via BGP-EVPN:

```
*A:PE-2# show service id 100 fdb detail
```

```
===============================================================================
Forwarding Database, Service 100
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
100        00:00:00:00:00:03 mpls:                   EvpnS:P  06/08/21 15:01:37
                             192.0.2.3:524279
           ldp:65540
100        00:00:00:00:00:04 mpls:                   EvpnS:P  06/08/21 15:01:37
                             192.0.2.4:524283
           ldp:65538
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

Likewise, the following FDB for B-VPLS 100 on PE-3 contains the BMAC addresses of PE-2 and PE-4:

```
*A:PE-3# show service id 100 fdb detail

===============================================================================
Forwarding Database, Service 100
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
100        00:00:00:00:00:02 mpls:                   EvpnS:P  06/08/21 15:16:08
                             192.0.2.2:524283
           ldp:65537
100        00:00:00:00:00:04 mpls:                   EvpnS:P  06/08/21 15:16:08
                             192.0.2.4:524283
           ldp:65539
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The following FDB for B-VPLS 100 on PE-4 contains the BMAC addresses of PE-2 and PE-3, but also the BMAC addresses of vES-23_401 and vES-23_402:

```
*A:PE-4# show service id 100 fdb detail

===============================================================================
Forwarding Database, Service 100
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
100        00:00:00:00:00:02 mpls:                   EvpnS:P  06/08/21 14:09:43
                             192.0.2.2:524283
           ldp:65538
100        00:00:00:00:00:03 mpls:                   EvpnS:P  06/08/21 14:50:11
                             192.0.2.3:524279
           ldp:65540
100        00:00:00:00:23:41 eES:                    EvpnS:P  06/08/21 14:50:02
                             MAX-ESI
100        00:00:00:00:23:42 eES:                    EvpnS:P  06/08/21 14:50:02
                             MAX-ESI
-------------------------------------------------------------------------------
```

```
No. of MAC Entries: 4
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

On PE-4, the following list of BGP EVPN routes for ES-BMAC 00:00:00:00:23:41 of vES-23_401 shows that PE-4 learned the ES-BMAC address via two PEs: PE-2 and PE-3.

```
*A:PE-4# show router bgp routes evpn mac mac-address 00:00:00:00:23:41
===============================================================================
 BGP Router ID:192.0.2.4        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.2:100       00:00:00:00:23:41 ESI-MAX
      0                   Static            LABEL 524283
                          n/a
                          192.0.2.2

u*>i  192.0.2.3:100       00:00:00:00:23:41 ESI-MAX
      0                   Static            LABEL 524279
                          n/a
                          192.0.2.3

-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

PE-4 also learned ES-BMAC 00:00:00:00:23:42 via PE-2 and PE-3, as follows:

```
*A:PE-4# show router bgp routes evpn mac mac-address 00:00:00:00:23:42
===============================================================================
 BGP Router ID:192.0.2.4        AS:64500       Local AS:64500
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN MAC Routes
===============================================================================
Flag  Route Dist.         MacAddr           ESI
      Tag                 Mac Mobility      Label1
                          Ip Address
                          NextHop
-------------------------------------------------------------------------------
u*>i  192.0.2.2:100       00:00:00:00:23:42 ESI-MAX
      0                   Static            LABEL 524283
                          n/a
                          192.0.2.2
```

```
u*>i  192.0.2.3:100       00:00:00:00:23:42 ESI-MAX
      0                   Static            LABEL 524279
                          n/a
                          192.0.2.3

-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

When a ping is initiated from CE-17 to CE-47, the ICMP packets are forwarded from PE-1 to PE-2, because the best route to 23.23.23.4 is via PE-2. PE-2 learns MAC address ca:fe:01:17:17:17 of CE-17 on the local I-VPLS SAP. PE-2 forwards the ICMP packets through I-VPLS 401 and B-VPLS 100 toward PE-4. PE-4 learns MAC ca:fe:01:17:17:17 of CE-17 via the ES-BMAC. When the reply is sent, PE-4 learns MAC address ca:fe:04:47:47:47 of CE-47 on the local SAP.

The FDB for I-VPLS 401 on PE-2 shows that MAC ca:fe:04:47:47:47 is learned on the local SAP and MAC ca:fe:04:47:47:47 can be reached via the B-VPLS to PE-4.

```
*A:PE-2# show service id 401 fdb detail

===============================================================================
Forwarding Database, Service 401
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
401        ca:fe:01:17:17:17 sap:lag-3:4.401         L/0      06/08/21 15:19:19
401        ca:fe:04:47:47:47 b-mpls:                 L/0      06/08/21 15:19:19
                             192.0.2.4:524283
           ldp:65538
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The following FDB for I-VPLS 401 on PE-3 shows that MAC ca:fe:04:47:47:47 is learned via BGP-EVPN from PE-4.

```
*A:PE-3# show service id 401 fdb detail
===============================================================================
Forwarding Database, Service 401
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
401        ca:fe:04:47:47:47 b-mpls:                 L/0      06/08/21 15:19:19
                             192.0.2.4:524283
           ldp:65539
-------------------------------------------------------------------------------
No. of MAC Entries: 1
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

The following FDB for I-VPLS 401 on PE-4 shows that MAC ca:fe:04:47:47:47 is learned on a local SAP, whereas MAC ca:fe:01:17:17:17 is learned via ES-BMAC 00:00:00:00:23:41 of vES-23_401.

```
*A:PE-4# show service id 401 fdb detail
```

```
===============================================================================
Forwarding Database, Service 401
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
401        ca:fe:01:17:17:17 eES-BMAC:               L/0      06/08/21 15:19:19
                             00:00:00:00:23:41
401        ca:fe:04:47:47:47 sap:1/2/1:4.401         L/0      06/08/21 15:19:19
-------------------------------------------------------------------------------
No. of MAC Entries: 2
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

Conclusion

VXLAN FPE is required to terminate non-system IPv4/IPv6 VXLAN tunnels. The examples in this chapter show how VXLAN FPE can be applied in Epipe services, to stitch static VXLAN to other services, such as I-VPLS services.

Three-byte EVI in EVPN Services

This chapter provides information about the three-byte EVI in EVPN services.

Topics in this chapter include:

Applicability

The information and configuration in this chapter are based on SR OS Release 22.10.R1. The three-byte EVI is supported in EVPN services in SR OS Release 21.10.R1 and later. Three-byte EVI values can be configured in VPLS, R-VPLS, B-VPLS, and Epipe services for MPLS, VXLAN, and SRv6 instances.

Overview

In SR OS implementations earlier than SR OS Release 21.10.R1, the EVPN instance (EVI) is defined as a two-byte integer value, providing up to 65535 unique identifiers. The EVI is a unique value per service that can be used for three purposes:

- service route target (RT) auto-derivation – autonomous system number (ASN):EVI; for example, 64496:10
- service route distinguisher (RD) auto-derivation – system IP address:EVI; for example, 192.0.2.1:10
- designated forwarder (DF) election, as described in the Preference-based and Non-revertive EVPN DF Election chapter

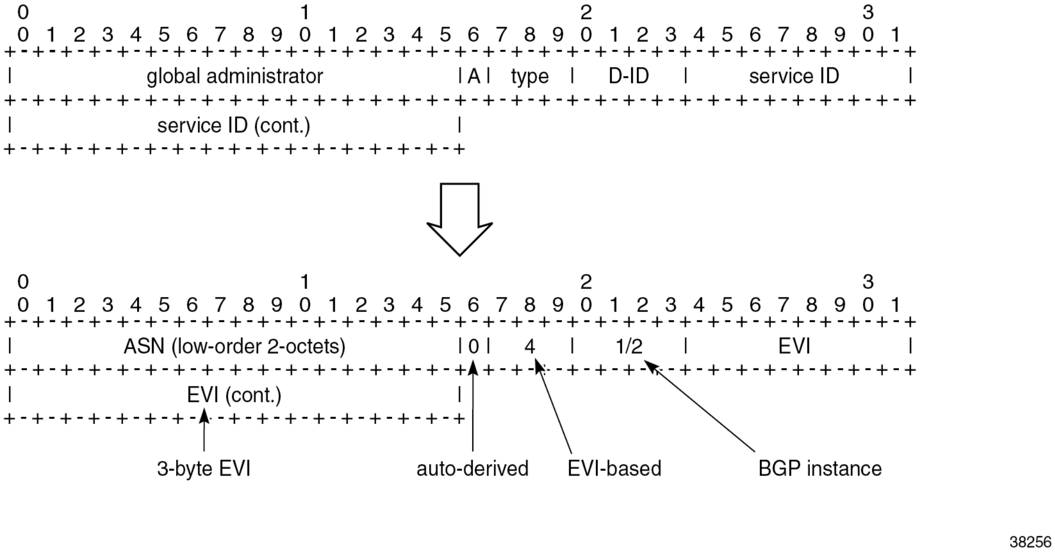
However, in large networks, more than 65535 EVI values are required if the EVI is desired to be unique network-wide. The three-byte EVI provides up to 16777215 values and is supported in SR OS Release 21.10.R1 and later.

All DF election procedures support the extended EVI range. The RD auto-derivation is only possible for the two-byte EVI; the RT auto-derivation for the three-byte EVI can be enabled with the **evi-three-byte-auto-rd** command.

Auto-derived RT

Figure 295: Auto-derived RT in RFC 8365 shows the RT auto-derivation for configured EVI values in the range up to 16777215.

Figure 295: Auto-derived RT in RFC 8365



For three-octet EVI values, the fields in the RT format are:

- the global administrator field, which contains (the lower two octets of) the autonomous system number (ASN)
- the single-bit field A, which indicates if the RT is auto-derived: A=0 for auto-derivation
- the three-bit type field, which indicates the space in which the three-byte service ID is defined:
  - 0: VID (802.1Q VLAN ID)
  - 1: VXLAN
  - 2: NVGRE
  - 3: I-SID
  - **4: EVI**
  - 5: dual-VID (QinQ VLAN ID)
- the four-bit D-ID field, which encodes the domain ID. For type 4 (EVI), the D-ID corresponds to the BGP instance ID in the EVPN service.
- the three-octet service ID, which is set to the EVI (for type 4)

As an example, in a dual-instance EVPN-VPLS service with the following characteristics:

- ASN 64496
- EVI 100002 (0x186A2)
- BGP 1 for EVPN-VXLAN; BGP 2 for EVPN-MPLS
- **evi-three-byte-auto-rt** enabled

The two auto-derived RTs are:

- 64496:1090619042 (0x4**1**0186A2) for BGP 1
- 64496:1107396258 (0x4**2**0186A2) for BGP 2

The RT can also be configured manually, for example, 64496:100002. A manually configured RT has precedence over an auto-derived RT.

Auto-derived RD

Each BGP instance in an EVPN service has an RD. Only for EVI values smaller than or equal to 65535, the RD for BGP instance 1 can be auto-derived out of the system IP address and the EVI, for example, 192.0.2.2:10. EVI values greater than 65535 do not generate RDs automatically.

The VPLS RD is selected based on the following precedence order:

- manually configured RD or auto-RD take precedence when configured,
- if there is no manual RD or auto-RD configuration, the RD is derived from the **bgp-ad>vpls-id**,
- if there is no manual RD, auto-RD, or VPLS ID configuration, the RD is derived from the EVI for EVI values up to 65535 and except for **bgp-mh** which does not support EVI-derived RD,

The Epipe RD is determined in a similar way, but there is no VPLS ID in Epipes.

The following error message is raised when attempting to enable **bgp-evpn** with an EVI value greater than 65535 without having configured a manual RD, auto-RD, or BGP-AD VPLS ID:

```
*A:PE-1>config>service>vpls>bgp-evpn>vxlan$ no shutdown
MINOR: SVCMGR #6554 no route-distinguisher configured - missing route-distinguisher or vpls-id
 or evi or evi is 3-byte
```

The RD configuration can be changed dynamically. When the RD changes, the active routes for the service are withdrawn and readvertised with the new RD.

EVI RT set for AD per-ES routes

As described in the EVPN for MPLS Tunnels chapter, Auto-discovery per Ethernet segment (AD per-ES) routes carry the ESI label and the multi-homing mode. When multiple EVIs are defined in an ES, the AD per-ES routes can be aggregated.

EVI RT set for AD per-ES routes with two-byte EVI

The following command enables the aggregation of AD per-ES routes for two-byte EVI values: **configure service system bgp-evpn ad-per-es-route-target evi-rt-set route-distinguisher <ip-address>**. The RD is specific for this EVI RT set feature. If enabled, a single AD per-ES route with the associated RD and a set of maximum 128 EVI RTs can be advertised. The EVI RTs are distributed in routes with the RD configured in the preceding command and one of the following *comm-val* values (the *comm-val* range is not configurable):

- EVIs from 1 to 128 – *comm-val* = 1
- EVIs from 129 to 256 – *comm-val* = 2
- ...
- EVIs from 65409 to 65535 – *comm-val* = 512

EVI RT set for AD per-ES routes with three-byte EVI

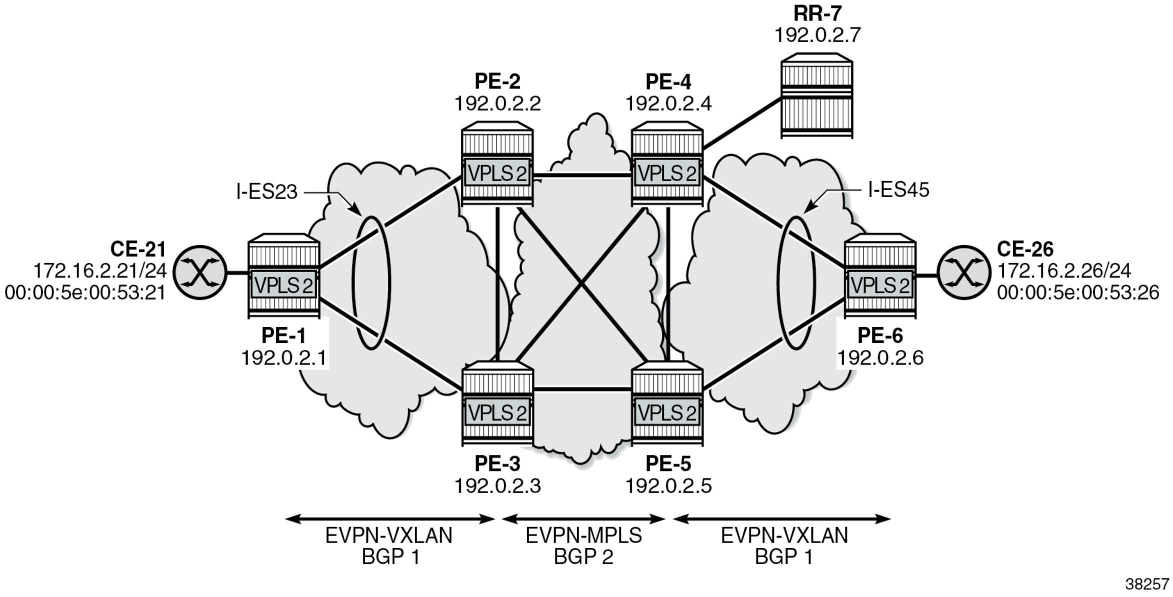
The command to enable AD per-ES route aggregation with extended EVI range is: **configure service system bgp-evpn ad-per-es-route-target evi-rt-set route-distinguisher <ip-address> extended-evi-range**. For three-byte EVIs, the *comm-val* range is extended from 512 to 65535 and the maximum number of AD per-ES routes that can be aggregated is increased from 128 to 257. The 257 RTs per route packing is done for any configured EVI, regardless of the value being greater than 65535 or not.

- EVIs from 1 to 257 – *comm-val* = 1
- EVIs from 258 to 514 – *comm-val* = 2
- ...
- EVIs from 16776961 to 16777215 – *comm-val* = 65281

Configuration

Figure 296: Example topology with dual-instance VPLS shows the example topology with dual-instance VPLS 2: VXLAN is used between PE-1, PE-2, PE-3 and also between PE-4, PE-5, and PE-6. MPLS is used between the core PEs PE-2, PE-3, PE-4, and PE-5.

Figure 296: Example topology with dual-instance VPLS



The initial configuration includes:

- cards, MDAs, ports
- router interfaces
- IS-IS level 2 between core PEs PE-2, PE-3, PE-4, PE-5, and RR-7
- IS-IS level 1 between PE-1, PE-2, and PE-3

- IS-IS level 1 between PE-6, PE-4, and PE-5
- SR-ISIS between core PEs PE-2, PE-3, PE-4, and PE-5

The BGP configuration and the used policies for dual-instance VPLSs in ESs are described in the EVPN Interconnect Ethernet Segmentschapter. Policies are required to prevent loops. RR-7 acts as route reflector for the core PEs PE-2, PE-3, PE-4, and PE-5. The policy and BGP configuration on PE-2 is as follows:

```
# on PE-2:
configure
    router Base
        policy-options
            begin
            community "vxlan"
                members "bgp-tunnel-encap:VXLAN"
            exit
            community "SOO-DCGW-23"
                members "origin:64500:23"
            exit
            policy-statement "allow only mpls"
                entry 10
                    from
                        community "vxlan"
                        family evpn
                    exit
                    action drop
                    exit
                exit
            exit
            policy-statement "allow only vxlan"
                entry 10
                    from
                        community "vxlan"
                        family evpn
                    exit
                    action accept
                    exit
                exit
                default-action drop
                exit
            exit
            policy-statement "drop SOO-DCGW-23"
                entry 10
                    from
                        community "SOO-DCGW-23"
                        family evpn
                    exit
                    action drop
                    exit
                exit
            exit
            policy-statement "add SOO to vxlan routes"
                entry 10
                    from
                        community "vxlan"
                        family evpn
                    exit
                    action accept
                        community add "SOO-DCGW-23"
                    exit
                exit
                default-action accept
                exit
            exit
            commit
        exit
        autonomous-system 64496
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "WAN"
                family evpn
                peer-as 64496
                export "allow only mpls"
                neighbor 192.0.2.7
                exit
            exit
            group "access1"
                family evpn
                peer-as 64496
                export "allow only vxlan"
                neighbor 192.0.2.1
                exit
                neighbor 192.0.2.3
                    import "drop SOO-DCGW-23"
                    export "add SOO to vxlan routes"
                exit
            exit
        exit
```

The all-active interconnect ES "I-ES23" is configured on PE-2 and PE-3; the single-active interconnect ES "I-ES45" is configured on PE-4 and PE-5. VPLS 1 with EVI 1 (0x1) and VPLS 2 with EVI 100002 (0x186A2) are configured on all PEs. Both VPLSs have BGP 1 for VXLAN and BGP 2 for MPLS (SR-ISIS) in the core. For VPLS 1, no extended EVI range is required. The RD can be auto-derived for BGP instance 1, but not for BGP instance 2. For VPLS 2, the EVI is greater than 65535, so the RD must always be configured (manual configuration or auto-RD). The RT is auto-derived in VPLS 1 and VPLS 2. For VPLS 2, the **evi-three-byte-auto-rt** command is configured to enable auto-derivation of RTs for EVI values up to 16777215. On all core PEs, **evi-rt-set** is enabled for the aggregation of AD per-ES routes. The service configuration on PE-2 is as follows:

```
# on PE-2:
configure
    service
        system
            bgp-auto-rd-range 192.0.2.2 comm-val 2000 to 2999
            bgp-evpn
                ad-per-es-route-target evi-rt-set route-distinguisher 10.0.2.2 extended-evi-
range
                ethernet-segment "I-ES23" virtual create
                    esi 00:00:00:00:00:23:23:00:00:01
                    service-carving
                        mode manual
                        manual
                            preference non-revertive create
                                value 150
                            exit
                            evi 1 to 200000
                        exit
                    exit
                    multi-homing all-active
                    network-interconnect-vxlan 1
                    service-id
                        service-range 1 to 2
                    exit
                    no shutdown
                exit
            exit
        exit
        vpls 1 name "VPLS-1" customer 1 create
            vxlan instance 1 vni 1 create
            exit
            bgp
                # route-distinguisher 192.0.2.2:1   # will be auto-derived
            exit
            bgp 2
                route-distinguisher auto-rd
            exit
            bgp-evpn
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
                mpls bgp 2
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution-filter
                            sr-isis
                        exit
                        resolution filter
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
        vpls 2 name "VPLS-2" customer 1 create
            vxlan instance 1 vni 2 create
            exit
            bgp
                route-distinguisher auto-rd    # RD cannot be auto-derived from EVI
            exit
            bgp 2
                route-distinguisher auto-rd
            exit
            bgp-evpn
                evi 100002
                vxlan bgp 1 vxlan-instance 1
                    evi-three-byte-auto-rt
                    no shutdown
                exit
                mpls bgp 2
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    evi-three-byte-auto-rt
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
```

When configuring the **evi-rt-set** command, the RD must be different from the RD in the auto-rd range. If not, the following error message is raised:

```
*A:PE-2>config>service>system>bgp-evpn# ad-per-es-route-target evi-rt-set route-distinguisher
 192.0.2.2 extended-evi-range
MINOR: SVCMGR #7905 Range is in use - auto-rd range exists with same ip-address
```

The following command shows the RD and RT values for both BGP instances in VPLS 1 on PE-2:

```
*A:PE-2# show service id 1 bgp

===============================================================================
BGP Information
===============================================================================
Bgp Instance         : 1
Vsi-Import           : None
Vsi-Export           : None
Route Dist           : None
Oper Route Dist      : 192.0.2.2:1
Oper RD Type         : derivedEvi
Rte-Target Import    : None                 Rte-Target Export: None
Oper RT Imp Origin   : derivedEvi           Oper RT Import   : 64496:1
Oper RT Exp Origin   : derivedEvi           Oper RT Export   : 64496:1
ADV Service MTU      : -1

Bgp Instance         : 2
Vsi-Import           : None
Vsi-Export           : None
Route Dist           : auto-rd
Oper Route Dist      : 192.0.2.2:2000
Oper RD Type         : auto
Rte-Target Import    : None                 Rte-Target Export: None
Oper RT Imp Origin   : derivedEvi           Oper RT Import   : 64496:1
Oper RT Exp Origin   : derivedEvi           Oper RT Export   : 64496:1
ADV Service MTU      : -1

PW-Template Id       : None
-------------------------------------------------------------------------------
===============================================================================
```

RD 192.0.2.2:1 for BGP instance 1 is auto-derived whereas RD 192.0.2.2:2000 for BGP instance 2 is the result of auto-RD. RT 64496:1 is auto-derived based on the system IP address and the EVI. It is possible to configure **evi-three-byte-auto-rt** in VPLS 1, even though the EVI value is smaller than 65535. To reconfigure VPLS 1 with evi-three-byte-auto-rd in both BGP instances, EVPN-VXLAN and EVPN-MPLS must be disabled, as follows:

```
# on PE-2:
configure
    service
        vpls "VPLS-1"
            bgp-evpn
                vxlan bgp 1
                    shutdown
                    evi-three-byte-auto-rt
                    no shutdown
                exit
                mpls bgp 2
                    shutdown
                    evi-three-byte-auto-rt
                    no shutdown
                exit
            exit
```

In this case, the auto-derivation is based on RFC 8365 and the value is 64496:1090519041 (0x41000001) for BGP 1 and 64496:1107296257 (0x42000001) for BGP 2.

```
*A:PE-2# show service id 1 bgp

===============================================================================
BGP Information
===============================================================================
Bgp Instance         : 1
Vsi-Import           : None
Vsi-Export           : None
Route Dist           : None
Oper Route Dist      : 192.0.2.2:1
Oper RD Type         : derivedEvi
Rte-Target Import    : None                 Rte-Target Export: None
Oper RT Imp Origin   : derivedEvi           Oper RT Import   : 64496:1090519041
Oper RT Exp Origin   : derivedEvi           Oper RT Export   : 64496:1090519041
ADV Service MTU      : -1

Bgp Instance         : 2
Vsi-Import           : None
Vsi-Export           : None
Route Dist           : auto-rd
Oper Route Dist      : 192.0.2.2:2000
Oper RD Type         : auto
Rte-Target Import    : None                 Rte-Target Export: None
Oper RT Imp Origin   : derivedEvi           Oper RT Import   : 64496:1107296257
Oper RT Exp Origin   : derivedEvi           Oper RT Export   : 64496:1107296257
ADV Service MTU      : -1

PW-Template Id       : None
-------------------------------------------------------------------------------
===============================================================================
```

The auto-derived RTs on the other nodes are the same as on PE-2 when the BGP instance is the same. On PE-2, PE-3, PE-4, and PE-5, BGP 1 is for VXLAN and BGP 2 is for MPLS in the core. That way, EVPN messages can be exchanged in BGP instance 2 between the core PEs.

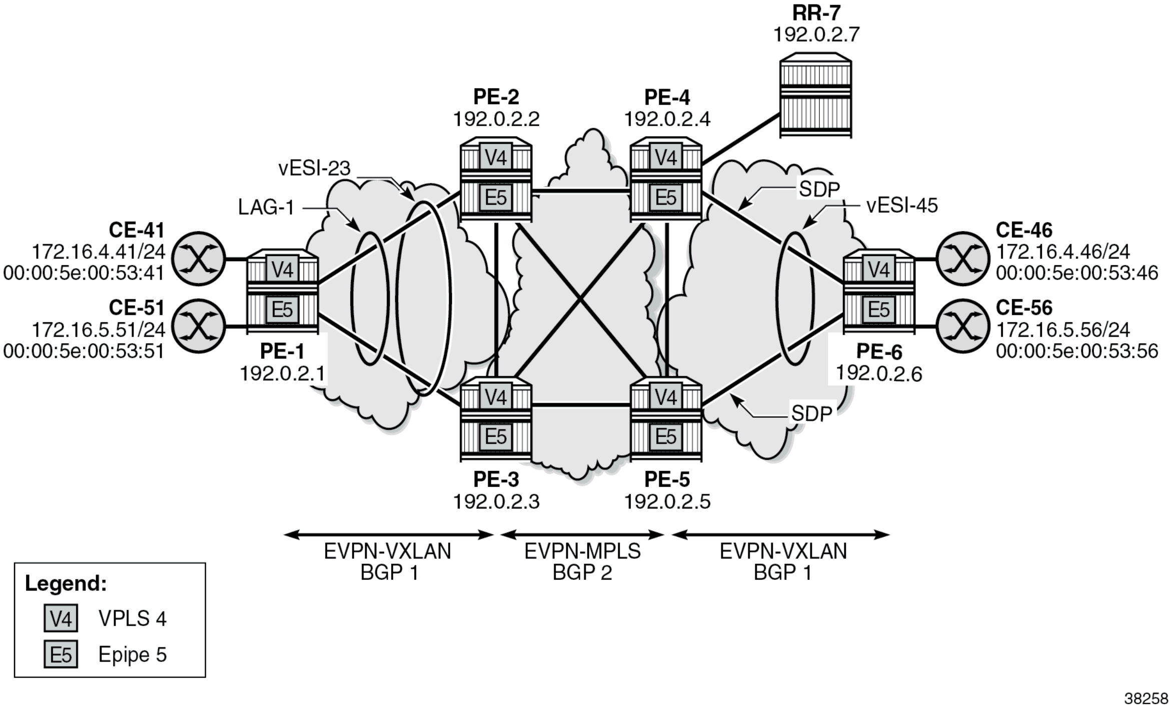
AD per-ES aggregation is enabled on the core nodes. The following command on PE-2 shows one AD per-ES with RD 10.0.2.4:390 (10.0.2.4 is the RD configured for **evi-rt-set** and 390 is the *comm-val* value for the EVI range from 99974 to 100230) and one AD per-EVI with RD 192.0.2.4:2002 (auto-RD).

```
*A:PE-2# show router bgp routes evpn auto-disc detail
===============================================================================
 BGP Router ID:192.0.2.2        AS:64496       Local AS:64496
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP EVPN Auto-Disc Routes
```

```
===============================================================================
---snip---
                            ## AD per-ES
Network        : n/a
Nexthop        : 192.0.2.4
Path Id        : None
From           : 192.0.2.7
Res. Nexthop   : 192.168.24.2
Local Pref.    : 100                    Interface Name : int-PE-2-PE-4
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64496:1107396258    # auto-derived RT
                 esi-label:524271/Single-Active
Cluster        : 192.0.2.7
Originator Id  : 192.0.2.4              Peer Router Id : 192.0.2.7
Flags          : Used Valid Best IGP
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : AUTO-DISC
ESI            : 00:00:00:00:00:45:45:00:00:01
Tag            : MAX-ET              # AD per-ES has MAX-ET
Route Dist.    : 10.0.2.4:390        # RD for evi-rt-set
MPLS Label     : LABEL 0
Route Tag      : 0
Neighbor-AS    : n/a
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h02m48s
---snip---

                            ## AD per-EVI
Network        : n/a
Nexthop        : 192.0.2.4
Path Id        : None
From           : 192.0.2.7
Res. Nexthop   : 192.168.24.2
Local Pref.    : 100                    Interface Name : int-PE-2-PE-4
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None                   IGP Cost       : 10
Connector      : None
Community      : target:64496:1107396258 bgp-tunnel-encap:MPLS      # auto-RT
Cluster        : 192.0.2.7
Originator Id  : 192.0.2.4              Peer Router Id : 192.0.2.7
Flags          : Used Valid Best IGP
Route Source   : Internal
AS-Path        : No As-Path
EVPN type      : AUTO-DISC
ESI            : 00:00:00:00:00:45:45:00:00:01
Tag            : 0                  # AD per-EVI has Ethernet tag 0
Route Dist.    : 192.0.2.4:2002     # RD for BGP 2 in VPLS 2
MPLS Label     : LABEL 524268
Route Tag      : 0
Neighbor-AS    : n/a
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 00h06m56s
```

Figure 297: Example topology with VPLS 4 and Epipe 5 shows an example topology with EVPN-MPLS in the core, all-active ES "vESI-23" on PE-2 and PE-3, and single-active ES "vESI-45" on PE-4 and PE-5. VPLS 4 with EVI 100004 is configured on all PEs with manually configured RD and RT.

Figure 297: Example topology with VPLS 4 and Epipe 5



The configuration on PE-2 is as follows:

```
# on PE-2:
configure
    service
        system
            bgp-evpn
                ethernet-segment "vESI-23" virtual create
                    esi 01:00:00:00:00:23:03:04:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 1
                    dot1q
                        q-tag-range 3 to 10
                    exit
                    no shutdown
                exit
            exit
        exit
        vpls 4 name "VPLS-4" customer 1 create
            bgp
                route-distinguisher 192.0.2.2:4
                route-target export target:64496:100004 import target:64496:100004
            exit
            bgp-evpn
```

```
                evi 100004
                mpls bgp 1
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    # evi-three-byte-auto-rt     # not required - manual RT
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap lag-1:4 create
                no shutdown
            exit
            no shutdown
        exit
```

With the configured RD 192.0.2.2:4 and RT 64496:100004, the following BGP information is retrieved on PE-2 for VPLS 4:

```
*A:PE-2# show service id 4 bgp

===============================================================================
BGP Information
===============================================================================
Bgp Instance         : 1
Vsi-Import           : None
Vsi-Export           : None
Route Dist           : 192.0.2.2:4
Oper Route Dist      : 192.0.2.2:4
Oper RD Type         : configured
Rte-Target Import    : 64496:100004         Rte-Target Export: 64496:100004
Oper RT Imp Origin   : configured           Oper RT Import   : 64496:100004
Oper RT Exp Origin   : configured           Oper RT Export   : 64496:100004
ADV Service MTU      : -1

PW-Template Id       : None
-------------------------------------------------------------------------------
===============================================================================
```

Epipe 5 is configured on all PEs with EVI 100005 (0x186A5) and **evi-three-byte-auto-rt** enabled. The configuration on PE-2 is as follows:

```
# on PE-2:
configure
    service
        epipe 5 name "Epipe-5" customer 1 create
            bgp
                route-distinguisher auto-rd
            exit
            bgp-evpn
                local-attachment-circuit AC-ESI-23-PE-1 create
                    eth-tag 231
                exit
                remote-attachment-circuit AC-ESI-45-PE-6 create
                    eth-tag 456
                exit
                evi 100005
                mpls bgp 1
```

```
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    evi-three-byte-auto-rt
                    no shutdown
                exit
            exit
            sap lag-1:5 create
                no shutdown
            exit
            no shutdown
```

The auto-derived RT is 64496:1090619045 (0x410186A5):

```
*A:PE-2# show service id 5 bgp

===============================================================================
BGP Information
===============================================================================
Vsi-Import           : None
Vsi-Export           : None
Route Dist           : auto-rd
Oper Route Dist      : 192.0.2.2:2003
Oper RD Type         : auto
Rte-Target Import    : None                 Rte-Target Export: None
Oper RT Imp Origin   : derivedEvi           Oper RT Import   : 64496:1090619045
Oper RT Exp Origin   : derivedEvi           Oper RT Export   : 64496:1090619045
ADV Service MTU      : -1
PW-Template Id       : None
-------------------------------------------------------------------------------
===============================================================================
```

Instead of VXLAN or MPLS, SRv6 can be used too. As an example,VPLS 6 with EVI 100006 (0x186A6) is configured on PE-1, PE-2, PE-4, and PE-6. SRv6 is configured between PE-4 and PE-6, as described in the *Segment Routing over IPv6* chapter in the *7750 SR and 7950 XRS Segment Routing and PCE Advanced Configuration Guide for Classic CLI*. The configuration on PE-2 is as follows:

```
# on PE-2
configure
    service
        vpls 6 name "VPLS-6" customer 1 create
            vxlan instance 1 vni 6 create
            exit
            segment-routing-v6 1 create
                locator "PE-2_loc"
                    function
                        end-dt2u
                        end-dt2m
                    exit
                exit
            exit
            bgp
                route-distinguisher auto-rd
            exit
            bgp 2
                route-distinguisher auto-rd
            exit
            bgp-evpn
                evi 100006
                vxlan bgp 1 vxlan-instance 1
                    evi-three-byte-auto-rt
                    no shutdown
                exit
                segment-routing-v6 bgp 2 srv6-instance 1 default-locator "PE-2_loc" create
                    ecmp 2
                    evi-three-byte-auto-rt
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
```

On PE-2, RT 64496:1090619046 (0x410186A6) is auto-derived for BGP 1 and RT 64496:1107396262 (0x420186A6) for BGP 2:

```
*A:PE-2# show service id 6 bgp

===============================================================================
BGP Information
===============================================================================
Bgp Instance         : 1
Vsi-Import           : None
Vsi-Export           : None
Route Dist           : auto-rd
Oper Route Dist      : 192.0.2.2:2004
Oper RD Type         : auto
Rte-Target Import    : None                Rte-Target Export: None
Oper RT Imp Origin   : derivedEvi          Oper RT Import   : 64496:1090619046
Oper RT Exp Origin   : derivedEvi          Oper RT Export   : 64496:1090619046
ADV Service MTU      : -1

Bgp Instance         : 2
Vsi-Import           : None
Vsi-Export           : None
Route Dist           : auto-rd
Oper Route Dist      : 192.0.2.2:2005
Oper RD Type         : auto
Rte-Target Import    : None                Rte-Target Export: None
Oper RT Imp Origin   : derivedEvi          Oper RT Import   : 64496:1107396262
Oper RT Exp Origin   : derivedEvi          Oper RT Export   : 64496:1107396262
ADV Service MTU      : -1

PW-Template Id       : None
-------------------------------------------------------------------------------
===============================================================================
```

Conclusion

In large networks, a three-byte EVI can be required as a unique identifier for services. RTs can be auto-derived based on a three-byte EVI, but RDs cannot be auto-derived that way.

VCCV BFD for Epipe Services

This chapter describes the VCCV BFD for Epipe services.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written based on SR OS Release 15.0.R7. The CLI in the current edition corresponds to SR OS Release 23.7.R1.

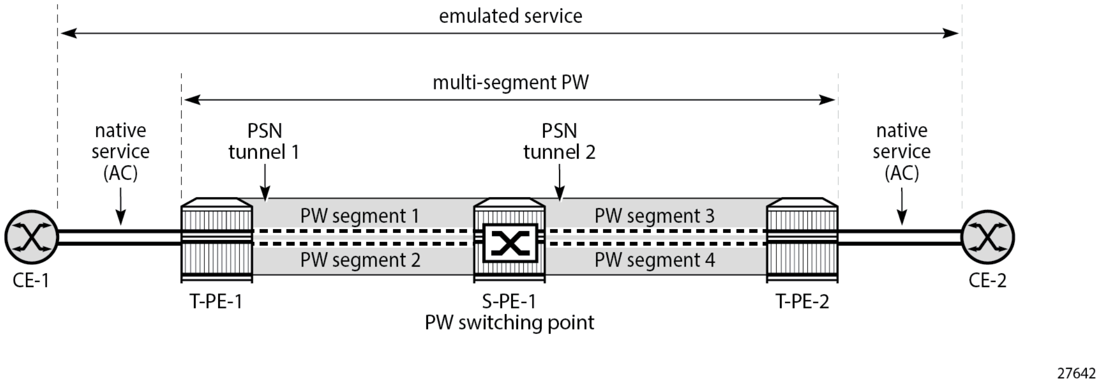
Overview

Virtual circuit connectivity verification (VCCV) is defined by RFC 5085. Bidirectional forwarding detection (BFD) is defined by RFC 5880.

VCCV is an end-to-end fault-detection tool for testing pseudowires (PWs), and typically requires an operator to take manual actions. The PWs can be used for virtual leased line (VLL), virtual private LAN service (VPLS), and Internet enhanced service (IES)/virtual private routed network (VPRN) services with Epipe or Ipipe spoke-SDPs.

SR OS supports RFC 5885 which specifies a method for carrying BFD messages in a PW-associated channel and is referred to as VCCV BFD in SR OS. Because the associated channel shares fate with the data plane, VCCV BFD monitors the PW between two terminating PEs (T-PEs), regardless of the number of provider routers or switching PEs (S-PEs) the PW may traverse; see Figure 298: PW reference model. When enabled, faults in individual PWs can be detected quickly, whether or not other provider routers or S-PEs also carry other PWs. VCCV BFD can monitor specific high-value services, where detecting forwarding failures (and potentially recovering from them) in a minimum amount of time is critical.

Figure 298: PW reference model



VCCV BFD avoids manual hop-by-hop troubleshooting of each element along the path of the PW, which minimizes the probability of not detecting silent failures on intermediate routers.

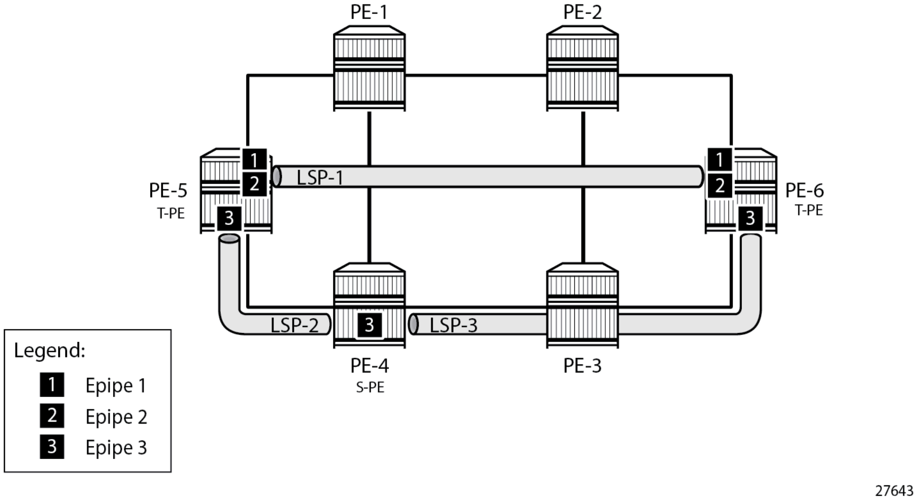
VCCV BFD sessions run end-to-end on a switched or single-hop PW, from T-PE to T-PE. They do not terminate on an intermediate S-PE; therefore, the TTL of the PW label on VCCV BFD packets is always set to 255, to ensure that the packets reach the far-end T-PE of a multi-segment PW.

BFD is only used for fault detection. While RFC 5885 provides a mode in which VCCV BFD can be used to signal PW status, this mode is only applicable for PWs that have no other status signaling mechanism in use. LDP status and static PW status signaling always take precedence over BFD-signaled PW status, and BFD-signaled PW status is not used on PWs that use LDP status or static PW status signaling mechanisms.

Configuration

Figure 299: Example topology shows the example topology with Epipes "Epipe-1" and "Epipe-2" using LSP "lsp-1" between PE-5 and PE-6 and Epipe "Epipe-3" using LSP "lsp-2" between PE-5 and S-PE PE-4 and LSP "lsp-3" between PE-4 and PE-6.

Figure 299: Example topology



The initial configuration includes:

- Cards, MDAs, and ports
- Router interfaces
- IS-IS as IGP on all interfaces (alternatively, OSPF can be used), with traffic engineering enabled
- MPLS paths and LSPs:
  - LSP "lsp-1" configured on PE-5 with primary path "path-5-1-2-6" and on PE-6 with primary path "path-6-2-1-5"
  - LSP "lsp-2" configured on PE-5 with primary path "path-5-4" and on PE-4 with primary path "path-4-5"
  - LSP "lsp-3" configured on PE-4 with primary path "path-4-3-6" and on PE-6 with primary path "path-6-3-4"

VCCV BFD configuration

Three steps are needed when configuring VCCV BFD:

1. Configure the BFD template
2. Apply the BFD template
3. Enable BFD

Step 1: configure BFD template

The **bfd-template** command provides the control packet timer values for the BFD.

The general command to define a BFD template is as follows:

```
configure
    router Base
```

```
        bfd
            bfd-template <name>
                transmit-interval <transmit-interval>
                receive-interval <receive-interval>
                echo-receive <echo-interval>
                multiplier <multiplier>
                type {cpm-np}
            exit
```

However, network processor BFD (cpm-np) is not supported for VCCV, and the minimum supported receive or transmit timer interval is 100 ms. An error is generated if a user tries to apply a BFD template with the **type cpm-np** command or any unsupported transmit or receive interval value. An error is also generated when the user attempts to commit changes to a BFD template that is already bound to a spoke-SDP.

Steps 2 and 3: apply BFD template and enable BFD

To apply and enable the BFD template to a spoke-SDP where LDP is used as the SDP signaling protocol for a service, the following command can be used, depending on the service:

```
configure
    service
        [epipe|cpipe|ipipe|vpls|ies|vprn] <service-id>
            spoke-sdp <sdp-binding-id>
                bfd
                    bfd-template <name>
                    bfd-enable
```

If BGP is used as the SDP signaling protocol, the following command is used:

```
configure
    service
        [epipe|vpls] <service-id>
            pw-template-binding <reference>
                bfd-template <name>
                bfd-enable
```

In this example, the following BFD templates are configured on PE-5 and PE-6:

```
# on PE-5, PE-6:
configure
    router Base
        bfd
            begin
            bfd-template "bfdt-1"
                transmit-interval 2000
                receive-interval 2000
                multiplier 5
            exit
            bfd-template "bfdt-2"
                transmit-interval 1000
                receive-interval 1000
            exit
            commit
        exit
```

These BFD templates are used in the Epipe services configured in the next section.

Service configuration

LDP VLL "Epipe-1"

The service "Epipe-1" is an LDP VLL running between PE-5 and PE-6, and uses manually configured SDP 56 on PE-5 and SDP 65 on PE-6, respectively, so the signaling is set to T-LDP. On PE-5, the spoke-SDP 56:1 has BFD template *bfdt-2* applied, and BFD is enabled. The configuration on PE-6 is similar.

```
# on PE-5
configure
    service
        sdp 56 mpls create
            signaling tldp    # default
            far-end 192.0.2.6
            lsp "lsp-1"
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        epipe 1 name "Epipe-1" customer 1 create
            sap 1/1/c4/1:1 create
                no shutdown
            exit
            spoke-sdp 56:1 create
                bfd
                    bfd-template "bfdt-2"
                    bfd-enable
                exit
                no shutdown
            exit
            no shutdown
        exit
```

Log 99 indicates the local discriminator value used for the VCCV BFD session, as follows:

```
88 2023/08/17 11:47:44.610 UTC MINOR: VRTR #2070 Base 127.0.0.1
"The vccv BFD session with Local Discriminator 1 on Svc 1 SdpBind 56:1 is up"
```

Configuring the spoke-SDP with a template with an invalid type (for example, type "cpm-np") leads to an error, as follows:

```
*A:PE-5>config>service>epipe>spoke-sdp>bfd# bfd-template "bfdt-cpm-np-100ms"
MINOR: SVCMGR #6260 Invalid bfd-template - incompatible bfd-template param
with VCCV: invalid type (oper)
```

Configuring the spoke-SDP with a template with an invalid transmit timer value also leads to an error, as follows:

```
*A:PE-5>config>service>epipe>spoke-sdp>bfd# bfd-template "bfdt-cpm-np-50ms"
MINOR: SVCMGR #6260 Invalid bfd-template - incompatible bfd-template param
with VCCV: invalid tx interval (oper)
```

BGP VPWS "Epipe-2"

The service "Epipe-2" is a BGP VPWS, also running between PE-5 and PE-6, using the manually configured SDPs 561 and 651 with the signaling set to BGP, as follows. Again, BFD template *bfdt-2* is used, but now the BFD template is referred to from the **pw-template-binding** context. See the BGP Virtual Private Wire Services chapter for more information.

```
# PE-5:
configure
    service
        sdp 561 mpls create
            signaling bgp
            far-end 192.0.2.6
            lsp "lsp-1"
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        pw-template 1 prefer-provisioned-sdp create
        exit
        epipe 2 name "Epipe-2" customer 1 create
            bgp
                route-distinguisher 65545:2
                route-target export target:65545:2 import target:65545:2
                pw-template-binding 1
                    bfd-template "bfdt-2"
                    bfd-enable
                exit
            exit
            bgp-vpws
                ve-name "PE-5"
                    ve-id 5
                exit
                remote-ve-name "PE-6"
                    ve-id 6
                exit
                no shutdown
            exit
            sap 1/1/c4/1:2 create
                no shutdown
            exit
            no shutdown
        exit
```

LDP VLL "Epipe-3" with switching node PE-4

The service "Epipe-3" is another LDP VLL running between PE-5 and PE-6, but switched at PE-4. It uses the manually configured SDPs 54 and 45 between PE-5 and PE-4, and SDPs 46 and 64 between PE-4 and PE-6. All these SDPs are using T-LDP for the signaling. On PE-5, the spoke-SDP 54:3 has BFD template "bfdt-1" applied, and **control-word** is active. This ensures that BFD packets get into the PW mapping to that spoke-SDP and that these packets are forwarded between the VC-switched spoke-SDPs at PE-4. The configuration on PE-6 is similar.

```
# on PE-5:
configure
    service
```

```
        sdp 54 mpls create
            signaling tldp        # default
            far-end 192.0.2.4
            lsp "lsp-2"
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        epipe 3 name "Epipe-3" customer 1 create
            sap 1/1/c4/1:3 create
                no shutdown
            exit
            spoke-sdp 54:3 create
                control-word
                bfd
                    bfd-template "bfdt-1"
                    bfd-enable
                exit
                no shutdown
            exit
            no shutdown
        exit
```

For PE-4 to switch traffic from one VC to another, the creation time keyword **vc-switching** is required, as follows:

```
# on PE-4:
configure
    service
        sdp 45 mpls create
            signaling tldp        # default
            far-end 192.0.2.5
            lsp "lsp-2"
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        sdp 46 mpls create
            signaling tldp        # default
            far-end 192.0.2.6
            lsp "lsp-3"
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        epipe 3 name "Epipe-3" customer 1 vc-switching create
            spoke-sdp 45:3 create
                no shutdown
            exit
            spoke-sdp 46:3 create
                no shutdown
            exit
            no shutdown
        exit
```

VCCV BFD verification

The following command shows that BFD template "bfdt-2" is applied to SDP 56:1 in the "Epipe-1" service and to SDP 561:4294967295 in the "Epipe-2" service:

```
*A:PE-5# show router bfd bfd-template "bfdt-2"

===============================================================================
BFD Template bfdt-2
===============================================================================
Template Name           : bfdt-2      Template Type             : auto
Transmit Timer          : 1000 msec    Receive Timer             : 1000 msec
Template Multiplier     : 3           Echo Receive Interval     : 100 msec

LSP-LDP  Association Count : 0
LSP-RSVP Association Count : 0
LSP-RSVP Template Association Count : 0
LSP-SR-TE Association Count : 0
LSP-SR-TE Template Association Count : 0
LSP-SR-TE Association Count : 0
LSP-SR-TE Template Association Count : 0
Static SR-Policy Association Count : 0
BGP SR-Policy Association Count : 0

Mpls-tp Association
None

-------------------------------------------------------------------------------
Service Associations
-------------------------------------------------------------------------------
SvcId               Sdp Bind                      BFD Enable     BFD Encap
-------------------------------------------------------------------------------
1                   56:1                          yes            ipv4
2                   561:4294967295                yes            ipv4
===============================================================================
```

The BFD configuration for SDP 56:1 on the "Epipe-1" service is listed in the detailed output for the SDP, as follows. The BFD template used is *bfdt-2*, BFD is enabled, and the BFD encapsulation used is IPv4. The peer VCCV CV bits indicate that the remote end supports LSP ping as well as BFD fault detection.

```
*A:PE-5# show service id 1 sdp 56:1 detail

===============================================================================
Service Destination Point (Sdp Id : 56:1) Details
===============================================================================
-------------------------------------------------------------------------------
 Sdp Id 56:1  -(192.0.2.6)
-------------------------------------------------------------------------------
Description     : (Not Specified)
SDP Id             : 56:1                     Type              : Spoke
Spoke Descr     : (Not Specified)
VC Type            : Ether                    VC Tag            : n/a
Admin Path MTU     : 0                        Oper Path MTU     : 8974
Delivery           : MPLS
Far End            : 192.0.2.6                Tunnel Far End    : n/a
Oper Tunnel Far End: 192.0.2.6
LSP Types          : RSVP
Hash Label         : Disabled                 Hash Lbl Sig Cap  : Disabled
Oper Hash Label    : Disabled
Entropy Label      : Disabled
```

```
Admin State        : Up                          Oper State        : Up
MinReqd SdpOperMTU : 1514
Adv Service MTU    : n/a
Acct. Pol          : None                        Collect Stats     : Disabled
Ingress Label      : 524284                      Egress Label      : 524284
Ingr Mac Fltr-Id   : n/a                         Egr Mac Fltr-Id   : n/a
Ingr IP Fltr-Id    : n/a                         Egr IP Fltr-Id    : n/a
Ingr IPv6 Fltr-Id  : n/a                         Egr IPv6 Fltr-Id  : n/a
Admin ControlWord  : Not Preferred               Oper ControlWord  : False
Admin BW(Kbps)     : 0                           Oper BW(Kbps)     : 0
BFD Template       : bfdt-2
BFD-Enabled        : yes                         BFD-Encap         : ipv4
BFD Fail Action    : none                        BFD Oper State    : connected
BFD WaitForUpTimer : 0 secs
BFD Time Remain    : 0 secs
Last Status Change : 08/17/2023 11:47:43         Signaling         : TLDP
Last Mgmt Change   : 08/17/2023 11:57:21
Endpoint           : N/A                         Precedence        : 4
ICB                : False
PW Status Sig      : Enabled
Force Vlan-Vc      : Disabled                    Force Qinq-Vc     : none
Class Fwding State : Down
Flags              : None
Local Pw Bits      : None
Peer Pw Bits       : None
Peer Fault Ip      : None
Peer Vccv CV Bits  : lspPing bfdFaultDet
Peer Vccv CC Bits  : mplsRouterAlertLabel

---snip---

-------------------------------------------------------------------------------
Control Channel Status
-------------------------------------------------------------------------------
PW Status          : disabled                    Refresh Timer     : <none>
Peer Status Expire : false
Request Timer      : <none>
Acknowledgement    : false

---snip---

-------------------------------------------------------------------------------
RSVP/Static LSPs
-------------------------------------------------------------------------------
Associated LSP List :
Lsp Name           : lsp-1
Admin State        : Up                          Oper State        : Up
Time Since Last Tr*: 00h15m42s

---snip---

Number of SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

The full set of VCCV BFD sessions running with the currently used parameters can be shown as follows:

```
*A:PE-5# show service vccv-bfd

===============================================================================
BFD Session
```

```
===============================================================================
Svc-Id                                            State      Tx Pkts    Rx Pkts
  Sdp-Id:Vc-Id                                  Multipl     Tx Intvl   Rx Intvl
  Protocols                                        Type     LAG Port     LAG ID
                                                                       LAG name
-------------------------------------------------------------------------------
1                                                    Up          150        150
  56:1                                                3         1000       1000
  vccv                                          central          N/A        N/A
  127.0.0.2
2                                                    Up          718        718
  561:4294967295                                      3         1000       1000
  vccv                                          central          N/A        N/A
  127.0.0.2
3                                                    Up          251        251
  54:3                                                5         2000       2000
  vccv                                          central          N/A        N/A
  127.0.0.2
-------------------------------------------------------------------------------
No. of System BFD sessions: 3
===============================================================================
```

The VCCV BFD sessions for a single service can be shown as follows:

```
*A:PE-5# show service id 3 vccv-bfd session

===============================================================================
BFD Session
===============================================================================
Svc-Id                                            State      Tx Pkts    Rx Pkts
  Sdp-Id:Vc-Id                                  Multipl     Tx Intvl   Rx Intvl
  Protocols                                        Type     LAG Port     LAG ID
                                                                       LAG name
-------------------------------------------------------------------------------
3                                                    Up          265        265
  54:3                                                5         2000       2000
  vccv                                          central          N/A        N/A
  127.0.0.2
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
```

Similar output can be obtained on PE-6.

Disconnecting the link between PE-1 and PE-2 affects the traffic taking the upper path; the VCCV BFD sessions for the services "Epipe-1" and "Epipe-2" go down, and so do the SDPs and the services. This is reflected in log 99, as follows:

```
142 2023/08/17 12:00:42.764 UTC WARNING: MPLS #2012 Base VR 1:
"LSP path lsp-1::path-5-1-2-6 is operationally disabled ('shutdown') because
resvTear"

143 2023/08/17 12:00:42.764 UTC WARNING: MPLS #2010 Base VR 1:
"LSP lsp-1 is operationally disabled ('shutdown') because noPathIsOperational"

144 2023/08/17 12:00:42.765 UTC MINOR: SVCMGR #2303 Base
"Status of SDP 56 changed to admin=up oper=down"

145 2023/08/17 12:00:42.765 UTC MINOR: SVCMGR #2303 Base
"Status of SDP 561 changed to admin=up oper=down"

146 2023/08/17 12:00:42.766 UTC MINOR: SVCMGR #2326 Base
```

```
"Status of SDP Bind 56:1 in service 1 (customer 1) local PW status bits changed
to psnIngressFault psnEgressFault "

147 2023/08/17 12:00:42.767 UTC MAJOR: SVCMGR #2316 Base
"Processing of a SDP state change event is finished and the status of all affected
 SDP Bindings on SDP 561 has been updated."

148 2023/08/17 12:00:42.767 UTC MAJOR: SVCMGR #2316 Base
"Processing of a SDP state change event is finished and the status of all affected
 SDP Bindings on SDP 56 has been updated."

149 2023/08/17 12:00:45.019 UTC MINOR: VRTR #2069 Base 127.0.0.1
"The vccv BFD session with Local Discriminator 9 on Svc 1 SdpBind 56:1 is down
 due to noHeartBeat "

150 2023/08/17 12:00:45.600 UTC MINOR: VRTR #2069 Base 127.0.0.1
"The vccv BFD session with Local Discriminator 2 on Svc 2 SdpBind 561:4294967295
 is down due to noHeartBeat "

151 2023/08/17 12:00:49.802 UTC MINOR: SVCMGR #2313 Base
"Status of SDP Bind 56:1 in service 1 (customer 1) peer PW status bits changed
to psnIngressFault psnEgressFault "
```

This status of the VCCV BFD sessions then is as follows:

```
*A:PE-5# show service vccv-bfd

===============================================================================
BFD Session
===============================================================================
Svc-Id                                            State      Tx Pkts    Rx Pkts
  Sdp-Id:Vc-Id                                  Multipl     Tx Intvl   Rx Intvl
  Protocols                                        Type     LAG Port     LAG ID
                                                                       LAG name
-------------------------------------------------------------------------------
1                                                  Down          230        230
  56:1                                                3         1000       1000
  vccv                                          central          N/A        N/A
  127.0.0.2
2                                                  Down          799        799
  561:4294967295                                      3         1000       1000
  vccv                                          central          N/A        N/A
  127.0.0.2
3                                                    Up          320        319
  54:3                                                5         2000       2000
  vccv                                          central          N/A        N/A
  127.0.0.2
-------------------------------------------------------------------------------
No. of System BFD sessions: 3
===============================================================================
```

Consequently, the "Epipe-1" and "Epipe-2" services are operationally down, as follows:

```
*A:PE-5# show service service-using epipe

===============================================================================
Services [epipe]
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
1            Epipe     Up   Down 1          Epipe-1
2            Epipe     Up   Down 1          Epipe-2
3            Epipe     Up   Up   1          Epipe-3
```

```
-------------------------------------------------------------------------------
Matching Services : 3
-------------------------------------------------------------------------------
===============================================================================
```

Conclusion

VCCV BFD can monitor specific high-value services, where detecting forwarding failures (and potentially recovering from them) in the minimal amount of time is critical. VCCV BFD complements other on-demand tools such as VCCV ping and VCCV trace by providing proactive detection of faults. VCCV ping and VCCV trace can later be used to localize and diagnose the root cause of the fault.

Virtual Ethernet Segments

This chapter provides information about Virtual Ethernet Segments.

Topics in this chapter include:

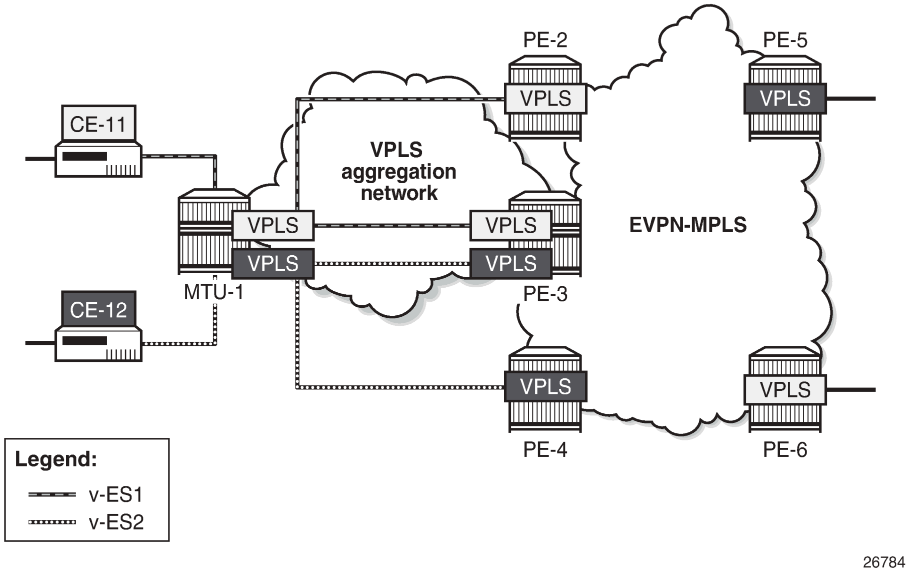
Applicability

This chapter was initially written based on SR OS Release 15.0.R3, but the CLI in the current edition is based on SR OS Release 21.2.R2. Virtual Ethernet segments are supported in SR OSRelease 15.0.R1, and later.

Overview

RFC 7432 describes the use and procedures for Ethernet segments (ESs) that can be associated with physical Ethernet ports and LAGs. The SR OS implementation also allows an ES to be associated with SDPs. ESs meet the redundancy requirements of directly connected CEs. However, ESs will not work when an aggregation network exists between CEs and ES PEs, which requires different ESs to be defined for the port, LAG, or SDP. *Draft-ietf-bess-evpn-virtual-eth-segment* describes how virtual ESs (vESs) can be defined with an Attachment Circuit (AC) level granularity. Figure 300: vESs for PWs shows an example where vES definition at the pseudowire (PW) granularity level is required:

Figure 300: vESs for PWs



When a Layer 2 aggregation network is used to get access to EVPN, the association of ACs that belong to the same ES and physical ports or SDPs can be arbitrary. For example, the SDP between MTU-1 and PE-3 (Figure 300: vESs for PWs) cannot be associated with only one ES, because it is being used by two different CEs that require different ESs. The association must be at spoke-SDP level. The RFC 7432 port/lag-based ES definition is not sufficient, so vESs need to be defined. Virtual ESs can be configured with up to eight ranges of one or more:

- VC-IDs (spoke-SDPs)
- Q-tags (dot1q)
- S-tags (qinq)
- C-tags for a fixed S-tag (qinq)

Mesh-SDPs are not allowed for an SDP used by a vES.

Virtual ESs are configured as Ethernet segments with the creation-time keyword **virtual**:

```
*A:PE-2>config>service>system>bgp-evpn# ethernet-segment ?
  - ethernet-segment <name> [create] [virtual]
  - no ethernet-segment <name>

 <name>                 : [32 chars max]
 <virtual>              : keyword

      dot1q             + Configure dot1q port or lag information
 [no] es-activation-* - Configure ethernet segment activation timer
 [no] es-orig-ip        - Configure ES route's originating IP address.
 [no] esi               - Configure ethernet segment identifier
 [no] lag               - Configure lag for service BGP EVPN ethernet segment
 [no] multi-homing      - Configure multi-homing for service BGP EVPN ethernet segment
 [no] network-interc* - Configure network interconnect vxlan information
 [no] oper-group        - Configure operational-group for the ethernet-segment
 [no] port              - Configure port for service BGP EVPN ethernet segment
 [no] pw-port           - Configure pw-port for service BGP EVPN ethernet segment
      qinq              + Configure qinq port or lag information
 [no] route-next-hop  - Configure next hop IP for ES and AD per-ES routes.
 [no] sdp             - Configure sdp for service BGP EVPN ethernet segment
      service-carving + Configure service carving mode for BGP EVPN ethernet segment
      service-id      + Configure service id vxlan information under ethernet segment
 [no] shutdown        - Enable/disable administrative state of the ethernet segment
 [no] source-bmac-lsb - Configure source  BMAC address LSB information
 [no] vc-id-range     - Configure VC ID range
```

Virtual ES "vESI-23_600" is associated with LAG 1 and one service-delimiting VLAN range is defined for the S-tag, as follows:

```
# on PE-2, PE-3:
configure
    service
        system
            bgp-evpn
                ethernet-segment "vESI-23_600" virtual create
                    esi 01:00:00:00:00:23:06:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode manual
                        manual
                            evi 2
                        exit
                    exit
                    multi-homing all-active
                    lag 1
                    qinq
                        s-tag-range 600 to 602
                    exit
                    no shutdown
                exit
```

The configured ES will match all the SAPs for which the top (outer) service-delimiting tag is within the 600 to 602 range.

When the ES is created as virtual, a port, LAG, or SDP needs to be created before any VLAN or VC-ID can be associated.

- For VC-ID, only spoke-SDPs are allowed, no mesh-SDPs. Manual spoke-SDP VC-IDs and BGP-AD VC-IDs can be included in the range.
- For dot1q, only those SAPs that match the service-delimiting VLAN range will be associated with the vES
- For qinq, the following two commands can be configured, with a mutually exclusive S-tag:
  - **s-tag-range** <*qtag1*> **to** <*qtag1*> - associates all qinq SAPs with outer tag between the configured qtags.
  - **s-tag** <*qtag1*> **c-tag-range** <*qtag2*> **to** <*qtag2*> - associates all qinq SAPs with outer qtag1 and inner qtag between the configured qtag2 values to the vES

  A mutually exclusive S-tag means that a value for the S-tag can be configured in either of the two commands, but not in both.

Table 13: Supported examples for Q-tag values between 1 and 4094 shows the supported examples for qtag values between 1 and 4094; Table 14: Supported examples for Q-tag values 0, *, and null shows the supported examples for qtag values 0, *, and null:

Table 13: Supported examples for Q-tag values between 1 and 4094

vES configuration for port 1/1/1	SAP association
dot1q qtag-range 100	1/1/1:100
dot1q qtag-range 100 to 102	1/1/1:100, 1/1/1:101, 1/1/1:102
qinq s-tag 100 c-tag-range 200	1/1/1:100.200
qinq s-tag 100 c-tag-range 200 to 202	1/1/1:100.200, 1/1/1:100.201, 1/1/1:100.202
qinq s-tag-range 100	All SAPs 1/1/1:100.x (x being 1 to 4094, 0, or *)
qinq s-tag-range 100 to 102	All SAPs 1/1/1:100.x, 1/1/1:101.x, 1/1/1:102.x (x being 1 to 4094, 0, or *)

*Table 14: Supported examples for Q-tag values 0, *, and null*

vES configuration for port 1/1/1	SAP association
dot1q qtag-range 0	1/1/1:0
dot1q qtag-range *	1/1/1:*
qinq s-tag 0 c-tag-range *	1/1/1:0.*
qinq s-tag * c-tag-range *	1/1/1:*.*
qinq s-tag * c-tag-range null	1/1/1:*.null

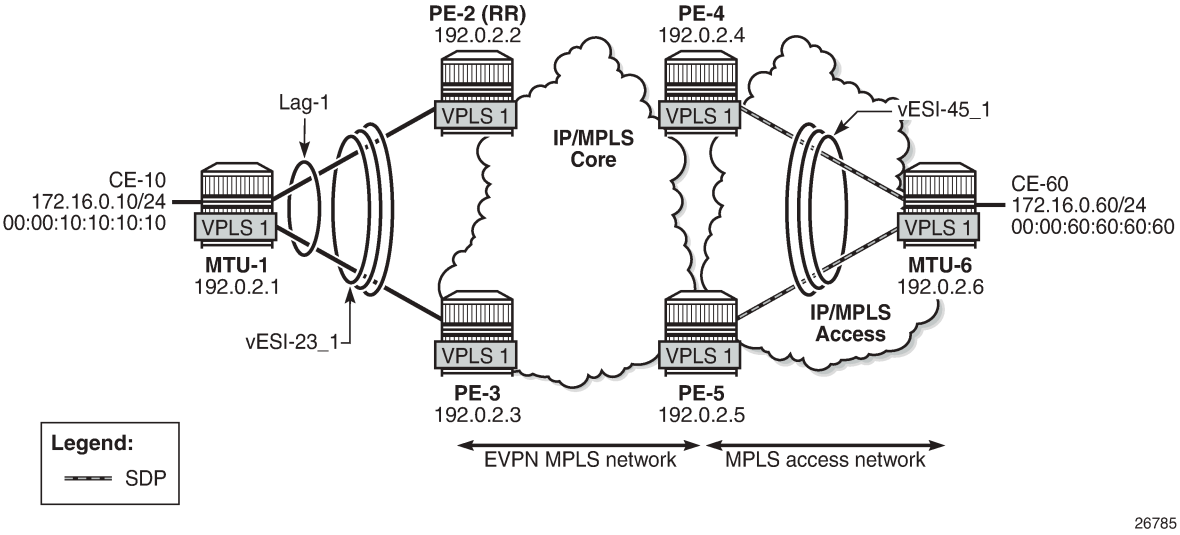
Considerations:

- The ranges can be modified on the fly: qtag-range, s-tag/c-tag-range, vc-id-range.
- For port-based vESs, PXC sub-ports are supported. For more information about PXC, see chapter *Port Cross-Connect (PXC)* in the *7450 ESS, 7750 SR, and 7950 XRS Interface Configuration Advanced Configuration Guide for Classic CLI*.
- Virtual ESs are supported in EVPN-MPLS, PBB-EVPN, and EVPN-VPWS
- Virtual ESs are supported in single-active and all-active EVPN multi-homing
  - Two all-active vESs must use different ES-BMAC addresses, even if they are defined in the same LAG.
- Virtual ESs implement CMAC flush procedures described in RFC 7623. Optionally, ISID-based CMAC-flush can be used where the single-active vES does not use ES-BMAC allocation. See chapter PBB-EVPN ISID-based CMAC Flush.
- Connection-profile-vlan SAPs (CP-SAPs) cannot be associated with a vES and cannot be configured on ports where vESs are defined. For more information about CP-SAPs, see chapter VLAN Range SAPs for VPLS and Epipe Services.

Configuration

Figure 301: Example topology shows the example topology with four core PEs in an EVPN-MPLS network and two MTUs. VPLS 1 is configured in all the nodes. EVPN is configured on the core PEs, not on the MTUs. LAG 1 is configured on MTU-1, PE-2, and PE-3 and associated with an all-active vES "ESI-23_1" on PE-2 and PE-3. A single-active vES "ESI-45_1" is configured on PE-4 and PE-5, associated with SDPs.

Figure 301: Example topology



The configuration is similar to the one in chapter EVPN for MPLS Tunnels, where the parameters are described in detail.

The initial configuration on the nodes includes the following:

- Cards, MDAs, ports
- Router interfaces
- IS-IS (alternatively, OSPF can be configured)
- LDP in the IP/MPLS core and IP/MPLS access network

LAG 1 is configured with qinq encapsulation. The LAG configuration on MTU-1 is as follows:

```
# on MTU-1:
configure
    lag 1 name "lag-1"
        mode access
        encap-type qinq
        port 1/1/1
        port 1/1/2
        lacp active administrative-key 32768
        no shutdown
```

BGP is configured on all PEs for address family EVPN. PE-2 is the Route Reflector (RR) and is configured as follows.

```
# on RR PE-2:
configure
    router Base
        autonomous-system 64500
        bgp
            vpn-apply-import
            vpn-apply-export
            enable-peer-tracking
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family evpn
                cluster 1.1.1.1
                peer-as 64500
                neighbor 192.0.2.3
                exit
                neighbor 192.0.2.4
                exit
                neighbor 192.0.2.5
                exit
            exit
```

VPLS 1 is configured on all nodes. On the PEs, BGP-EVPN is enabled for MPLS. The following is configured on PE-2:

```
# on PE-2:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap lag-1:1.1 create
                no shutdown
            exit
            no shutdown
        exit
```

The configuration on the other PEs is similar, but on PE-4 and PE-5, a spoke-SDP is configured instead of a SAP. The service configuration on PE-4 is as follows:

```
# on PE-4:
configure
    service
```

```
        sdp 46 mpls create
            far-end 192.0.2.6
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        vpls 1 name "VPLS 1" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    ingress-replication-bum-label
                    ecmp 2
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            spoke-sdp 46:1 create
                no shutdown
            exit
            no shutdown
        exit
```

Virtual ESs must be created with the **virtual** keyword; if not, the following error is raised after an attempt to define a range:

```
*A:PE-2>config>service>system>bgp-evpn>eth-seg>qinq# s-tag-range 1
MINOR: SVCMGR #8070 Cannot create range - ethernet-segment is not virtual
```

On PE-2 and PE-3, the two following two all-active multi-homing vESs are created, each with a unique ESI:

```
# on PE-2, PE-3:
configure
    service
        system
            bgp-evpn
                ethernet-segment "vESI-23_1" virtual create
                    esi 01:00:00:00:00:23:01:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing all-active
                    lag 1
                    qinq
                        s-tag-range 1
                        s-tag-range 500 to 501
                        s-tag 495 c-tag-range 100 to 102
                    exit
                    no shutdown
                exit
                ethernet-segment "vESI-23_600" virtual create
                    esi 01:00:00:00:00:23:06:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode manual
                        manual
                            evi 2
                        exit
                    exit
                    multi-homing all-active
                    lag 1
                    qinq
                        s-tag-range 600 to 602
                    exit
                    no shutdown
                exit
```

When attempting to configure another vES with the ESI of an existing ES/vES, the following error is raised:

```
*A:PE-2>config>service>system>bgp-evpn# ethernet-segment "vESI-23_610" virtual create
*A:PE-2>config>service>system>bgp-evpn>eth-seg# esi 01:00:00:00:00:23:06:00:00:01
MINOR: SVCMGR #8047 Ethernet segment id is not valid - ESI already in use by another ethernet
 segment
```

Multiple vESs can be defined on the same LAG. However, the ranges should not overlap. The following error is raised after attempting to configure an additional range in vES "ESI-23_600" that uses S-tag 600 in combination with a range of C-tags. S-tag 600 is already included in the first range: **s-tag-range 600 to 602**. The error message points out that this range is of a different type: the existing range defines only S-tags, whereas the new range defines a range of C-tags for S-tag 600.

```
*A:PE-2>config>service>system>bgp-evpn>eth-seg>qinq# s-tag 600 c-tag-range 100 to 111
MINOR: SVCMGR #8070 Cannot create range - range overlaps with existing range of a different
 type
```

When attempting to define **s-tag-range 1** in "vESI-23_2", when S-tag 1 is already defined in "vESI-23_1", the following error is raised:

```
*A:PE-2>config>service>system>bgp-evpn>eth-seg>qinq# s-tag-range 1
MINOR: SVCMGR #8070 Cannot create range - range overlaps with existing range in ethernet-
segment vESI-23_1
```

On PE-4, the following single-active multi-homing vESs are configured. The configuration on PE-5 contains a different SDP.

```
# on PE-4:
configure
    service
        system
            bgp-evpn
                ethernet-segment "vESI-45_1" virtual create
                    esi 01:00:00:00:00:45:01:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    sdp 46
                    vc-id-range 1
                    vc-id-range 500 to 501
                    no shutdown
                exit
                ethernet-segment "vESI-45_2" virtual create
```

```
                    esi 01:00:00:00:00:45:02:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode manual
                        manual
                            evi 2
                        exit
                    exit
                    multi-homing single-active
                    sdp 46
                    vc-id-range 2
                    no shutdown
                exit
```

The configured ESs and vESs can be retrieved as follows:

```
*A:PE-2# show service system bgp-evpn ethernet-segment

===============================================================================
Service Ethernet Segment
===============================================================================
Name                             ESI                           Admin     Oper
-------------------------------------------------------------------------------
vESI-23_1                        01:00:00:00:00:23:01:00:00:01 Enabled   Up
vESI-23_600                      01:00:00:00:00:23:06:00:00:01 Enabled   Up
-------------------------------------------------------------------------------
Entries found: 2
===============================================================================
```

The following information for the first entry in the list shows that it is a virtual ES.

```
*A:PE-2# show service system bgp-evpn ethernet-segment name "vESI-23_1"

===============================================================================
Service Ethernet Segment
===============================================================================
Name                     : vESI-23_1
Eth Seg Type             : Virtual
Admin State              : Enabled            Oper State         : Up
ESI                      : 01:00:00:00:00:23:01:00:00:01
Multi-homing             : allActive          Oper Multi-homing  : allActive
ES SHG Label             : 524280
Source BMAC LSB          : <none>
Lag Id                   : 1
ES Activation Timer      : 3 secs
Oper Group               : (Not Specified)
Svc Carving              : auto               Oper Svc Carving   : auto
Cfg Range Type           : primary
===============================================================================
```

Virtual ES "vESI-23_1" on PE-2 has the following S-tag ranges and S/C-tag ranges:

```
*A:PE-2# show service system bgp-evpn ethernet-segment name "vESI-23_1" virtual-ranges

===============================================================================
Q-Tag Ranges
===============================================================================
Q-Tag Start         Q-Tag End           Last Changed
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
No entries found
===============================================================================
===============================================================================
VC-Id Ranges
===============================================================================
VC-Id Start         VC-Id End           Last Changed
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
No entries found
===============================================================================

===============================================================================
S-Tag Ranges
===============================================================================
S-Tag Start         S-Tag End           Last Changed
-------------------------------------------------------------------------------
1                   1                   04/19/2021 12:21:18
500                 501                 04/19/2021 12:21:18
-------------------------------------------------------------------------------
Number of Entries: 2
===============================================================================

===============================================================================
S-Tag C-Tag Ranges
===============================================================================
S-Tag Start         C-Tag Start         C-Tag End       Last Changed
-------------------------------------------------------------------------------
495                 100                 102             04/19/2021 12:21:18
-------------------------------------------------------------------------------
Number of Entries: 1
===============================================================================

===============================================================================
Vxlan Instance Service Ranges
===============================================================================
Svc Range Start          Svc Range End            Last Changed
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
No entries found
===============================================================================
```

The ranges in the vES can be modified while the vES is operationally up, for example, an S-tag range can be added as follows:

```
# on PE-2:
configure
    service
        system
            bgp-evpn
                ethernet-segment "vESI-23_1"
                    qinq
                        s-tag-range 10
```

The S-tag ranges can be verified with the following command. Compared with the preceding output, the S-tag 10 has been added:

```
*A:PE-2# show service system bgp-evpn ethernet-segment name "vESI-23_1" virtual-ranges | match
 S-Tag post-lines 9
S-Tag Ranges
===============================================================================
S-Tag Start         S-Tag End           Last Changed
-------------------------------------------------------------------------------
1                   1                   04/19/2021 12:21:18
```

```
10                  10                  04/19/2021 12:27:11
500                 501                 04/19/2021 12:21:18
-------------------------------------------------------------------------------
Number of Entries: 3
===============================================================================

===============================================================================
S-Tag C-Tag Ranges
===============================================================================
S-Tag Start         C-Tag Start         C-Tag End       Last Changed
-------------------------------------------------------------------------------
495                 100                 102             04/19/2021 12:21:18
-------------------------------------------------------------------------------
Number of Entries: 1
===============================================================================

===============================================================================
Vxlan Instance Service Ranges
===============================================================================
```

On PE-4, the same **show** command shows the range of VC-IDs, as follows:

```
*A:PE-4# show service system bgp-evpn ethernet-segment name "vESI-45_1" virtual-ranges

===============================================================================
Q-Tag Ranges
===============================================================================
Q-Tag Start         Q-Tag End           Last Changed
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
No entries found
===============================================================================

===============================================================================
VC-Id Ranges
===============================================================================
VC-Id Start         VC-Id End           Last Changed
-------------------------------------------------------------------------------
1                   1                   04/19/2021 12:24:50
500                 501                 04/19/2021 12:24:50
-------------------------------------------------------------------------------
Number of Entries: 2
===============================================================================

===============================================================================
S-Tag Ranges
===============================================================================
S-Tag Start         S-Tag End           Last Changed
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
No entries found
===============================================================================

===============================================================================
S-Tag C-Tag Ranges
===============================================================================
S-Tag Start         C-Tag Start         C-Tag End       Last Changed
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
No entries found
===============================================================================

===============================================================================
Vxlan Instance Service Ranges
===============================================================================
Svc Range Start          Svc Range End            Last Changed
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
No entries found
===============================================================================
```

Connection-profile-vlan SAPs (CP-SAPs) cannot be associated with a vES and cannot be configured on ports where vESs are defined. CP-SAP 10 is created on PE-3, as follows:

```
# on PE-3:
configure
    connection-profile-vlan 10 create
        vlan-range 5 to 100
        vlan-range 495
    exit
```

The following vES is configured on PE-3:

```
# on PE-3:
configure
    service
        system
            bgp-evpn
                ethernet-segment "vESI-23_10" virtual create
                    esi 01:00:00:00:00:23:10:00:00:01
                    es-activation-timer 3
                    service-carving
                        mode auto
                    exit
                    multi-homing single-active
                    port 1/2/3
                    qinq
                        s-tag-range 100
                    exit
                    no shutdown
                exit
```

This vES can only be configured when no CP-SAPs are defined on port 1/2/3. The following error message is raised when a CP-SAP is configured on port 1/2/3 already and the vES is configured afterward:

```
*A:PE-3>config>service>system>bgp-evpn>eth-seg# port 1/2/3
MINOR: SVCMGR #8048 Ethernet segment access port/lag/sdp/vxlan-instance/pw-port is not valid -
 not allowed when connection profile saps configured on port/lag
```

When attempting to configure CP-SAP 1/2/3:cp-10 in VPLS 1 with port 1/2/3 associated with a vES, the following error message is raised.

```
*A:PE-3>config>service>vpls# sap 1/2/3:100.cp-10 create
MINOR: SVCMGR #6044 Cannot create sap - sap type not allowed when port is associated with
 virtual ethernet-segment
```

Conclusion

Regular ESs and vESs can be associated with ports, LAGs, and SDPs; in case of vES, ranges of Q-tags, S-tags, C-tags, or VC-IDs can be defined. The granularity for vES is per AC. Multiple vESs with different ESIs can be defined on the same port, LAG, or SDP.

VLAN Range SAPs for VPLS and Epipe Services

This chapter provides information about VLAN range SAPs for VPLS and Epipe services.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS Release 14.0.R6, but the CLI in the current edition is based on SR OS Release 21.2.R1. Connection-Profile VLAN SAPs (CP SAPs) are supported in SR OS Release 14.0.R1, and later.

Overview

Backhaul services through metro Ethernet networks require bundled interface support. In SR OS terminology, bundling refers to Connection-Profile VLAN SAPs (CP SAPs)—special SAPs that capture the traffic of a range of CE VLAN IDs (VIDs) entering an Ethernet port. CP SAPs are fully compatible with Metro Ethernet Forum (MEF) 10.3 bundling service attributes and RFC 7432 EVPN VLAN bundle service interfaces. CP SAPs are supported in Layer 2 services only, and can be configured together with other SAPs and/or SDP-bindings.

For frames with an ingress VID contained in the range configured in the SAP's CP, the behavior is similar to default SAPs, such as 1/1/1:*, where "*" spans the entire VID range from 0 to 4095 and serves as a wildcard. However, unlike a default SAP, a CP SAP cannot co-exist with a VLAN SAP that is in the same range and on the same port or LAG. For example, 1/1/1:* and 1/1/1:100 can co-exist whereas 1/1/2:cp-1 (where cp-1 corresponds to the VLAN range from 1 to 200) and 1/1/2:100 cannot co-exist.

The VLAN manipulation between VLAN SAPs, default SAPs, and CP SAPs is compared in Table 15: VLAN manipulation in SAPs .

Table 15: VLAN manipulation in SAPs

	VLAN SAP	Default SAP	CP SAP
Service-delimiting VLAN	Yes For example: VLAN 100 in 1/1/1:100	No	No
Push/pop VLAN tags in egress/ ingress frames	Yes	No	No

	VLAN SAP	Default SAP	CP SAP
VLAN translation	Yes	No	No

Figure 302: Customer VID is popped and pushed by VLAN SAPs - VLAN translation shows how dot1q VLAN SAPs pop the customer VLAN tag in ingress frames and push the VLAN tag in egress frames. Therefore, frames are untagged between PE-1 and PE-2. VLAN translation is possible when the VIDs in the VLAN tags that are popped or pushed at the SAPs are different at ingress and egress, as follows.

Figure 302: Customer VID is popped and pushed by VLAN SAPs - VLAN translation

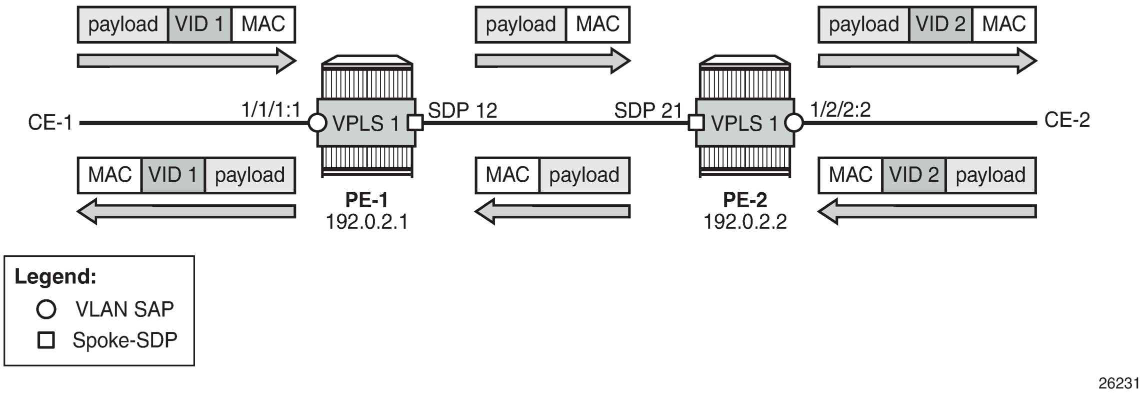


Figure 303: Customer VID is preserved between dot1q CP SAPs - no VLAN translation shows that dot1q CP SAPs do not pop or push the CE VID. Frames keep the same tag end-to-end; therefore, VLAN translation is not possible.

Figure 303: Customer VID is preserved between dot1q CP SAPs - no VLAN translation

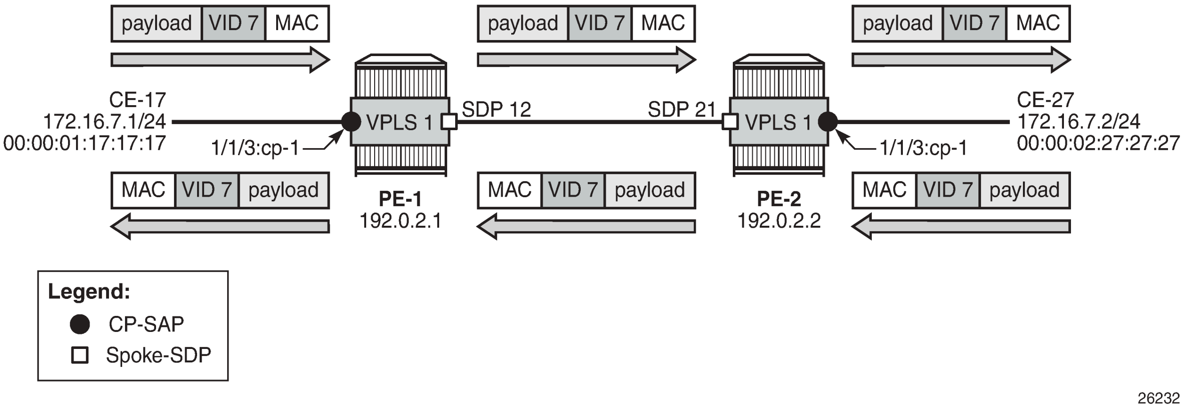
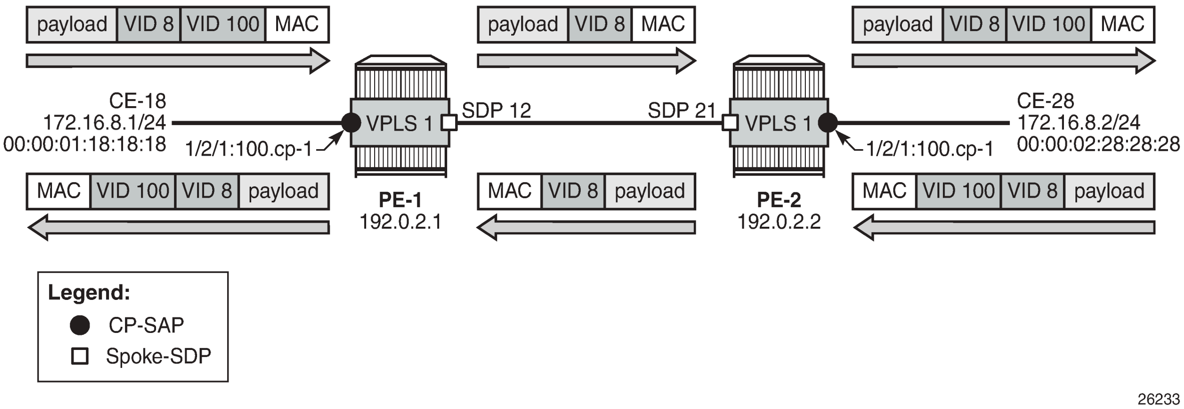


Figure 304: Customer VID is preserved between QinQ CP SAPs - no VLAN translation shows that QinQ CP SAPs only pop or push the service delimiting VID (VID 100), but not the customer VID in the CP range, as follows:

Figure 304: Customer VID is preserved between QinQ CP SAPs - no VLAN translation



VID 100 is service delimiting and can be different in both SAPs, but the customer VID in the VLAN range of the CP is not.

Connection profile VLAN

Note:
The **connection-profile-vlan** context is different from the connection-profile used for ATM connectivity.

CP SAPs refer to connection profiles (connection-profile-vlan) that can contain up to 32 ranges of customer VIDs. Connection profiles are configured with the following command:

```
*A:PE-1>config# connection-profile-vlan ?
  - connection-profile-vlan <conn-prof-id> [create]
  - no connection-profile-vlan <conn-prof-id>

 <conn-prof-id>       : [1..8000]


 [no] description     - Configure a connection profile VLAN description
 [no] vlan-range      - Configure a connection profile vlan range
```

VLAN ranges in a CP contain one or more consecutive VIDs, as follows:

```
*A:PE-1>config>connprofvlan$ vlan-range ?
  - no vlan-range <from>
  - vlan-range <from> [to <to>]

 <from>               : [1..4094]
 <to>                 : [1..4094]
```

Following is an example of a CP configuration containing three non-overlapping VLAN ranges:

```
# on PE-1:
configure
    connection-profile-vlan 10 create
        vlan-range 5 to 100
        vlan-range 150 to 300
        vlan-range 350
    exit
```

Overlapping ranges are not allowed within the same CP. An error message indicates that the VLAN range from 7 to 9 overlaps with the previously configured ranges (in this example, the VLAN range from 5 to 100), as follows:

```
*A:PE-1>config>connprofvlan# vlan-range 7 to 9
MINOR: SVCMGR #9012 Overlapping range
```

New non-overlapping VLAN ranges can be added to the CP defined in an existing and operationally up SAP. The CP's VLAN ranges can also be removed on the fly. When a user wants to extend a VLAN range, for example, VLAN range 350 becoming a range from 350 to 400, the existing VLAN range is not overwritten and a message is raised indicating that the VLAN ranges overlap, as follows:

```
*A:PE-1>config>connprofvlan# vlan-range 350 to 400
MINOR: SVCMGR #9012 Overlapping range
```

The existing VLAN range of 350 can be preserved when the CP SAP is operational and a new VLAN range from 351 to 400 can be added, as follows:

```
*A:PE-1>config>connprofvlan# vlan-range 351 to 400
```

The following example shows four VLAN ranges in CP 10, with a timestamp of the last change for each VLAN range:

```
*A:PE-1# show connection-profile-vlan 10

===============================================================================
Connection Profile 10 Information
===============================================================================
Description : (Not Specified)
Last Change : 03/30/2021 07:35:20


===============================================================================
Connection Profile Vlan Eth Information
===============================================================================
Range Start         Range End                     Last Change
-------------------------------------------------------------------------------
5                   100                           03/30/2021 07:58:07
150                 300                           03/30/2021 07:58:07
350                 350                           03/30/2021 07:58:07
351                 400                           03/30/2021 07:59:58
===============================================================================
===============================================================================
```

If a VLAN tag combination matches different SAPs, the highest priority SAP will be picked regardless of the operational status. For completeness, the following two tables show the SAP lookup matching order for dot1q and QinQ ports.

Table 16: SAP lookup order for dot1q ports

Incoming frame qtag VID value	SAP lookup precedence order (:0 and :* are mutually exclusive on the same port)			
	:X	**:CP**	**:0**	**:***
X (belongs to the CP range)	1st	1st		2nd
0			1st	1st
<untagged>			1st	1st

Table 17: SAP lookup order for QinQ ports

Incoming frame qtag1.qtag2	System/port settings = new-qinq-untagged-sap SAP lookup precedence order (assumption: X and Y are defined in CP ranges)							
	:X.Y	**:X.0**	**:X.CP**	**:CP.***	**:X.***	**:0.***	**:*.null**	**:*.***
X.Y	1st		1st	2nd	2nd			3rd
X.0		1st		2nd	2nd			3rd
0.Y						1st		2nd
0.0						1st		2nd
X		1st		2nd	2nd		3rd	4th
0						1st	2nd	3rd
<untagged>						1st	2nd	3rd

For example, ingress frames with VIDs 100.20 are classified as part of CP SAP 1/2/1:100.cp-10, not of CP SAP 1/2/3:cp-10.*. Only when SAP 1/2/1:100.cp-10 is removed from the configuration, frames with VIDs 100.20 will go to SAP 1/2/3:cp-10.*.

Assign CP SAPs to VPLS or Epipe services

Like ordinary SAPs, CP SAPs can be assigned to VPLS or Epipe services, as follows. The VPLS and Epipe can be EVPN services or not. In the following example, VPLS 1 has BGP-EVPN enabled, whereas Epipe 2 does not:

```
# on PE-1:
configure
    service
        sdp 12 mpls create
            far-end 192.0.2.2
            ldp
            keep-alive
```

```
                shutdown
            exit
            no shutdown
        exit
        vpls 1 name "VPLS 1" customer 1 create
            bgp
            exit
            bgp-evpn
                evi 1
                mpls bgp 1
                    ingress-replication-bum-label
                    auto-bind-tunnel
                        resolution any
                    exit
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/1/3:cp-10 create
                no shutdown
            exit
            sap 1/2/1:1.11 create
                no shutdown
            exit
            sap 1/2/1:100.cp-10 create
                no shutdown
            exit
            sap 1/2/3:cp-10.* create
                no shutdown
            exit
            no shutdown
        exit
        epipe 2 name "Epipe 2" customer 1 create
            sap 1/2/1:200.cp-10 create
                no shutdown
            exit
            spoke-sdp 12:2 create
                no shutdown
            exit
            no shutdown
        exit
```

CP SAPs are configured in the same way as VLAN SAPs and default SAPs, with the following restrictions:

- A CP can be defined for inner or outer tags as shown in the preceding configuration, but not both at the same time, as follows:

```
*A:PE-1>config>service>vpls# sap 1/2/1:cp-3.cp-10 create
MINOR: CLI SAP-id has an invalid port number or encapsulation value.
```

- If a CP is defined for the outer VID, the inner VID cannot be a specific VID. The inner VID can only be a "*" (where the inner tag can have any value) or a "0" (where the inner tag can be 0 or null), as follows:

```
*A:PE-1>config>service>vpls# sap 1/2/1:cp-3.4 create
MINOR: CLI SAP-id has an invalid port number or encapsulation value.
```

```
*A:PE-1>config>service>vpls# sap 1/2/1:cp-10.* create
```

```
*A:PE-1>config>service>vpls>sap# exit
```

```
*A:PE-1>config>service>vpls# sap 1/2/1:cp-3.0 create
*A:PE-1>config>service>vpls>sap$ exit
```

- No VLAN SAP can be added on a port in dot1q (or a combination of port and service-delimiting VLAN in case of QinQ) when the VLAN is included in the VLAN range in a CP SAP on the same port. One of the VLAN ranges in CP 10 contains all VIDs from 5 to 100. Therefore, it is not allowed to configure a VLAN SAP with VID 100 on port 1/1/3, where a CP SAP is configured with CP 10, as follows:

```
*A:PE-1>config>service>vpls# sap 1/1/3:100 create
MINOR: SVCMGR #1602 The SAP-id is already in use - 1/1/3:100 is already configured
```

- No CP SAPs can be added with overlapping VLAN ranges on the same port for dot1q (or on the same port- and service-delimiting tag for QinQ), as follows. CP 1 contains VLAN range from 7 to 9, which overlaps with VLAN range from 5 to 100 in CP 10.

```
# on PE-1:
configure
    connection-profile-vlan 1 create
        vlan-range 7 to 9
    exit
```

```
*A:PE-1>config>service>vpls# sap 1/1/3:cp-1 create
MINOR: SVCMGR #1602 The SAP-id is already in use - 1/1/3:cp-10 is already configured
```

```
*A:PE-1>config>service>vpls# sap 1/2/1:100.cp-1 create
MINOR: SVCMGR #1602 The SAP-id is already in use - 1/2/1:100.cp-10 is already configured
```

  However, the CP can be referred to by SAPs on other ports for dot1q or for QinQ on other combinations of port and service-delimiting VLAN, as follows:

```
*A:PE-1>config>service>vpls# sap 1/2/1:101.cp-1 create
*A:PE-1>config>service>vpls>sap$ exit
```

- CP SAPs can be added when they contain non-overlapping VLAN ranges on the same port, as follows. CP 3 contains one VLAN range with only one VID: 3. This VLAN range (3) does not overlap with any VLAN range in the CP SAPs assigned to VPLS 1.

```
# on PE-1:
configure
    connection-profile-vlan 3 create
        vlan-range 3
    exit
```

```
*A:PE-1>config>service>vpls# sap 1/1/3:cp-3 create
*A:PE-1>config>service>vpls>sap# exit
```

```
*A:PE-1>config>service>vpls# sap 1/2/1:100.cp-3 create
*A:PE-1>config>service>vpls>sap# exit
```

VPLS 1 contains the following SAPs. There is no overlap between the VLAN ranges on a port (or port and service-delimiting tag for QinQ).

```
*A:PE-1# show service id 1 sap

===============================================================================
SAP(Summary), Service 1
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
1/1/3:cp-3                      1          1     none    1     none   Up   Up
1/1/3:cp-10                     1          1     none    1     none   Up   Up
1/2/1:cp-3.0                    1          1     none    1     none   Up   Up
1/2/1:1.11                      1          1     none    1     none   Up   Up
1/2/1:cp-10.*                   1          1     none    1     none   Up   Up
1/2/1:101.cp-1                  1          1     none    1     none   Up   Up
1/2/1:100.cp-3                  1          1     none    1     none   Up   Up
1/2/1:100.cp-10                 1          1     none    1     none   Up   Up
1/2/3:cp-10.*                   1          1     none    1     none   Up   Up
-------------------------------------------------------------------------------
Number of SAPs : 9
-------------------------------------------------------------------------------
===============================================================================
```

Constraints to be considered when applying CP SAPs in Layer 2 services are described in the Release Notes, section "Known Limitations" - "Services General".

Consumed resources for CP SAPs

The following SAPs are used on PE-1: nine SAPs are used in VPLS 1 and one SAP is used in Epipe 2:

```
*A:PE-1# show service sap-using

===============================================================================
Service Access Points
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
1/1/3:cp-3                      1          1     none    1     none   Up   Up
1/1/3:cp-10                     1          1     none    1     none   Up   Up
1/2/1:cp-3.0                    1          1     none    1     none   Up   Up
1/2/1:1.11                      1          1     none    1     none   Up   Up
1/2/1:cp-10.*                   1          1     none    1     none   Up   Up
1/2/1:101.cp-1                  1          1     none    1     none   Up   Up
1/2/1:100.cp-3                  1          1     none    1     none   Up   Up
1/2/1:100.cp-10                 1          1     none    1     none   Up   Up
1/2/3:cp-10.*                   1          1     none    1     none   Up   Up
1/2/1:200.cp-10                 2          1     none    1     none   Up   Up
-------------------------------------------------------------------------------
Number of SAPs : 10
-------------------------------------------------------------------------------
===============================================================================
```

Regular and default SAPs consume one SAP instance each, whereas CP SAPs consume a number of SAP instances equal to the number of VLANs in the range. The following shows that there are ten SAP entries (nine in use by VPLS 1 and one in use by Epipe 2), which can be regular, default, or CP SAP entries:

```
*A:PE-1# tools dump resource-usage system

===============================================================================
Resource Usage Information for System
===============================================================================
                                                  Total   Allocated        Free
-------------------------------------------------------------------------------
                 SAP Ingress QoS Policies |        3071           1        3070
                  SAP Egress QoS Policies |        3071           1        3070
            Ingress Queue-Group Templates |        2047           4        2043
             Egress Queue-Group Templates |        2047           5        2042
        Egress Port Queue-Group Instances |      163839           8      163831
         Ingress FP Queue-Group Instances |       16383           0       16383
              Fast Depth Monitored Queues |       50000           0       50000
                        Egress Port VPort |       40959           0       40959
        Dynamic Services Next-Hop Entries +      511999           0      511999
                   IPSec Next-Hop Entries -      500000           0      500000
              Subscriber Next-Hop Entries -      500000           0      500000
                              SAP Entries +      262143          10      262133
                     (in use by:   Apipe) -                       0
                     (in use by:   Cpipe) -                       0
                     (in use by:   Epipe) -                       1
                     (in use by:   Fpipe) -                       0
                     (in use by:   Ipipe) -                       0
                     (in use by:     Ies) -                       0
                     (in use by:  Mirror) -                       0
                     (in use by:    Vpls) -                       9
                     (in use by:    Vprn) -                       0
===============================================================================
```

However, the number of SAP instances consumed for card 1 FP 1 exceeds the number of SAP entries in the system, as follows:

```
*A:PE-1# tools dump resource-usage card 1 fp 1

===============================================================================
Resource Usage Information for Card Slot #1 FP #1
===============================================================================
                                                  Total   Allocated        Free
-------------------------------------------------------------------------------
---snip---
                               SAP Instances |      63999       1497      62502
---snip---
===============================================================================
```

The calculation of the number of SAP instances is as follows. In this example, CP 10 is used in five SAPs (four in VPLS 1 and one in Epipe 2) and contains the following VLAN ranges:

```
*A:PE-1# show connection-profile-vlan 10
===============================================================================
Connection Profile 10 Information
===============================================================================
Description : (Not Specified)
Last Change : 03/30/2021 07:35:20


===============================================================================
Connection Profile Vlan Eth Information
```

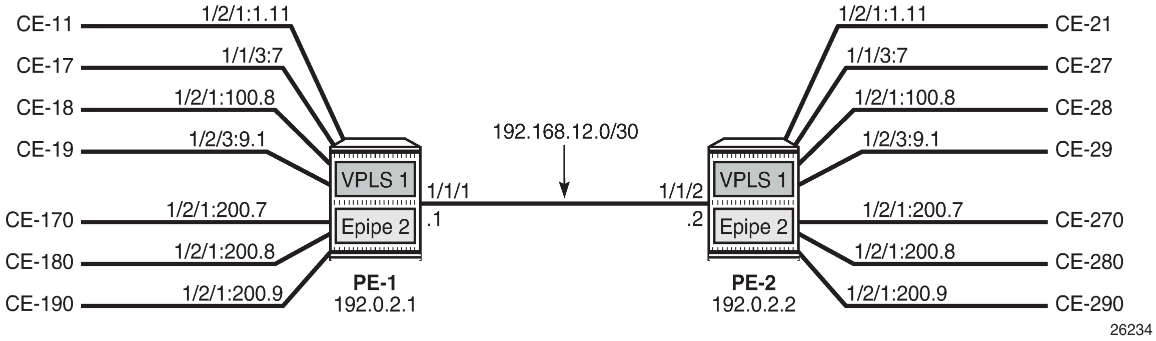
```
===============================================================================
Range Start         Range End                     Last Change
-------------------------------------------------------------------------------
5                   100                           03/30/2021 07:58:07
150                 300                           03/30/2021 07:58:07
350                 350                           03/30/2021 07:58:07
351                 400                           03/30/2021 07:59:58
===============================================================================
===============================================================================
```

The number of VLANs in the VLAN ranges of CP 10 equals 298. For each of the five SAP entries with CP 10, 298 SAP instances are used, for a total of 1490. As well, there is one CP SAP using CP 1 with three VLANs in the VLAN range from 7 to 9 (for three more SAP instances). Three CP SAPs use CP 3 with only VID 3 in the VLAN range (for three more SAP instances), and one SAP is a regular SAP that consumes one SAP instance. Therefore, the total number of SAP instances is 1497.

Configuration

Figure 305: Example topology shows the example topology used in this chapter.

Figure 305: Example topology



The initial configuration on the PEs includes the following:

- Cards, MDAs, ports
- Router interfaces
- IS-IS (or OSPF) between the PEs
- LDP between the PEs

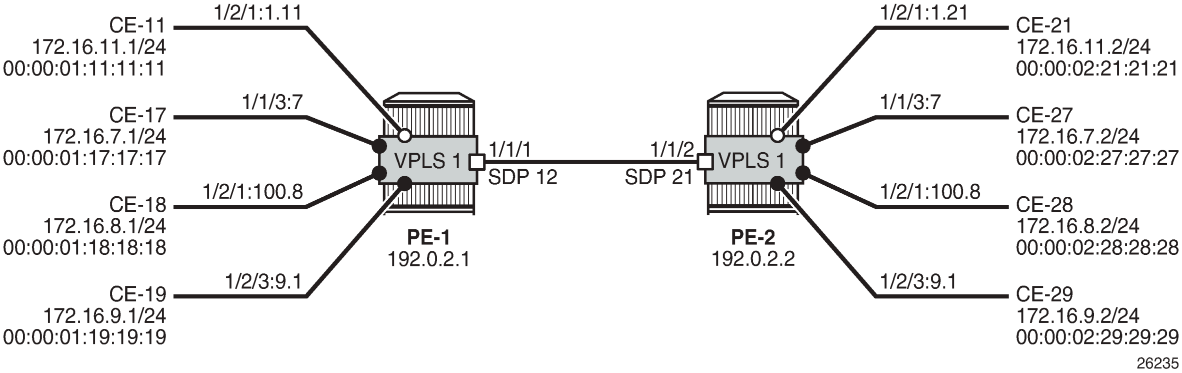
In this example, no BGP is configured and no BGP-EVPN will be configured in the VPLS and Epipe services. However, VLAN ranges can be applied in EVPN VPLS and EVPN Epipe services.

VLAN ranges in VPLS services

Figure 306: Example topology for VLAN ranges in VPLS 1 shows the example topology for VPLS 1 with a combination on VLAN SAPs and CP SAPs. The port:VID represents the port to which the CE is connected and the VID sent by the CE; for example, CE-17 is connected to port 1/1/3 on PE-1 and sends frames with

VID 7. When VLAN ranges are used, the port:VID 1/1/3:7 does not represent the configured SAP, which is 1/1/3:cp-1.

Figure 306: Example topology for VLAN ranges in VPLS 1



The service configuration for VPLS 1 on PE-1 is as follows:

```
# on PE-1:
configure
    service
        sdp 12 mpls create
            far-end 192.0.2.2
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        vpls 1 name "VPLS 1" customer 1 create
            stp
                shutdown
            exit
            sap 1/1/3:cp-1 create
                no shutdown
            exit
            sap 1/2/1:1.11 create
                no shutdown
            exit
            sap 1/2/1:100.cp-1 create
                no shutdown
            exit
            sap 1/2/3:cp-1.* create
                no shutdown
            exit
            spoke-sdp 12:1 create
                no shutdown
            exit
            no shutdown
        exit
```

The configuration of VPLS 1 on PE-2 is as follows:

```
# on PE-2:
configure
    service
```

```
        sdp 21 mpls create
            far-end 192.0.2.1
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        vpls 1 name "VPLS 1" customer 1 create
            stp
                shutdown
            exit
            sap 1/1/3:cp-1 create
                no shutdown
            exit
            sap 1/2/1:1.21 create
                no shutdown
            exit
            sap 1/2/1:100.cp-1 create
                no shutdown
            exit
            sap 1/2/3:cp-1.* create
                no shutdown
            exit
            spoke-sdp 21:1 create
                no shutdown
            exit
            no shutdown
        exit
```

When the CEs send traffic to each other, such as ICMP echo requests, the MAC addresses are learned in the SAPs, and the forwarding database (FDB) on PE-1 is as follows:

```
*A:PE-1# show service id 1 fdb detail

===============================================================================
Forwarding Database, Service 1
===============================================================================
ServId     MAC               Source-Identifier       Type     Last Change
            Transport:Tnl-Id                         Age
-------------------------------------------------------------------------------
1          00:00:01:11:11:11 sap:1/2/1:1.11          L/0      03/30/21 09:18:14
1          00:00:01:17:17:17 sap:1/1/3:cp-1          L/0      03/30/21 09:17:58
1          00:00:01:18:18:18 sap:1/2/1:100.cp-1      L/0      03/30/21 09:17:58
1          00:00:01:19:19:19 sap:1/2/3:cp-1.*        L/0      03/30/21 09:18:14
1          00:00:02:21:21:21 sdp:12:1                L/0      03/30/21 09:19:12
1          00:00:02:27:27:27 sdp:12:1                L/0      03/30/21 09:19:02
1          00:00:02:28:28:28 sdp:12:1                L/0      03/30/21 09:19:02
1          00:00:02:29:29:29 sdp:12:1                L/0      03/30/21 09:19:12
-------------------------------------------------------------------------------
No. of MAC Entries: 8
-------------------------------------------------------------------------------
Legend:  L=Learned O=Oam P=Protected-MAC C=Conditional S=Static Lf=Leaf
===============================================================================
```

VLAN manipulation in dot1q SAPs

Figure 307: Customer VIDs are popped and pushed by dot1q VLAN SAPs shows the VLAN manipulation for VLAN SAPs. CE-17 and CE-18 are connected to VLAN SAPs, where the VLAN tag with VID 7 will

be popped or pushed. VLAN translation is possible, but does not apply. The configuration of the SAPs in VPLS 1 on PE-1 and PE-2 is modified as follows:

```
# on PE-1, PE-2:
configure
    service
        vpls "VPLS 1"
            sap 1/1/3:cp-1
                shutdown
            exit
            no sap 1/1/3:cp-1
            sap 1/1/3:7 create
                no shutdown
            exit
```

Figure 307: Customer VIDs are popped and pushed by dot1q VLAN SAPs

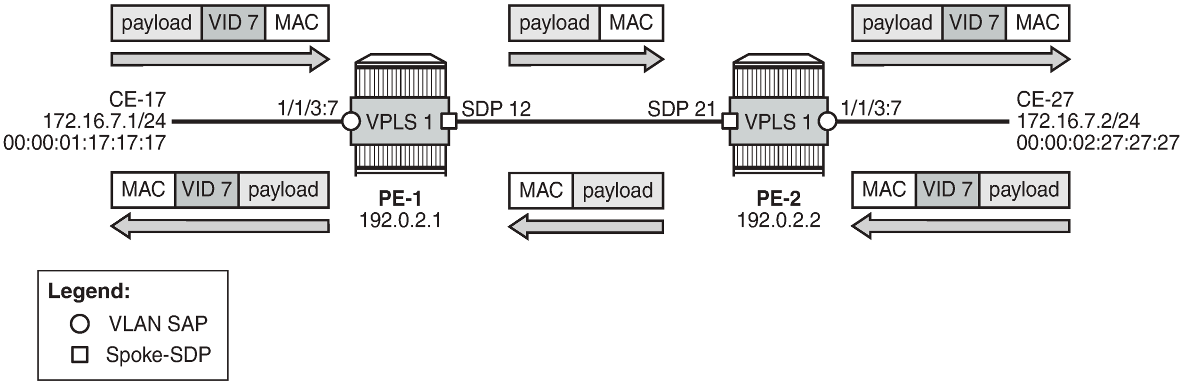
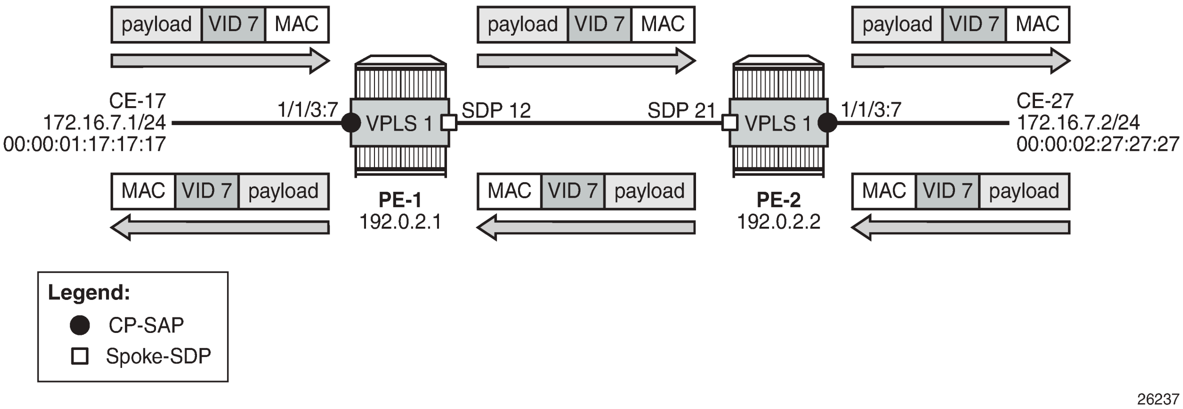


Figure 308: Customer VID is preserved between two dot1q CP SAPs shows how the customer VID 7 is preserved between CE-17 and CE-27 when CP SAPs are used instead of VLAN SAPs. The configuration for the SAPs is modified as follows:

```
# on PE-1, PE-2:
configure
    service
        vpls "VPLS 1"
            sap 1/1/3:7
                shutdown
            exit
            no sap 1/1/3:7
            sap 1/1/3:cp-1 create
                no shutdown
            exit
```

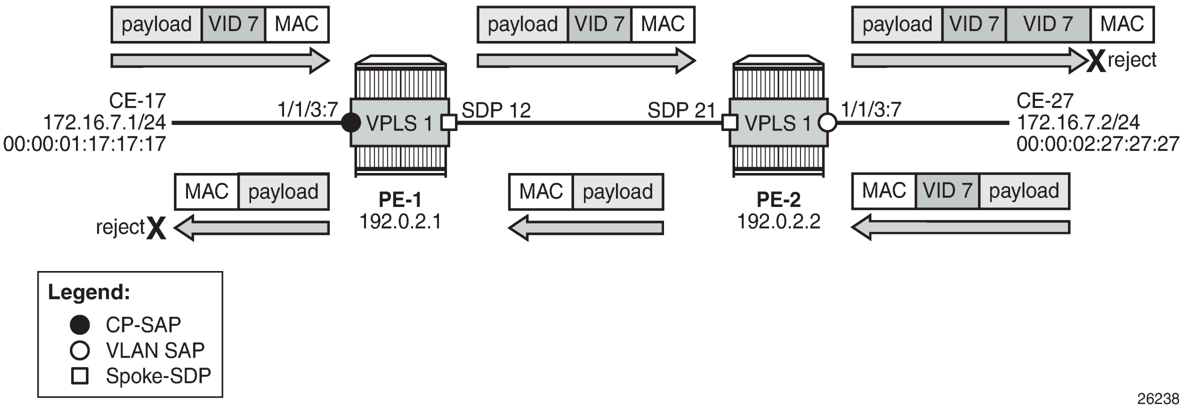
CE-17 sends frames with VID 7 to dot1q CP SAP 1/1/3:cp-1 in VPLS 1 on PE-1, and this CP SAP preserves the VLAN tag. When the frames with VID 7 reach the egress CP SAP 1/1/3:cp-1 of VPLS 1 on PE-2, the egress CP SAP preserves the VID, and the frames are forwarded to CE-27. Traffic in the opposite direction is treated in the same way: the customer VID is preserved between the CEs.

Figure 308: Customer VID is preserved between two dot1q CP SAPs



No traffic is possible between a CP SAP in VPLS 1 on PE-1 and a VLAN SAP in VPLS 1 on PE-2, as shown in Figure 309: No traffic between dot1q CP SAP and dot1q VLAN SAP.

Figure 309: No traffic between dot1q CP SAP and dot1q VLAN SAP



The CP SAP 1/1/3:cp-1 in VPLS 1 on PE-1 remains unchanged, whereas the SAP in VPLS 1 on PE-2 is reconfigured as VLAN SAP 1/1/3:7 for VLAN 7, as follows:

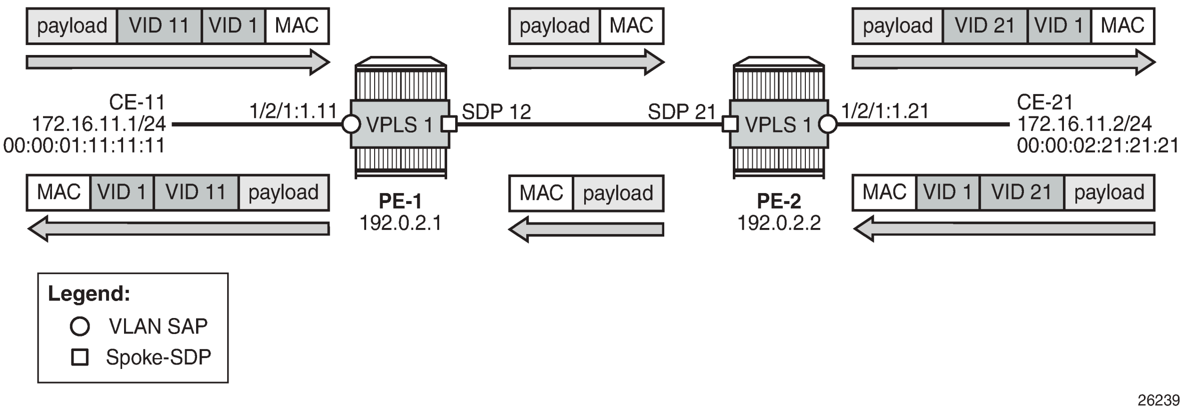
```
# on PE-2:
configure
    service
        vpls "VPLS 1"
            sap 1/1/3:cp-1
                shutdown
            exit
            no sap 1/1/3:cp-1
            sap 1/1/3:7 create
                no shutdown
            exit
```

Frames from CE-17 are forwarded by CP SAP 1/1/3:cp-1 in VPLS 1 on PE-1 without any changes to the VLAN tag. The tagged frames reach the VLAN SAP 1/1/3:7, where another VLAN tag with VID 7 is pushed onto the frame. The receiver CE-27 rejects the double-tagged frame. When CE-27 sends traffic to CE-17, the VLAN SAP 1/1/3:7 in VPLS 1 on PE-2 pops the VLAN tag and the frame is forwarded untagged to PE-1. The CP SAP 1/1/3:cp-1 on PE-1 does not push any VLAN tag and the frame is forwarded untagged to CE-17, where it is rejected.

VLAN manipulation in QinQ SAPs

Figure 310: Traffic between two QinQ VLAN SAPs - VLAN translation shows the VLAN manipulation in QinQ VLAN SAPs that pop and push the VLAN labels. In the example, the customer VID is translated.

Figure 310: Traffic between two QinQ VLAN SAPs - VLAN translation



CE-11 sends double-tagged traffic to QinQ VLAN SAP 1/2/1:1.11 in VPLS 1 on PE-1. This VLAN SAP pops both labels and forwards the frame untagged to PE-2. The egress VLAN SAP 1/2/1:1.21 in VPLS 1 on PE-2 pushes a label stack with two labels: the inner label with VID 21 and the outer label with VID 1. Both VIDs can be translated, but in this example, only the inner label gets another VID.

Figure 311: No traffic between two QinQ CP SAPs - VLAN translation not supported shows that VLAN translation is not possible between two QinQ CP SAPs. In the example, the outer tag with VID 1 is popped by the CP SAPs (VLAN translation is possible for this VLAN tag, but not done here) and the inner tag with VID 11 or 21 is preserved by the CP SAPs, which implies that the received frames will be rejected.

In this example, CP 2 is configured on both PE-1 and PE-2 with one VLAN range with one VID (11 or 21), as follows:

```
# on PE-1:
configure
    connection-profile-vlan 2 create
        vlan-range 11
    exit
```

```
# on PE-2:
configure
    connection-profile-vlan 2 create
        vlan-range 21
    exit
```

The VLAN SAP 1/2/1:1.11 is replaced by CP SAP 1/2/1:1.cp-2, as follows:

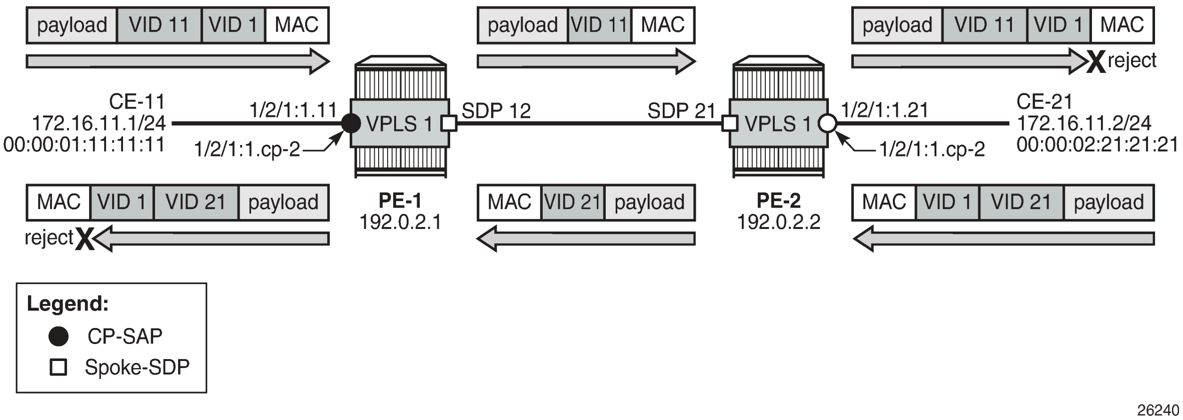
```
# on PE-1:
```

```
configure
    service
        vpls "VPLS 1"
            sap 1/2/1:1.11
                shutdown
            exit
            no sap 1/2/1:1.11
            sap 1/2/1:1.cp-2 create
        exit
```

Likewise, the VLAN 1/2/1:1.21 is replaced by CP SAP 1/2/1:1.cp-2, as follows:

```
# on PE-2:
configure
    service
        vpls "VPLS 1"
            sap 1/2/1:1.21
                shutdown
            exit
            no sap 1/2/1:1.21
            sap 1/2/1:1.cp-2 create
        exit
```

Figure 311: No traffic between two QinQ CP SAPs - VLAN translation not supported

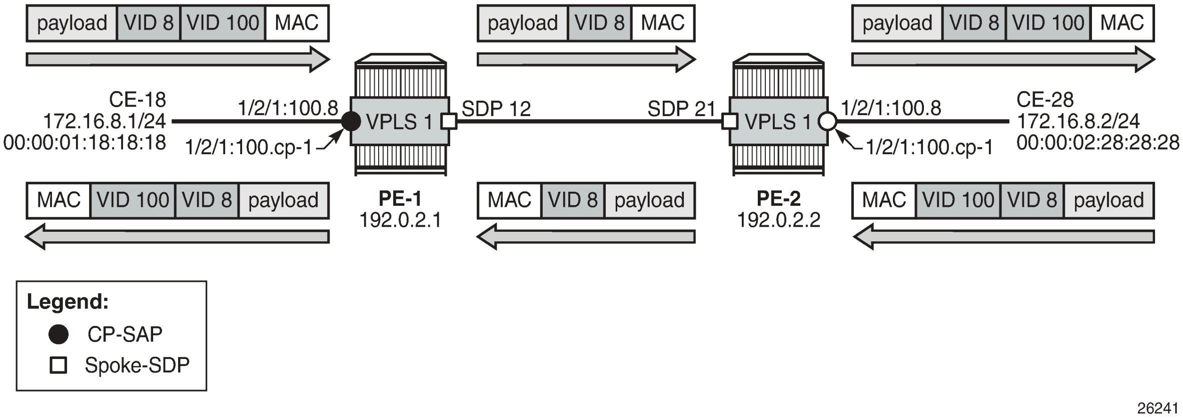


CE-11 sends double-tagged frames to SAP 1/2/1:1.cp-2 in VPLS 1 on PE-1. This CP SAP pops the outer tag with VID 1, but preserves the VLAN tag with VID 11. The single-tagged frame is sent to PE-2 where CP SAP 1/2/1:1.cp-2 pushes an outer tag with VID 1 onto the frame. This double-tagged frame is sent to CE-12 where it is rejected, because an inner label with VID 21 is expected.

When CE-21 sends frames to CE-11, the frames will be double-tagged with inner tag VID 21 and outer tag 1. The outer tag is popped by the ingress SAP 1/2/1:1.cp-2 in VPLS 1 on PE-2, but the inner tag is preserved. The egress SAP 1/2/1:1.cp-2 in VPLS 1 on PE-1 preserves the inner tag with VID 21 and pushes an outer tag with VID 1. This double-tagged frame is rejected by CE-11, because another inner tag is expected, with VID 11 instead of VID 21.

Figure 312: Traffic between two QinQ CP SAPs - no VLAN translation shows how traffic is sent between two QinQ CP SAPs without VLAN translation. Both CE-18 and CE-28 send double-tagged frames with inner tag VID 8 and outer tag VID 100. The tag with VID 100 need not be the same on both CEs, because it is popped and pushed by the CP SAPs; only the tag with VID 8 from the VLAN range must be unchanged.

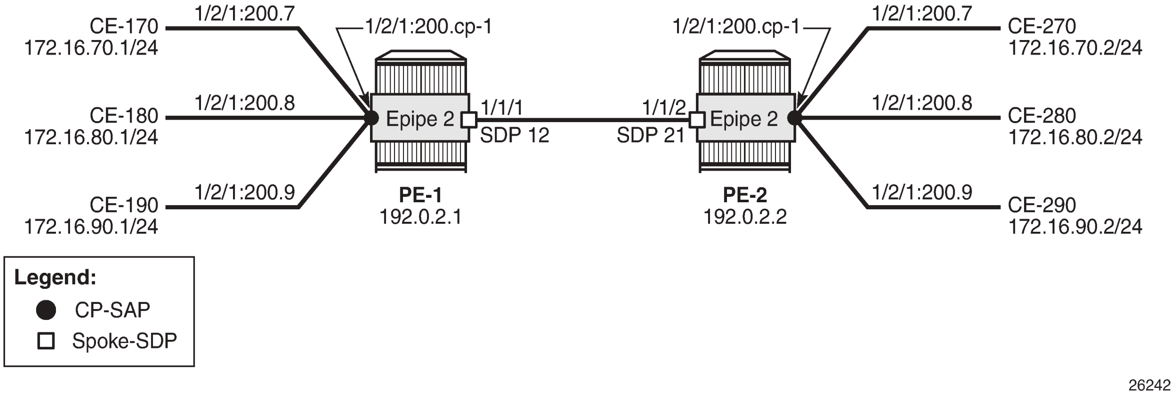
Figure 312: Traffic between two QinQ CP SAPs - no VLAN translation



VLAN ranges in Epipe services

Figure 313: Example topology for VLAN ranges in Epipe 2 shows the example topology for VLAN ranges in Epipe 2.

Figure 313: Example topology for VLAN ranges in Epipe 2



Epipe 2 is configured with one CP SAP and a spoke-SDP, as follows:

```
# on PE-1:
configure
    service
        sdp 12 mpls create
            far-end 192.0.2.2
            ldp
            keep-alive
                shutdown
            exit
        no shutdown
        exit
        epipe 2 name "Epipe 2" customer 1 create
            sap 1/2/1:200.cp-1 create
                no shutdown
            exit
            spoke-sdp 12:2 create
                no shutdown
            exit
            no shutdown
        exit
```

CE-170 and CE-270 send double-tagged frames with inner VID 7 and outer VID 200. The inner VID 7 is preserved by the CP SAPs; therefore, CE-170 can only communicate with CE-270, not with any other CE at the other end, because they have different customer VIDs.

Conclusion

CP SAPs can be used to build services that can be bundled as per MEF 10.3 and RFC 7432. Multiple customer VIDs can be mapped to one CP-SAP.

VXLAN Forwarding Path Extension

This chapter provides information about VXLAN Forwarding Path Extension.

Topics in this chapter include:

Applicability

This chapter was initially written based on SR OS Release 15.0.R4, but the CLI in the current edition corresponds to SR OS Release 21.2.R2. Virtual eXtensible Local Area Network (VXLAN) Forwarding Path Extension (FPE) is supported in SR OS Release 14.0.R4, and later. IPv6 addresses are supported for EVPN-VXLAN BGP peering in SR OS Release 15.0.R1, and later.

Overview

Use cases

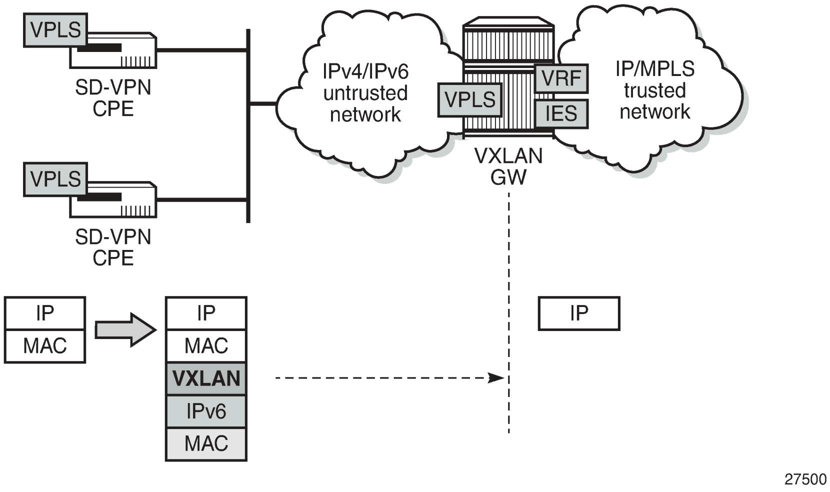
VXLAN Forwarding Path Extension (FPE) is an SR OS feature that enables VXLAN tunnels to terminate on non-system IPv4 and IPv6 Destination Addresses (DAs). The non-system IPv4/IPv6 VXLAN termination feature can be applied in the following use cases:

- VXLAN Gateway (GW) in Software-Defined VPNs (SD-VPNs)
- VXLAN IPv6 underlay for Data Centers (DCs)

VXLAN GW in SD-VPNs

Traffic transported on a VXLAN is usually connected to a trusted environment through a VPRN running in a private IP/MPLS network. The VXLAN GW system IP address is used for all internal management and MPLS termination in the trusted network. However, in this use case, SR OS routers are expected to be used as a VXLAN GW in SD-VPNs where the VXLAN GW terminates untrusted VXLAN tunnels initiated on the SD-VPN CPEs and forwards packets to a trusted IP/MPLS network, as shown in Figure 314: VXLAN GW in an SD-VPN.

Figure 314: VXLAN GW in an SD-VPN

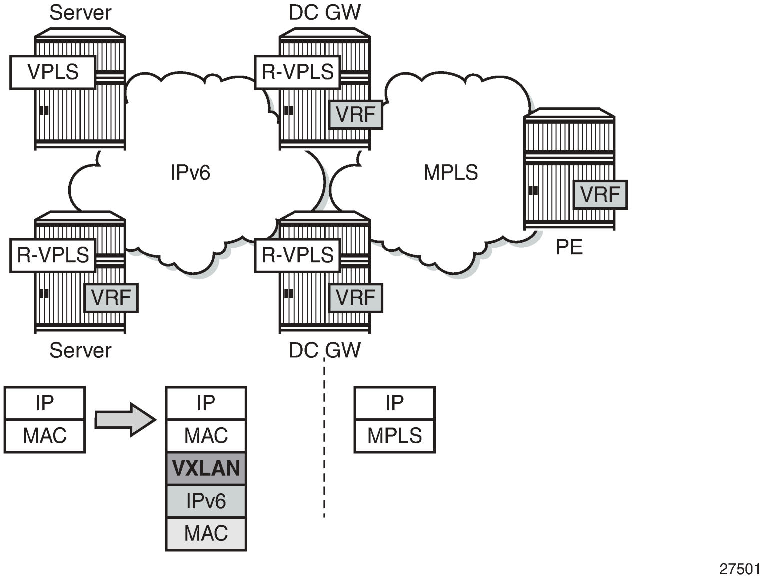


For security reasons, service providers will not expose system IP addresses to the untrusted IP network. Therefore, an IPv4 or IPv6 loopback address will be defined and used for VXLAN termination. The VXLAN tunnel can be terminated in a VPLS, an Epipe, or an R-VPLS service connected to a VPRN.

VXLAN IPv6 underlay for DCs

Some service providers migrate their entire network infrastructure to IPv6, including the DC network, so the DC GW must be able to terminate a VXLAN over an IPv6 infrastructure. Layer 2 (VPLS termination) and Layer 3 (R-VPLS termination) DC interconnect are both supported. Figure 315: VXLAN IPv6 underlay for DC shows the VXLAN IPv6 underlay for DC.

Figure 315: VXLAN IPv6 underlay for DC



VXLAN FPE function

The following applies to VXLAN FPE:

- In an SR OS node, VXLAN tunnels can be terminated in four different VXLAN Tunnel Endpoints (VTEPs):
  - System IPv4 address
  - Up to three non-system IPv4/IPv6 addresses

  This limit is based on the number of supported source IP addresses that can be used for VXLAN encapsulation.
- The preceding four terminating IP addresses can be used in addition to the Assisted Replication IP address (AR IP). The AR IP does not count against this limit of four VTEPs. See chapter Layer 2 Multicast Optimization for EVPN-VXLAN — Assisted Replication for more information about AR.
- VXLAN FPE requires PXC ports; see chapter *Port Cross-connect (PXC)*.
  - Ingress traffic from a VXLAN with an IP DA equal to a loopback address will be redirected to the PXC port where the IP header will get additional processing.
  - Usually, only the ingress traffic from the VXLAN is redirected to the PXC port. The egress traffic to the VXLAN tunnel can go straight out of the egress network port, except for R-VPLS traffic toward an IPv6 VXLAN that is redirected to the PXC port.
- The VPLS/R-VPLS functionality is not impacted by the choice of VTEP termination (system IP address or not).

Provisioning model

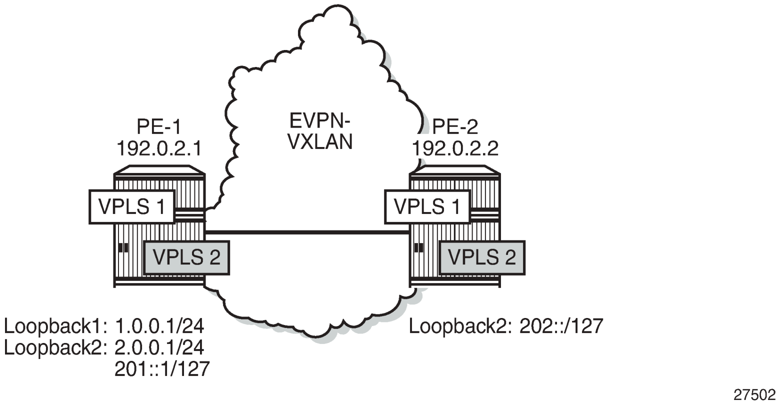
Non-system IP VXLAN termination and VXLAN IPv6 underlay are both provisioned as per the following steps:

1. Create an FPE
2. Associate the FPE with VXLAN termination
3. Configure a router loopback interface
4. Configure non-system VXLAN termination VTEP addresses
5. Add the service configuration

Configuration

Figure 316: Example topology for VXLAN FPE shows the example topology with two PEs in an EVPN-VXLAN network. The loopback addresses in the base router will be used for non-system IP VXLAN termination.

Figure 316: Example topology for VXLAN FPE



The initial configuration includes the cards, MDAs, ports, router interfaces and IGP. BGP is configured for address family EVPN, for example on PE-1 as follows:

```
# on PE-1:
configure
    router
        autonomous-system 64500
        bgp
            rapid-withdrawal
            split-horizon
            rapid-update evpn
            group "internal"
                family evpn
                peer-as 64500
                neighbor 192.0.2.2
                exit
            exit
        exit
```

In this example, the BGP peering is IPv4-based, but EVPN-VXLAN routes can also be exchanged between IPv6 BGP peers.

Non-system IP VXLAN termination

Create FPEs

PXC is used as a simple back-to-back cross-connect. An FPE uses the PXC ports assigned in the FPE path, either a PXC port or a LAG-based PXC. For non-system IP VXLAN terminations between VPLSs, the PXC is only required on the ingress (from VXLAN, or from PE-1 to GW PE-2). The following PXCs are created on PE-1:

```
# on PE-1:
configure
    port-xc
        pxc 1 create
            port 1/2/1
            no shutdown
        exit
        pxc 2 create
            port 1/2/2
            no shutdown
        exit
```

The PXC auto-created sub-ports and ports are enabled as follows.

```
# on PE-1:
configure
    port pxc-1.a
        no shutdown
    exit
    port pxc-1.b
        no shutdown
    exit
    port 1/2/1
        no shutdown
    exit
```

```
*A:PE-1# show port pxc 1

===============================================================================
Ports on Port Cross Connect 1
===============================================================================
Port          Admin Link Port    Cfg  Oper LAG/ Port Port Port   C/QS/S/XFP/
Id            State      State   MTU  MTU  Bndl Mode Encp Type   MDIMDX
-------------------------------------------------------------------------------
pxc-1.a       Up    Yes  Up      1574 1574    - hybr dotq xgige
pxc-1.b       Up    Yes  Up      1574 1574    - hybr dotq xgige
===============================================================================
```

The following FPEs use the PXCs.

```
# on PE-1, PE-2:
configure
    fwd-path-ext
        fpe 1 create
            path pxc 1
        exit
        fpe 2 create
            path pxc 2
        exit
```

These FPEs are created without defining a range of SDP IDs. SDP IDs are required in case of R-VPLS services terminating IPv6 VXLAN tunnels, where the FPE is also used at the egress and an internal static SDP is created to allow for the required extra processing.

When the FPE has no VXLAN termination associated, no internal router interfaces are created, so the only router interfaces are the system interface and the interface between PE-1 and PE-2, as follows.

```
*A:PE-1# show router interface

===============================================================================
Interface Table (Router: Base)
===============================================================================
Interface-Name                   Adm       Opr(v4/v6)  Mode    Port/SapId
   IP-Address                                                  PfxState
-------------------------------------------------------------------------------
int-PE-1-PE-2                    Up        Up/Down     Network 1/1/1
```

```
   192.168.12.1/30                                             n/a
system                           Up        Up/Down     Network system
   192.0.2.1/32                                                n/a
-------------------------------------------------------------------------------
Interfaces : 2
===============================================================================
```

Associate the FPEs with VXLAN termination

The following command associates the FPEs with VXLAN termination.

```
# on PE-1, PE-2:
configure
    fwd-path-ext
        sdp-id-range from 10000 to 10127
        fpe 1 create
            path pxc 1
            vxlan-termination
        exit
        fpe 2 create
            path pxc 2
            vxlan-termination
        exit
```

When attempting to associate the FPEs with VXLAN termination without configuring a range of SDP IDs for FPE, the following error is raised:

```
*A:PE-1>config>fwd-path-ext>fpe# vxlan-termination
MINOR: FPE #1021 sdp-id-range is not configured
```

After the FPEs are associated with VXLAN terminations, the system creates two internal router interfaces per FPE, one per PXC sub-port:

```
*A:PE-1# show router interface

===============================================================================
Interface Table (Router: Base)
===============================================================================
Interface-Name                   Adm       Opr(v4/v6)  Mode    Port/SapId
   IP-Address                                                  PfxState
-------------------------------------------------------------------------------
_tmnx_fpe_1.a                    Up        Up/Up       Network pxc-1.a:1
   fe80::100/64                                                PREFERRED
_tmnx_fpe_1.b                    Up        Up/Up       Network pxc-1.b:1
   fe80::101/64                                                PREFERRED
_tmnx_fpe_2.a                    Up        Up/Up       Network pxc-2.a:1
   fe80::200/64                                                PREFERRED
_tmnx_fpe_2.b                    Up        Up/Up       Network pxc-2.b:1
   fe80::201/64                                                PREFERRED
int-PE-1-PE-2                    Up        Up/Down     Network 1/1/1
   192.168.12.1/30                                             n/a
system                           Up        Up/Down     Network system
   192.0.2.1/32                                                n/a
-------------------------------------------------------------------------------
Interfaces : 6
===============================================================================
```

Configure router loopback interfaces

The following loopback interfaces are configured in PE-1 and added to the IS-IS context:

```
# on PE-1:
configure
    router Base
        interface "loopback1"
            address 1.0.0.1/24
            loopback
        exit
        interface "loopback2"
            address 2.0.0.1/31
            loopback
            ipv6
                address 201::/127
            exit
        exit
        isis
            interface "loopback1"
            exit
            interface "loopback2"
            exit
        exit
```

A non /32 or /128 subnet must be assigned to the loopback interface, because the system cannot terminate VXLAN on a local interface address. In the preceding example, all addresses in the subnet 1.0.0.0/24 can be used for VXLAN tunnel termination, except for 1.0.0.1. The subnet will be advertised by the IGP. The subnet can be as small as /31 or /127, as for example for interface "loopback2".

In this scenario, only one loopback interface with an IPv4 address is sufficient: interface "loopback1" with IPv4 address 1.0.0.1/24. There is no need to configure loopback interfaces in the GW PE-2, because VXLAN FPE is only required in the ingress (from VXLAN to GW).

Configure non-system VTEP addresses

Up to three non-system VTEP addresses can be configured to terminate VXLAN tunnels and their corresponding FPEs; on PE-1 as follows:

```
# on PE-1:
configure
    service
        system
            vxlan
                tunnel-termination 1.0.0.2 fpe 1 create
                tunnel-termination 2.0.0.2 fpe 2 create
                tunnel-termination 201::1 fpe 2 create
            exit
        exit
```

No non-system VTEP addresses need to be configured on PE-2.

When attempting to configure the IP address of the loopback interface as a VXLAN tunnel termination, the following error is raised:

```
*A:PE-1>config>service>system>vxlan# tunnel-termination 1.0.0.1 fpe 1 create
```

```
MINOR: SVCMGR #8353 VXLAN Tunnel termination IP address cannot be configured - IP address in
 use by another application or matches a local interface IP address
```

When attempting to configure more than three non-system VTEP addresses, the following error is raised:

```
*A:PE-1>config>service>system>vxlan# tunnel-termination 1.0.0.100 fpe 1 create
MINOR: SVCMGR #8353 VXLAN Tunnel termination IP address cannot be configured - Reached system
 limit of VXLAN tunnel-termination addresses
```

When the non-system VTEP addresses are configured, an internal loopback interface "_tmnx_vli_vxlan_1_131077" is created that can respond to ICMP requests.

```
*A:PE-1# show router interface

===============================================================================
Interface Table (Router: Base)
===============================================================================
Interface-Name                   Adm       Opr(v4/v6)  Mode    Port/SapId
   IP-Address                                                  PfxState
-------------------------------------------------------------------------------
_tmnx_fpe_1.a                    Up        Up/Up       Network pxc-1.a:1
   fe80::100/64                                                PREFERRED
_tmnx_fpe_1.b                    Up        Up/Up       Network pxc-1.b:1
   fe80::101/64                                                PREFERRED
_tmnx_fpe_2.a                    Up        Up/Up       Network pxc-2.a:1
   fe80::200/64                                                PREFERRED
_tmnx_fpe_2.b                    Up        Up/Up       Network pxc-2.b:1
   fe80::201/64                                                PREFERRED
_tmnx_vli_vxlan_1_131077         Up        Up/Up       Network loopback
   1.0.0.2/32                                                  n/a
   2.0.0.2/32                                                  n/a
   201::1/128                                                  PREFERRED
   fe80::f:ffff:fe00:0/64                                      PREFERRED
int-PE-1-PE-2                    Up        Up/Down     Network 1/1/1
   192.168.12.1/30                                             n/a
loopback1                        Up        Up/Down     Network loopback
   1.0.0.1/24                                                  n/a
loopback2                        Up        Up/Up       Network loopback
   2.0.0.1/31                                                  n/a
   201::/127                                                   PREFERRED
   fe80::f:ffff:fe00:0/64                                      PREFERRED
system                           Up        Up/Down     Network system
   192.0.2.1/32                                                n/a
-------------------------------------------------------------------------------
Interfaces : 9
===============================================================================
```

The system does not verify whether there is a local base router loopback interface with a subnet corresponding to the VTEP address. If a tunnel termination address is configured and the FPE is up, the system will start terminating VXLAN traffic and responding using ICMP for that address, regardless of the presence of a loopback interface in the base router. It is also possible that a non-loopback interface has an IP address in the configured subnet.

Configure the VPLS

A VPLS will be configured with EVPN-VXLAN enabled. By default, the system IP address will be used as the source VTEP of the VXLAN-encapsulated frames. This default behavior can be overruled by the **vxlan-**

src-vtep command in the VPLS. The IP address corresponds to the non-system VTEP address configured in the preceding step (VXLAN tunnel termination). VPLS 1 is configured on PE-1 as follows:

```
# on PE-1:
configure
    service
        vpls 1 name "EVI-1" customer 1 create
            vxlan instance 1 vni 1 create
            exit
            vxlan-src-vtep 1.0.0.2
            bgp
            exit
            bgp-evpn
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            sap 1/1/2:1 create
                no shutdown
            exit
            no shutdown
        exit
```

When attempting to configure an IP address different from the VTEP addresses, the following error is raised:

```
*A:PE-1# configure service vpls 1 vxlan-src-vtep 1.0.0.99
MINOR: SVCMGR #8351 VXLAN Tunnel termination IP address does not exist
```

A different VTEP address can be configured as **vxlan-src-vtep** in different services on the same PE, as follows:

```
# on PE-1:
configure
    service
        vpls 2 name "EVI-2" customer 1 create
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 2 create
            exit
            vxlan-src-vtep 201::1
            bgp
            exit
            bgp-evpn
                evi 2
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
```

The configuration of VPLS 1 on PE-2 does not include any VTEP address, because it is not required in the egress, as follows:

```
# on PE-2:
configure
    service
        vpls 1 name "VPLS 1" customer 1 create
            vxlan instance 1 vni 1 create
            exit
            bgp
            exit
            bgp-evpn
                evi 1
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
```

When a vxlan-src-vtep is configured in VPLS 1 on PE-1, this VTEP address will be used as the IP source VTEP for VPLS 1 and BGP will use this VTEP to the BGP NLRI next-hop, as shown in the following BGP route update messages.

The following BGP EVPN inclusive multicast route sent by PE-1 shows the configured source VTEP address 1.0.0.2 as NLRI next-hop, as originator address, and as tunnel endpoint.

```
# on PE-1:
1 2021/05/06 09:40:06.914 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 77
    Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 1.0.0.2
        Type: EVPN-INCL-MCAST Len: 17 RD: 192.0.2.1:1, tag: 0, orig_addr len: 32,
                              orig_addr: 1.0.0.2
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:1
        bgp-tunnel-encap:VXLAN
    Flag: 0xc0 Type: 22 Len: 9 PMSI:
        Tunnel-type Ingress Replication (6)
        Flags: (0x0)[Type: None BM: 0 U: 0 Leaf: not required]
        MPLS Label 1
        Tunnel-Endpoint 1.0.0.2
"
```

The following BGP EVPN-MAC route sent by PE-1 shows the configured VTEP for VPLS 1 as NLRI next-hop:

```
# on PE-1:
8 2021/05/06 09:41:43.212 UTC MINOR: DEBUG #2001 Base Peer 1: 192.0.2.2
"Peer 1: 192.0.2.2: UPDATE
Peer 1: 192.0.2.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 81
    Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
        Address Family EVPN
        NextHop len 4 NextHop 1.0.0.2
        Type: EVPN-MAC Len: 33 RD: 192.0.2.1:1 ESI: ESI-0, tag: 0, mac len: 48
                       mac: ca:fe:01:10:10:10, IP len: 0, IP: NULL, label1: 1
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 16 Extended Community:
        target:64500:1
        bgp-tunnel-encap:VXLAN
"
```

A BGP peer policy might override the NLRI next-hop created due to the **vxlan-src-vtep** configuration.

The following shows that the source VTEP address on PE-1 is 1.0.0.2:

```
*A:PE-1# show service id 1 vxlan
===============================================================================
VPLS VXLAN
===============================================================================
Vxlan Src Vtep IP: 1.0.0.2

===============================================================================
Vxlan Instance
===============================================================================
VXLAN Instance                VNI         AR          Oper-flags   VTEP
                                                                   security
-------------------------------------------------------------------------------
1                             1           none        none         disabled
-------------------------------------------------------------------------------
Number of Entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

The following command on PE-1 shows that the egress VTEP is 192.0.2.2:

```
*A:PE-1# show service id 1 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
Instance    VTEP Address                            Egress VNI  EvpnStatic Num
 Mcast       Oper State                              L2 PBR     SupBcasDom MACs
-------------------------------------------------------------------------------
1           192.0.2.2                               1           evpn      0
 BUM         Up                                      No          No
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

The following shows that no source VTEP address is configured on PE-2:

```
*A:PE-2# show service id 1 vxlan
===============================================================================
VPLS VXLAN
===============================================================================
Vxlan Src Vtep IP: N/A

===============================================================================
Vxlan Instance
===============================================================================
VXLAN Instance                VNI         AR          Oper-flags   VTEP
                                                                   security
-------------------------------------------------------------------------------
1                             1           none        none         disabled
-------------------------------------------------------------------------------
Number of Entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

The following command on PE-2 shows that the egress VTEP is 1.0.0.2:

```
*A:PE-2# show service id 1 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
Instance    VTEP Address                            Egress VNI  EvpnStatic Num
 Mcast       Oper State                              L2 PBR     SupBcasDom MACs
-------------------------------------------------------------------------------
1           1.0.0.2                                 1           evpn       1
 BUM         Up                                      No          No
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

Underlay IPv6 VXLAN termination

The configuration for underlay IPv6 VXLAN termination is similar to the non-system IP VXLAN termination. In the following example, R-VPLS 2 is configured; therefore, non-system VTEP addresses are configured in PE-2 as well as in PE-1. The changes required in PE-1 are as follows.

- IPv6 must be enabled on the router interfaces
- IPv6 native routing is configured in IS-IS
- IPv6 addresses are loopback address 201::/127 and VTEP address 201::1

```
# on PE-1:
configure
    port-xc
```

```
        pxc 2 create
            port 1/2/2
            no shutdown
        exit
    exit
    port pxc-2.a
        no shutdown
    exit
    port pxc-2.b
        no shutdown
    exit
    port 1/2/2
        no shutdown
    exit
    fwd-path-ext
        sdp-id-range from 10000 to 10127
        fpe 2 create
            path pxc 2
            vxlan-termination
        exit
    exit
    router Base
        interface "int-PE-1-PE-2
            address 192.168.12.1/30
            port 1/1/1
            ipv6
            exit
        exit
        interface "loopback2"
            address 2.0.0.1/31
            loopback
            ipv6
                address 201::/127
            exit
        exit
        isis 0
            ipv6-routing native
            interface "loopback2"
            exit
        exit
    exit
    service
        system
            vxlan
                tunnel-termination 201::1 fpe 2 create
            exit
        exit
        vpls 2 name "EVI-2" customer 1 create
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 2 create
            exit
            vxlan-src-vtep 201::1
            bgp
            exit
            bgp-evpn
                evi 2
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
```

The service configuration on PE-2 is as follows.

```
# on PE-2:
configure
    service
        system
            vxlan
                tunnel-termination 202:: fpe 2 create
            exit
        exit
        vpls 2 name "EVI-2" customer 1 create
            allow-ip-int-bind
            exit
            vxlan instance 1 vni 2 create
            exit
            vxlan-src-vtep 202::
            bgp
            exit
            bgp-evpn
                evi 2
                vxlan bgp 1 vxlan-instance 1
                    no shutdown
                exit
            exit
            stp
                shutdown
            exit
            no shutdown
        exit
```

The routing table for IPv6 on PE-1 shows that an internal static route is configured for the source VTEP 201::1 using the FPE internal interface "_tmnx_fpe_2.a". The route to egress VTEP 202:: is an IS-IS route.

```
*A:PE-1# show router route-table ipv6

===============================================================================
IPv6 Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
201::/127                                     Local   Local     00h16m34s  0
       loopback2                                                    0
201::1/128                                    Remote  Static    00h11m53s  5
       fe80::201-"_tmnx_fpe_2.a"                                    1
202::/127                                     Remote  ISIS      00h00m06s  15
       fe80::14:1ff:fe01:1-"int-PE-1-PE-2"                          10
-------------------------------------------------------------------------------
No. of Routes: 3
```

Likewise, the routing table for IPv6 on PE-2 shows an internal static route for source VTEP 202:: using the FPE internal interface "_tmnx_fpe_2.a":

```
*A:PE-2# show router route-table ipv6

===============================================================================
IPv6 Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
201::/127                                       Remote  ISIS      00h00m06s  15
       fe80::10:1ff:fe01:1-"int-PE-2-PE-1"                            10
202::/127                                       Local   Local     00h00m12s  0
       loopback2                                                      0
202::/128                                       Remote  Static    00h00m13s  5
       fe80::201-"_tmnx_fpe_2.a"                                      1
-------------------------------------------------------------------------------
No. of Routes: 3
```

When non-system IPv6 VTEP addresses are used in an R-VPLS, VTEP addresses need to be configured on ingress and egress VXLAN. The system creates an internal SDP binding for the egress processing. A range of SDP IDs has been configured from 10000 to 10127. The following command lists all SDP bindings for FPE:

```
*A:PE-2# show service sdp-using | match "Fpe"
2          10002:2            Fpe    fpe_2.b              Up    524286  524286
```

The internal SDP has ID 10002 and the far-end is fpe_2.b. The following command shows that the SDP source is FPE.

```
*A:PE-2# show service sdp 10002 detail | match "Sdp" pre-lines 4 post-lines 10

===============================================================================
Service Destination Point (Sdp Id : 10002) Details
===============================================================================
-------------------------------------------------------------------------------
 Sdp Id 10002  -fpe_2.b
-------------------------------------------------------------------------------
Description           : (Not Specified)
SDP Id               : 10002                  SDP Source         : fpe
Admin Path MTU       : 0                      Oper Path MTU      : 1552
Delivery             : MPLS
Far End              : fpe_2.b                Tunnel Far End     : n/a
Oper Tunnel Far End  : n/a
LSP Types            : FPE

Admin State          : Up                     Oper State         : Up
```

The following command on PE-1 shows that the source VTEP is 201::1:

```
*A:PE-1# show service id 2 vxlan
===============================================================================
VPLS VXLAN
===============================================================================
Vxlan Src Vtep IP: 201::1

===============================================================================
Vxlan Instance
===============================================================================
VXLAN Instance                VNI         AR          Oper-flags   VTEP
                                                                   security
-------------------------------------------------------------------------------
1                             2           none        none         disabled
-------------------------------------------------------------------------------
Number of Entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

The following command on PE-1 shows that the egress VTEP is 202::.

```
*A:PE-1# show service id 2 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
Instance    VTEP Address                            Egress VNI  EvpnStatic Num
 Mcast       Oper State                              L2 PBR     SupBcasDom MACs
-------------------------------------------------------------------------------
1           202::                                   2           evpn       0
 BUM         Up                                      No          No
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================

===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

The following command on PE-2 shows that the source VTEP is 202::.

```
*A:PE-2# show service id 2 vxlan
===============================================================================
VPLS VXLAN
===============================================================================
Vxlan Src Vtep IP: 202::

===============================================================================
Vxlan Instance
===============================================================================
VXLAN Instance                VNI         AR          Oper-flags   VTEP
                                                                   security
-------------------------------------------------------------------------------
1                             2           none        none         disabled
-------------------------------------------------------------------------------
Number of Entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

The following command on PE-2 shows that the egress VTEP is 201::1.

```
*A:PE-2# show service id 2 vxlan destinations

===============================================================================
Egress VTEP, VNI
===============================================================================
Instance    VTEP Address                            Egress VNI  EvpnStatic Num
 Mcast       Oper State                              L2 PBR     SupBcasDom MACs
-------------------------------------------------------------------------------
1           201::1                                  2           evpn       0
 BUM         Up                                      No          No
-------------------------------------------------------------------------------
Number of Egress VTEP, VNI : 1
-------------------------------------------------------------------------------
===============================================================================
===============================================================================
BGP EVPN-VXLAN Ethernet Segment Dest
===============================================================================
Instance  Eth SegId                       Num. Macs     Last Change
-------------------------------------------------------------------------------
No Matching Entries
===============================================================================
```

Conclusion

VXLAN FPE is required to terminate VXLAN tunnels on non-system IPv4/IPv6 addresses and to configure IPv6 underlay.

Customer document and product support

Customer documentation

Customer documentation welcome page

Technical support

Product support portal

Documentation feedback

Customer documentation feedback