7450 Ethernet Service Switch
7750 Service Router
7950 Extensible Routing System

Releases up to 24.3.R2

Router Configuration Advanced Configuration Guide for MD CLI

3HE 20803 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.

6PE Next-Hop Resolution

This chapter provides information about 6PE next hop resolution.

Topics in this chapter include:

Applicability

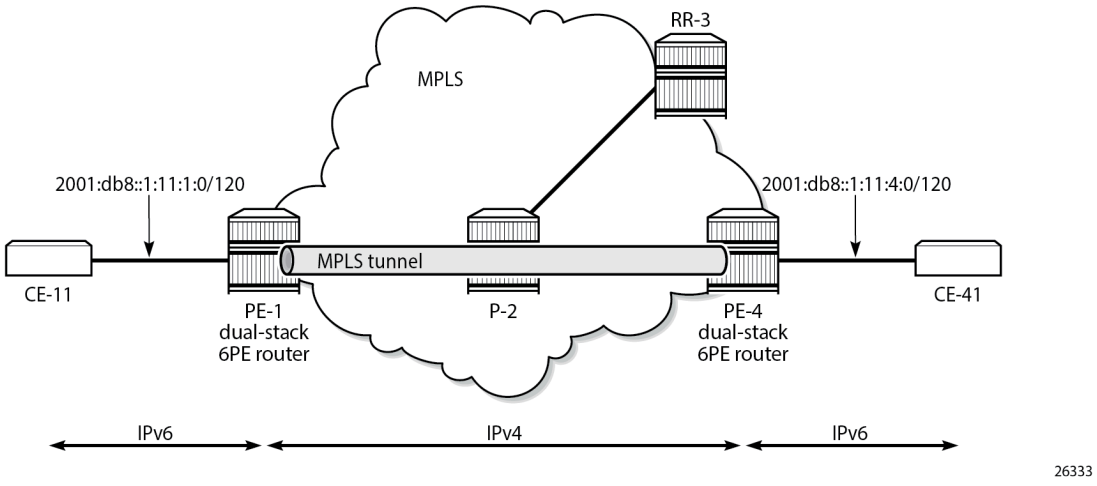
This chapter was initially written based on SR OS Release 14.0.R7, but the MD-CLI in the current edition corresponds to SR OS Release 23.7.R1.

In Releases earlier than 14.0.R1, only label distribution protocol label switched paths (LDP LSPs) could be used to resolve IPv6 provider edge (6PE) next hops. Additional options for 6PE next hop resolution are supported in SR OS Release 14.0.R1, and later. In this chapter, examples are shown with 6PE next hop resolution to different kinds of MPLS tunnels, such as LDP, RSVP-TE, SR-ISIS, and BGP tunnels.

Overview

IPv6 provider edge (6PE) enables IPv6 communication between IPv6 domains over an IPv4 multi-protocol label switching (MPLS) cloud. IPv6 packets are forwarded in an MPLS tunnel from one dual-stack 6PE router to another, as shown in Figure 1: IPv6 provider edge (6PE).

Figure 1: IPv6 provider edge (6PE)



The 6PE route next hop resolution is configured using the following command:

```
*[ex:/configure router "Base" bgp next-hop-resolution labeled-routes transport-tunnel family
 label-ipv6]
A:admin@PE-1#  resolution ?

 resolution <keyword>
 <keyword> - (none|filter|any)
 Default   - filter

    Resolution mode for binding BGP routes to tunnel types
```

With 6PE next hop resolution set to **any**, the tunnels are selected based on availability and tunnel table manager (TTM) preference. The order of preference of TTM tunnels is: RSVP, SR-TE, LDP, SR-OSPF, SR-ISIS, and UDP.

For LDP to be used, it is sufficient to enable LDP on the interfaces in the MPLS network.

For RSVP-TE to be used, an RSVP-TE LSP to the 6PE next-hop destination must be available or configured. For segment routing to be used, an SR-signaled path to the 6PE next hop destination must be available or configured. For BGP labeled routes to be used, the 6PE next hop must have been learned via a BGP peering carrying labeled unicast routes and placed in the active route table.

With 6PE next hop resolution set to filter (default), a subset of protocols is required, and LDP is automatically included in the protocol list in the resolution filter. The following **info** command shows an empty list of protocols when no resolution filter has been defined; the **info detail** command shows that LDP is (implicitly) included.

```
*[ex:/configure router "Base" bgp next-hop-resolution labeled-routes transport-tunnel family
 label-ipv6 resolution-filter]
A:admin@PE-1# info

*[ex:/configure router "Base" bgp next-hop-resolution labeled-routes transport-tunnel family
 label-ipv6 resolution-filter]
A:admin@PE-1# info detail
    bgp false
    ldp true
    rsvp false
    sr-isis false
    sr-ospf false
    sr-te false
    udp false
    sr-policy false
    rib-api false
    mpls-fwd-policy false
    sr-ospf3 false
```

If the 6PE next hop can be resolved to an LDP tunnel, this tunnel is preferred to a BGP tunnel.

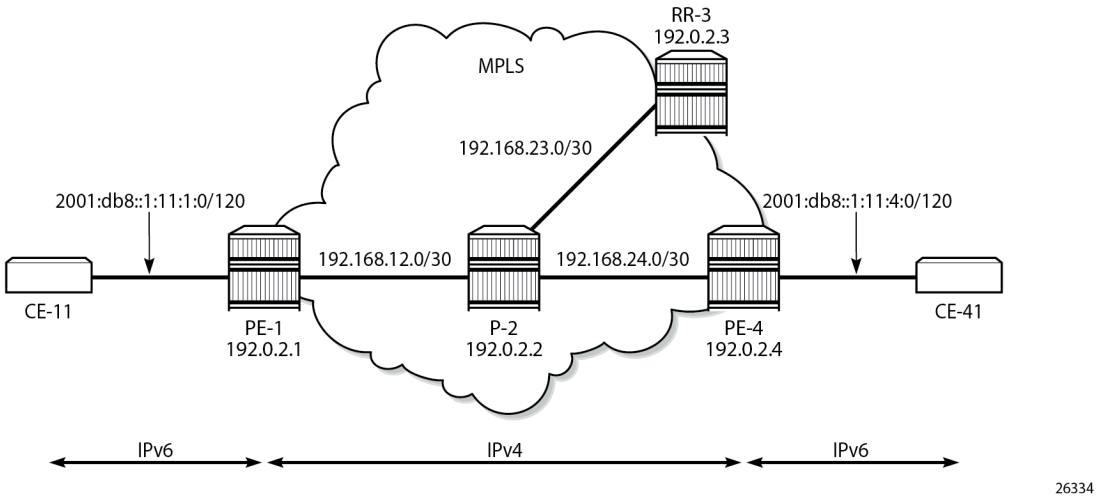
It is possible to explicitly exclude LDP from the list, as follows:

```
*[ex:/configure router "Base" bgp next-hop-resolution labeled-routes transport-tunnel family
 label-ipv6 resolution-filter]
A:admin@PE-1# ldp false

*[ex:/configure router "Base" bgp next-hop-resolution labeled-routes transport-tunnel family
 label-ipv6 resolution-filter]
A:admin@PE-1# info
    bgp true
    ldp false
```

Configuration

Figure 2: Example topology shows the example topology with two dual-stack 6PE routers (PE-1 and PE-4), a core router (P-2), and a route reflector (RR-3). IPv4 is used in the core network; IPv6 is used between the CEs and the PEs.

Figure 2: Example topology



The initial configuration on the nodes is as follows:

- Cards, MDAs, ports
- Router interfaces
- IS-IS as IGP in the core IPv4 network (alternatively, OSPF can be used)
- LDP enabled on the interfaces between the PEs and P-2, but not toward RR-3
- MPLS and RSVP enabled on the interfaces between the PEs and P-2, but not toward RR-3

BGP configuration

BGP is configured for the label-IPv6 address family on PE-1, PE-4, and RR-3, but not on P-2. The BGP configuration on both PEs defines how the 6PE next hops will be resolved: the resolution filter contains three options (LDP, RSVP, and SR-ISIS). The BGP configuration is identical on PE-1 and PE-4.

```
# on PE-1, PE-4:
configure {
    router "Base" {
        autonomous-system 64496
        bgp {
            split-horizon true
            next-hop-resolution {
                labeled-routes {
                    transport-tunnel {
                        family label-ipv6 {
                            resolution-filter {
```

```
                                # ldp true          # default
                                rsvp true
                                sr-isis true
                            }
                            # resolution filter     #default
                        }
                    }
                }
            }
            group "IBGP" {
                peer-as 64496
                export {
                    policy ["export-6pe"]
                }
            }
            neighbor "192.0.2.3" {
                group "IBGP"
                family {
                    label-ipv6 true
                }
            }
```

The export policy "export-6pe" exports the IPv6 prefixes that are local to the PE, for example, on PE-1: 2001:db8::1:11:1:0/120, and is defined as follows:

```
# on PE-1, PE-4:
configure {
    policy-options {
        policy-statement "export-6pe" {
            entry 10 {
                from {
                    protocol {
                        name [direct]
                    }
                }
                action {
                    action-type accept
                }
            }
            default-action {
                action-type reject
            }
        }
```

The BGP configuration on RR-3 does not include any export policy or any next-hop resolution settings, as follows:

```
# on RR-3:
configure {
    router "Base"
        autonomous-system 64496
        bgp {
            split-horizon true
            group "IBGP" {
                peer-as 64496
                cluster {
                    cluster-id 192.0.2.3
                }
            }
            neighbor "192.0.2.1" {
                group "IBGP"
                family {
```

```
                    label-ipv6 true
                }
            }
            neighbor "192.0.2.4" {
                group "IBGP"
                family {
                    label-ipv6 true
                }
            }
```

IES configuration

On PE-1, an IES is configured with IPv6 addresses on the interface toward CE-11, as follows:

```
# on PE-1:
configure {
    service {
        ies "IES-1" {
            admin-state enable
            description "6PE"
            service-id 1
            customer "1"
            interface "int-PE-1-CE-11" {
                sap 1/1/c3/1:1 {
                }
                ipv6 {
                    address 2001:db8::1:11:1:1 {
                        prefix-length 120
                    }
                }
            }
        }
```

The configuration on PE-4 is similar; the IPv6 address on interface "int-PE-4-CE-41" is different: 2001:db8::1:11:4:1/120.

A BGP labeled IPv6 tunnel, which is active in the IPv6 routing table, is established between the PEs, as follows:

```
[/]
A:admin@PE-1# show router route-table ipv6

===============================================================================
IPv6 Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
2001:db8::1:11:1:0/120                        Local   Local     00h04m06s  0
       int-PE-1-CE-11                                             0
2001:db8::1:11:4:0/120                        Remote  BGP_LABEL 00h03m56s  170
       192.0.2.4 (tunneled)                                       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
===============================================================================
```

CE-11 can send IPv6 packets with source address 2001:db8::1:11:1:11 to destination address 2001:db8::1:11:4:41 on CE-41, as follows:

```
[/]
A:admin@PE-1# ping 2001:db8::1:11:4:41 router-instance "CE-11" source-address
 2001:db8::1:11:1:11
PING 2001:db8::1:11:4:41 56 data bytes
64 bytes from 2001:db8::1:11:4:41 icmp_seq=1 hlim=62 time=8.79ms.
64 bytes from 2001:db8::1:11:4:41 icmp_seq=2 hlim=62 time=3.69ms.
64 bytes from 2001:db8::1:11:4:41 icmp_seq=3 hlim=62 time=3.47ms.
64 bytes from 2001:db8::1:11:4:41 icmp_seq=4 hlim=62 time=3.65ms.
64 bytes from 2001:db8::1:11:4:41 icmp_seq=5 hlim=62 time=2.60ms.

---- 2001:db8::1:11:4:41 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 2.60ms, avg = 4.44ms, max = 8.79ms, stddev = 2.21ms
```

6PE next hop resolved to an LDP tunnel

On PE-1, the route for prefix 2001:db8::1:11:4:0/120 uses a tunnel to 6PE next hop 192.0.2.4, as follows:

```
[/]
A:admin@PE-1# show router route-table 2001:db8::1:11:4:0/120

===============================================================================
IPv6 Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
2001:db8::1:11:4:0/120                        Remote  BGP_LABEL 00h04m17s  170
       192.0.2.4 (tunneled)                                        20
-------------------------------------------------------------------------------
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
===============================================================================
```

LDP is enabled on the interfaces between the PEs and P-2, which is sufficient for 6PE next hop resolution to an LDP tunnel. RSVP-TE tunnels have a higher priority, but no MPLS LSPs have been configured yet on the PEs. The tunnel table on PE-1 shows that the only tunnel to 6PE next hop 192.0.2.4 is an LDP tunnel, as follows:

```
[/]
A:admin@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
192.0.2.4/32          ldp       MPLS  65538     9      192.168.12.2   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
===============================================================================
```

Alternatively, the following show command can be used: the only tunnel on slot 1 (card 1) to 6PE next hop 192.0.2.4 is an LDP tunnel:

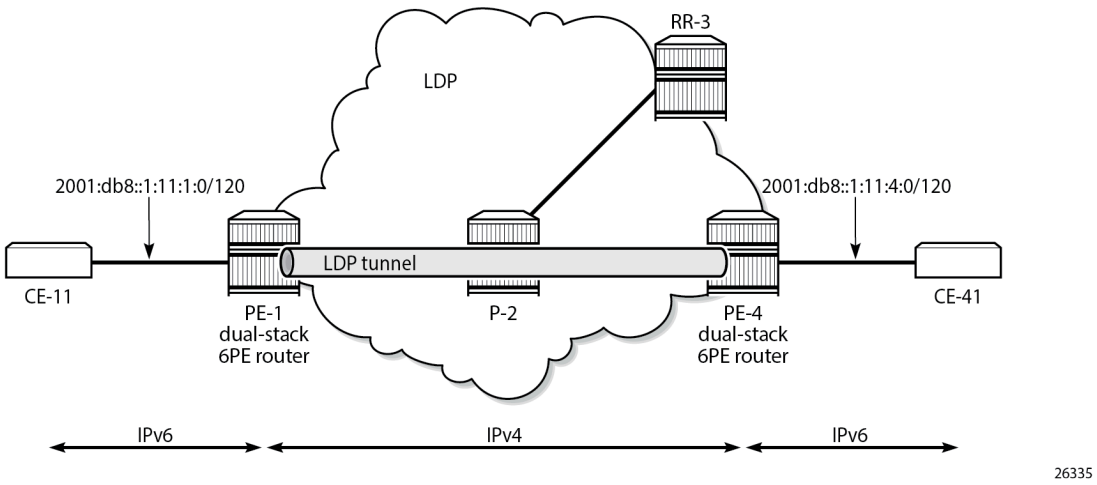
```
[/]
A:admin@PE-1# show router fp-tunnel-table 1 192.0.2.4/32

===============================================================================
IPv4 Tunnel Table Display

Legend:
label stack is ordered from bottom-most to top-most
B - FRR Backup
===============================================================================
Destination                                Protocol         Tunnel-ID
  Lbl/SID
    NextHop                                               Intf/Tunnel
  Lbl/SID (backup)
    NextHop   (backup)
-------------------------------------------------------------------------------
192.0.2.4/32                               LDP               -
  524286
    192.168.12.2                                         1/1/c1/1:1000
-------------------------------------------------------------------------------
Total Entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

The extended route information for IPv6 prefix 2001:db8::1:11:4:0/120 shows that the 6PE next hop 192.0.2.4 is resolved to an LDP tunnel:

```
[/]
A:admin@PE-1# show router route-table 2001:db8::1:11:4:0/120 extensive

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix             : 2001:db8::1:11:4:0/120
  Protocol              : BGP_LABEL
  Age                   : 00h05m42s
  Preference            : 170
  Indirect Next-Hop     : 192.0.2.4
    Label               : 2
    QoS                 : Priority=n/c, FC=n/c
    Source-Class        : 0
    Dest-Class          : 0
    ECMP-Weight         : N/A
    Resolving Next-Hop  : 192.0.2.4 (LDP tunnel)
      Metric            : 20
      ECMP-Weight       : N/A
-------------------------------------------------------------------------------
No. of Destinations: 1
===============================================================================
```

Figure 3: 6PE next hop resolved to an LDP tunnel shows that the 6PE next hop is resolved to an LDP tunnel. No other tunnels are available in the IPv4 core network.

Figure 3: 6PE next hop resolved to an LDP tunnel



6PE next hop resolved to an RSVP-TE tunnel

MPLS and RSVP are enabled on the interfaces between the PEs and P-2. On both PEs, an RSVP-TE LSP is configured toward the peer PE; for example, on PE-1:

```
# on PE-1:
configure {
    router Base
        mpls {
            path "empty" {
                admin-state enable
            }
            lsp "LSP-PE-1-PE-4" {
                admin-state enable
                type p2p-rsvp
                to 192.0.2.4
                primary "empty" {
                }
            }
```

The configuration is similar on PE-4. No additional configuration is required on P-2.

The following output shows that two tunnels are available to 6PE next hop 192.0.2.4/32: an LDP tunnel and an RSVP-TE tunnel:

```
[/]
A:admin@PE-1# show router fp-tunnel-table 1 192.0.2.4/32

===============================================================================
IPv4 Tunnel Table Display

Legend:
label stack is ordered from bottom-most to top-most
B - FRR Backup
===============================================================================
Destination                                Protocol         Tunnel-ID
  Lbl/SID
```

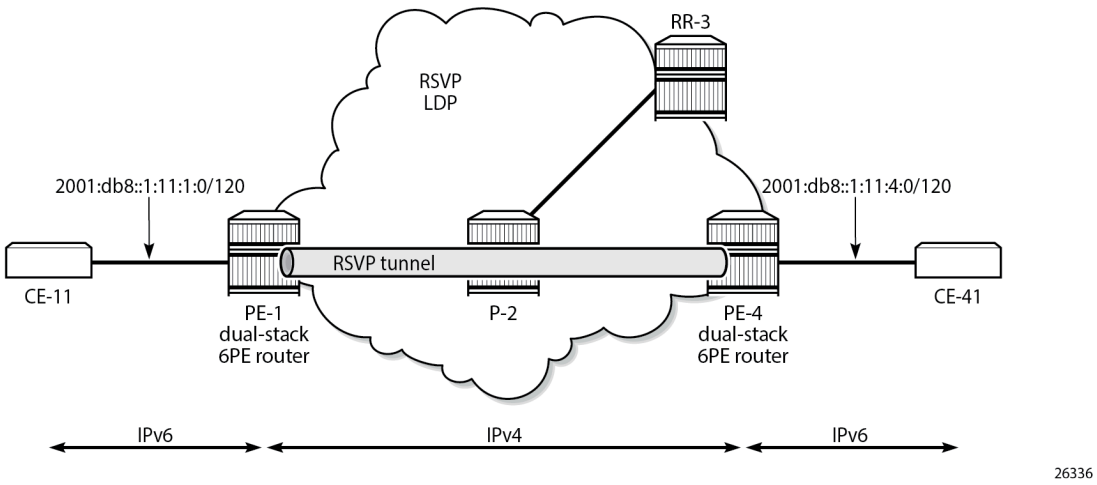
```
    NextHop                                                 Intf/Tunnel
  Lbl/SID (backup)
    NextHop   (backup)
-------------------------------------------------------------------------------
192.0.2.4/32                                LDP               -
  524286
    192.168.12.2                                           1/1/c1/1:1000
192.0.2.4/32                                RSVP              1
  524284
    192.168.12.2                                           1/1/c1/1:1000
-------------------------------------------------------------------------------
Total Entries : 2
-------------------------------------------------------------------------------
===============================================================================
```

For 6PE next hop resolution, RSVP-TE tunnels are preferred to any other tunnel type in the tunnel table, so the BGP next hop 192.0.2.4 will be resolved to an RSVP-TE tunnel, as follows:

```
[/]
A:admin@PE-1# show router route-table 2001:db8::1:11:4:0/120 extensive

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix             : 2001:db8::1:11:4:0/120
  Protocol              : BGP_LABEL
  Age                   : 00h00m43s
  Preference            : 170
  Indirect Next-Hop     : 192.0.2.4
    Label               : 2
    QoS                 : Priority=n/c, FC=n/c
    Source-Class        : 0
    Dest-Class          : 0
    ECMP-Weight         : N/A
    Resolving Next-Hop  : 192.0.2.4 (RSVP tunnel:1)
      Metric            : 20
      ECMP-Weight       : N/A
-------------------------------------------------------------------------------
No. of Destinations: 1
===============================================================================
```

Figure 4: 6PE next hop resolved to an RSVP-TE tunnel shows that the 6PE next hop 192.0.2.4 is resolved to an RSVP-TE tunnel, even though an LDP tunnel is available too.

Figure 4: 6PE next hop resolved to an RSVP-TE tunnel



6PE next hop resolved to an SR-ISIS tunnel

Segment routing is enabled for IS-IS on PE-1, P-2, and PE-4. The configuration is similar on each of these nodes; the only difference is the IPv4 node SID index on the system interface. The SR-ISIS configuration on PE-1 is as follows:

```
# on PE-1:
configure {
    router "Base" {
        mpls-labels {
            sr-labels {
                start 20000
                end 20099
            }
        }
        isis 0 {
            advertise-router-capability area
            segment-routing {
                admin-state enable
                prefix-sid-range {
                    start-label 20000
                    max-index 99
                }
            }
            interface "system" {
                ipv4-node-sid {
                    index 1
                }
            }
        }
```

For more information about SR-ISIS, see the "Segment Routing with IS-IS Control Plane" in *7750 SR and 7950 XRS Segment Routing and PCE Advanced Configuration Guide for MD CLI* chapter.

The following output shows that three tunnels are available toward 6PE next hop 192.0.2.4/32:

```
[/]
```

```
A:admin@PE-1# show router fp-tunnel-table 1 192.0.2.4/32

===============================================================================
IPv4 Tunnel Table Display

Legend:
label stack is ordered from bottom-most to top-most
B - FRR Backup
===============================================================================
Destination                                Protocol         Tunnel-ID
  Lbl/SID
    NextHop                                               Intf/Tunnel
  Lbl/SID (backup)
    NextHop   (backup)
-------------------------------------------------------------------------------
192.0.2.4/32                               LDP               -
  524286
    192.168.12.2                                         1/1/c1/1:1000
192.0.2.4/32                               RSVP              1
  524284
    192.168.12.2                                         1/1/c1/1:1000
192.0.2.4/32                               SR-ISIS-0         524291
  20004
    192.168.12.2                                         1/1/c1/1:1000
-------------------------------------------------------------------------------
Total Entries : 3
-------------------------------------------------------------------------------
===============================================================================
```

RSVP-TE tunnels are preferred; therefore, the 6PE next hop 192.0.2.4 is resolved to the RSVP-TE tunnel, as follows:

```
[/]
A:admin@PE-1# show router route-table 2001:db8::1:11:4:0/120 extensive

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix             : 2001:db8::1:11:4:0/120
  Protocol              : BGP_LABEL
  Age                   : 00h02m12s
  Preference            : 170
  Indirect Next-Hop     : 192.0.2.4
    Label               : 2
    QoS                 : Priority=n/c, FC=n/c
    Source-Class        : 0
    Dest-Class          : 0
    ECMP-Weight         : N/A
    Resolving Next-Hop  : 192.0.2.4 (RSVP tunnel:1)
      Metric            : 20
      ECMP-Weight       : N/A
-------------------------------------------------------------------------------
No. of Destinations: 1
===============================================================================
```

To verify that LDP tunnels are preferred over SR-ISIS tunnels, the RSVP-TE LSPs are disabled, as follows:

```
# on PE-1:
configure {
    router "Base"
        mpls {
            lsp "LSP-PE-1-PE-4" {
```

```
                admin-state disable
            }
```

The following output shows that two tunnels are available toward 6PE next hop 192.0.2.4/32: an LDP tunnel and an SR-ISIS tunnel.

```
[/]
A:admin@PE-1# show router fp-tunnel-table 1 192.0.2.4/32

================================================================================
IPv4 Tunnel Table Display

Legend:
label stack is ordered from bottom-most to top-most
B - FRR Backup
================================================================================
Destination                                Protocol         Tunnel-ID
  Lbl/SID
    NextHop                                               Intf/Tunnel
  Lbl/SID (backup)
    NextHop   (backup)
--------------------------------------------------------------------------------
192.0.2.4/32                               LDP               -
  524286
    192.168.12.2                                         1/1/c1/1:1000
192.0.2.4/32                               SR-ISIS-0         524291
  20004
    192.168.12.2                                         1/1/c1/1:1000
--------------------------------------------------------------------------------
Total Entries : 2
--------------------------------------------------------------------------------
================================================================================
```

For 6PE next-hop resolution, the LDP tunnel is preferred over the SR-ISIS tunnel, as follows:

```
[/]
A:admin@PE-1# show router route-table 2001:db8::1:11:4:0/120 extensive

================================================================================
Route Table (Router: Base)
================================================================================
Dest Prefix             : 2001:db8::1:11:4:0/120
  Protocol              : BGP_LABEL
  Age                   : 00h00m31s
  Preference            : 170
  Indirect Next-Hop     : 192.0.2.4
    Label               : 2
    QoS                 : Priority=n/c, FC=n/c
    Source-Class        : 0
    Dest-Class          : 0
    ECMP-Weight         : N/A
    Resolving Next-Hop  : 192.0.2.4 (LDP tunnel)
      Metric            : 20
      ECMP-Weight       : N/A
--------------------------------------------------------------------------------
No. of Destinations: 1
================================================================================
```

When LDP is disabled on interface "int-PE-1-P-2" on PE-1, the only remaining tunnel is an SR-ISIS tunnel, as follows:

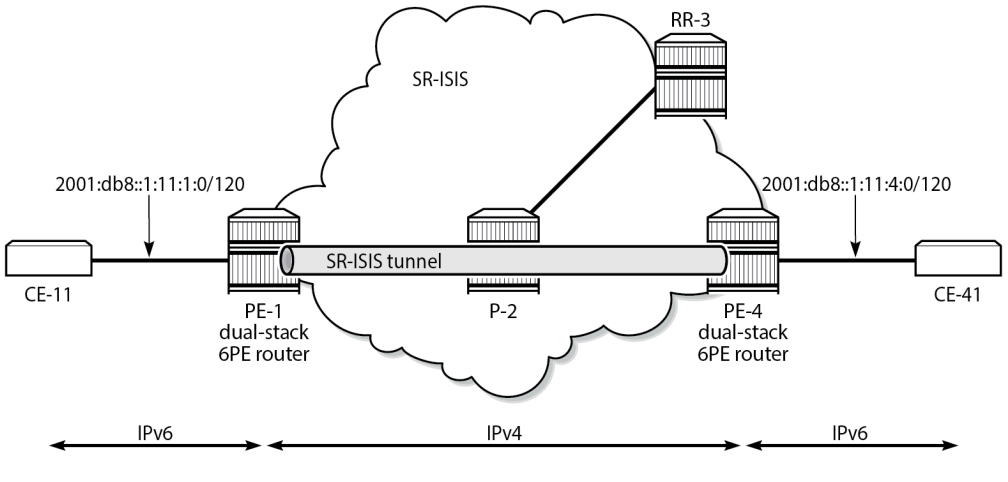
```
# on PE-1:
configure {
    router "Base" {
        ldp {
            interface-parameters {
                interface "int-PE-1-P-2" {
                    admin-state disable
                }
            }

[/]
A:admin@PE-1# show router fp-tunnel-table 1 192.0.2.4/32

===============================================================================
IPv4 Tunnel Table Display

Legend:
label stack is ordered from bottom-most to top-most
B - FRR Backup
===============================================================================
Destination                                Protocol         Tunnel-ID
  Lbl/SID
    NextHop                                               Intf/Tunnel
  Lbl/SID (backup)
    NextHop   (backup)
-------------------------------------------------------------------------------
192.0.2.4/32                               SR-ISIS-0         524291
  20004
    192.168.12.2                                         1/1/c1/1:1000
-------------------------------------------------------------------------------
Total Entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

The 6PE next hop 192.0.2.4 is resolved to an SR-ISIS tunnel, as follows:

```
[/]
A:admin@PE-1# show router route-table 2001:db8::1:11:4:0/120 extensive

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix             : 2001:db8::1:11:4:0/120
  Protocol              : BGP_LABEL
  Age                   : 00h00m55s
  Preference            : 170
  Indirect Next-Hop     : 192.0.2.4
    Label               : 2
    QoS                 : Priority=n/c, FC=n/c
    Source-Class        : 0
    Dest-Class          : 0
    ECMP-Weight         : N/A
    Resolving Next-Hop  : 192.0.2.4 (SR-ISIS tunnel:524291)
      Metric            : 20
      ECMP-Weight       : N/A
-------------------------------------------------------------------------------
No. of Destinations: 1
===============================================================================
```

Figure 5: 6PE next hop resolved to an SR-ISIS tunnel shows that the 6PE next hop 192.0.2.4 is resolved to an SR-ISIS tunnel after the RSVP-TE LSPs are disabled and LDP is disabled on the interfaces between the PEs and P-2. No other tunnels are available.

Figure 5: 6PE next hop resolved to an SR-ISIS tunnel



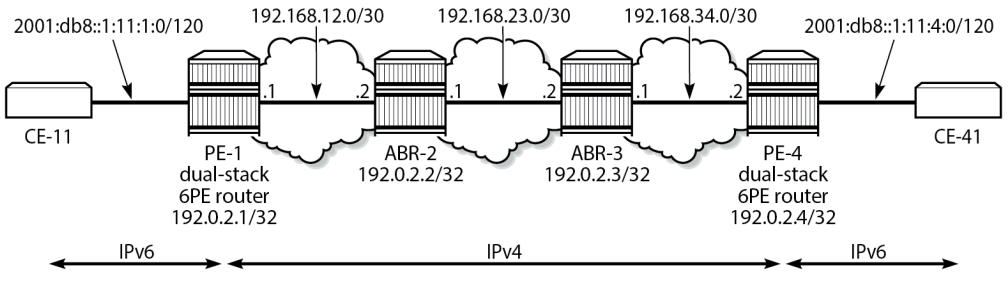
6PE next-hop resolution to a BGP IPv4 tunnel

The preceding example cannot be extended with BGP labeled IPv4 tunnels. The reason is that for BGP to work, some underlying MPLS signaling protocol is required, such as RSVP-TE or LDP. Because BGP tunnels have a very low preference, they will not be used when an LDP or RSVP-TE tunnel is available to the 6PE next hop.

This section shows a seamless MPLS example where 6PE next hops are resolved to BGP labeled IPv4 routes, because no LDP tunnel is available to the 6PE next hop in a different IGP topology (in this example, LDP is configured, not RSVP-TE). For a description of this seamless MPLS implementation, see the "Seamless MPLS: Isolated IGP/LDP Domains and Labeled BGP" chapter in the *7450 ESS, 7750 SR, and 7950 XRS MPLS Advanced Configuration Guide for MD CLI*.

Figure 6: Example topology for seamless MPLS shows the example topology for seamless MPLS with two aggregation networks and one core network.

Figure 6: Example topology for seamless MPLS

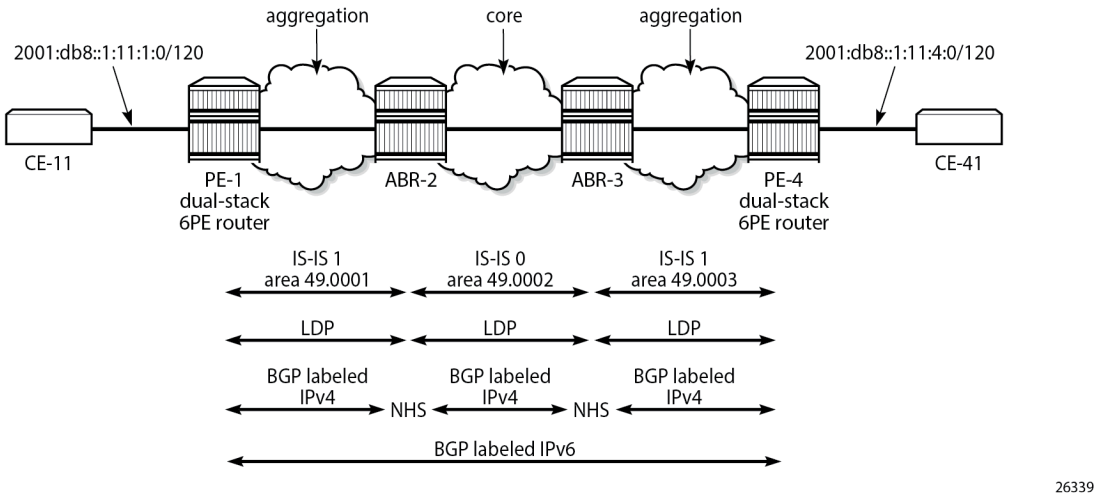


Different IS-IS instances are configured: IS-IS instance 0 is configured in the core, whereas IS-IS instance 1 is configured in the aggregation networks. On the area border routers (ABRs) ABR-2 and ABR-3, two instances of IS-IS are configured: IS-IS instance 0 for the core and IS-IS instance 1 for the aggregation network. PE-1 and PE-4 will only learn routes to destinations within their respective aggregation networks; ABRs learn routes within one aggregation network and the core network. LDP is configured on all interfaces, but PE-1 will not have an LDP binding for prefix 192.0.2.4/32, as shown in the following output. Therefore, 6PE next hop 192.0.2.4 cannot be resolved to an LDP tunnel.

```
[/]
A:admin@PE-1# show router ldp bindings active prefixes 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
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op
IngLbl                                  EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.1/32                            Pop
524287                                    --
  --                                      --

192.0.2.2/32                            Push
  --                                    524287
192.168.12.2                            1/1/c1/1:1000

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
```

Figure 7: Configured protocols for seamless MPLS shows the configured protocols for this example: IS-IS instances, LDP, BGP labeled IPv4 with the ABRs as route reflector with **next-hop-self** (NHS) option, and BGP labeled IPv6 peering between PE-1 and PE-4.

Figure 7: Configured protocols for seamless MPLS



The following initial configuration on ABR-2 includes two IS-IS instances in different areas. IS-IS instance 0 with area ID 49.0002 is configured in the core network; IS-IS instance 1 with area ID 49.0001 is configured in the aggregation network between PE-1 and ABR-2. LDP is configured on each router interface.

```
# on ABR-2:
configure {
    router "Base" {
        interface "int-ABR-2-ABR-3" {
            port 1/1/c3/1:1000
            ipv4 {
                primary {
                    address 192.168.23.1
                    prefix-length 30
                }
            }
        }
        interface "int-ABR-2-PE-1" {
            port 1/1/c2/1:1000
            ipv4 {
                primary {
                    address 192.168.12.2
                    prefix-length 30
                }
            }
        }
        interface "system" {
            ipv4 {
                primary {
                    address 192.0.2.2
                    prefix-length 32
                }
            }
        }
        isis 0 {
            admin-state enable
            level-capability 2
            area-address [49.0002]
            interface "int-ABR-2-ABR-3" {
                interface-type point-to-point
```

```
            }
            interface "system" {
            }
        }
        isis 1 {
            admin-state enable
            level-capability 2
            area-address [49.0001]
            interface "int-ABR-2-PE-1" {
                interface-type point-to-point
            }
            interface "system" {
            }
        }
        ldp {
            interface-parameters {
                interface "int-ABR-2-ABR-3" {
                    ipv4 {
                    }
                }
                interface "int-ABR-2-PE-1" {
                    ipv4 {
                    }
                }
            }
        }
```

The configuration is similar on the other nodes. Only the ABRs have two IS-IS instances configured; the PEs only have one IS-IS instance.

BGP needs to be configured for the label-IPv4 and label-IPv6 address families:

- The label-IPv4 address family is used with the ABRs as RR in the aggregation network. Each ABR is configured with the **next-hop-self** option. BGP label-IPv4 peering is between the ABRs without RR.
- The label-IPv6 address family is used between PE-1 and PE-4. The BGP session can only be established after the BGP labeled IPv4 routes have been exchanged between PE-1 and PE-4.

BGP is configured on PE-1 as follows:

```
# on PE-1:
configure {
    router "Base" {
        autonomous-system 64496
        bgp {
            split-horizon true
            next-hop-resolution {
                labeled-routes {
                    transport-tunnel {
                        family label-ipv6 {
                            resolution-filter {
                                bgp true
                                # ldp true    # LDP is by default included
                            }
                        }
                    }
                }
            }
            group "IBGPv4" {
                peer-as 64496
                export {
                    policy ["export-sys"]
                }
            }
```

```
            group "IBGPv6" {
                peer-as 64496
                export {
                    policy ["export-6pe"]
                }
            }
            neighbor "192.0.2.2" {
                group "IBGPv4"
                family {
                    label-ipv4 true
                }
            }
            neighbor "192.0.2.4" {
                group "IBGPv6"
                family {
                    label-ipv6 true
                }
            }
        }
```

The configuration is similar on PE-4, but the neighbor IP addresses are different.

The resolution filter will include LDP as well as BGP, because it is added automatically. However, no LDP tunnel will be available from PE-1 to PE-4, or vice versa; therefore, BGP labeled IPv4 will be used.

The "export-sys" policy exports the IPv4 system address of the PE and is defined as follows:

```
# on PE-1, PE-4:
configure {
    policy-options {
        prefix-list "system" {
            prefix 192.0.2.0/24 type longer {
            }
        }
        policy-statement "export-sys" {
            entry 10 {
                from {
                    prefix-list ["system"]
                    protocol {
                        name [direct]
                    }
                }
                action {
                    action-type accept
                }
            }
            default-action {
                action-type reject
            }
        }
```

The "export-6pe" policy exports the local labeled IPv6 routes and is the same in the preceding examples:

```
# on PE-1, PE-4:
configure {
    policy-options {
        policy-statement "export-6pe" {
            entry 10 {
                from {
                    protocol {
                        name [direct]
                    }
                }
```

```
                action {
                    action-type accept
                }
            }
            default-action {
                action-type reject
            }
        }
```

The BGP configuration on ABR-2 has two different groups for BGP labeled IPv4 peering: one toward the aggregation network—with the ABR as RR—and one toward the core, as follows:

```
# on ABR-2:
configure {
    router "Base" {
        autonomous-system 64496
        bgp {
            advertise-inactive true
            split-horizon true
            group "IBGPv4-agg" {
                next-hop-self true
                peer-as 64496
                cluster {
                    cluster-id 192.0.2.2
                }
            }
            group "IBGPv4-core" {
                next-hop-self true
                peer-as 64496
            }
            neighbor "192.0.2.1" {
                group "IBGPv4-agg"
                family {
                    label-ipv4 true
                }
            }
            neighbor "192.0.2.3" {
                group "IBGPv4-core"
                family {
                    label-ipv4 true
                }
            }
        }
```

The configuration is similar on ABR-3, but the neighbor IP addresses and the cluster ID are different.

The ABRs are configured with the **next-hop-self** option for both groups. The 6PE next hop 192.0.2.4 will have next hop ABR-2 on PE-1, which can be resolved to an LDP tunnel. On ABR-2, 6PE next hop 192.0.2.4 will have ABR-3 as next hop, which can be resolved to an LDP tunnel. On ABR-3, the 6PE next hop 192.0.2.4 can be resolved to an LDP tunnel (no active BGP route to 192.0.2.4/32 on ABR-3 because the route via IS-IS is preferred).

The **advertise-inactive** option is required for ABR-2 to export a BGP route for prefix 192.0.2.1/32, which is not active on ABR-2, because an IS-IS route is available for this prefix and IS-IS routes are preferred over BGP routes.

The IES configuration is the same as in the preceding example.

When the labeled IPv4 routes are exchanged between PE-1 and PE-4, the BGP labeled session using IPv6 peering can be established between PE-1 and PE-4, as follows:

```
[/]
```

```
A:admin@PE-1# 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
-------------------------------------------------------------------------------
192.0.2.2
Def. Inst       64496      12    0 00h03m34s 1/1/1 (Lbl-IPv4)
                           12    0
192.0.2.4
Def. Inst       64496       8    0 00h01m47s 1/1/1 (Lbl-IPv6)
                            8    0

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

For IPv6 prefix 2001:db8::1:11:4:0/120 on PE-1, 6PE next hop 192.0.2.4 is resolved to a BGP tunnel, as follows:

```
[/]
A:admin@PE-1# show router route-table 2001:db8::1:11:4:0/120 extensive

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix             : 2001:db8::1:11:4:0/120
  Protocol              : BGP_LABEL
  Age                   : 00h01m40s
  Preference            : 170
  Indirect Next-Hop     : 192.0.2.4
    Label               : 2
    QoS                 : Priority=n/c, FC=n/c
    Source-Class        : 0
    Dest-Class          : 0
    ECMP-Weight         : N/A
    Resolving Next-Hop  : 192.0.2.4 (BGP tunnel)
      Metric            : 1000
      ECMP-Weight       : N/A
-------------------------------------------------------------------------------
No. of Destinations: 1
===============================================================================
```

The BGP labeled IPv4 route to 192.0.2.4 has different next hops in different nodes, because both ABRs set the **next-hop-self** option. On PE-1, the BGP labeled IPv4 route for prefix 192.0.2.4 has next hop 192.0.2.2 and uses an LDP tunnel to reach ABR-2 within the aggregation network, as follows:

```
[/]
A:admin@PE-1# show router fp-tunnel-table 1 192.0.2.4/32

===============================================================================
IPv4 Tunnel Table Display

Legend:
label stack is ordered from bottom-most to top-most
B - FRR Backup
===============================================================================
Destination                                 Protocol         Tunnel-ID
```

```
  Lbl/SID
    NextHop                                                  Intf/Tunnel
  Lbl/SID (backup)
    NextHop   (backup)
-------------------------------------------------------------------------------
192.0.2.4/32                                 BGP              -
  524282
    192.0.2.2                                                LDP
-------------------------------------------------------------------------------
Total Entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

On ABR-2, the BGP labeled route to 192.0.2.4/32 has next hop 192.0.2.3 and uses an LDP tunnel in the core network to reach ABR-3, as follows:

```
[/]
A:admin@ABR-2# show router fp-tunnel-table 1 192.0.2.4/32

===============================================================================
IPv4 Tunnel Table Display

Legend:
label stack is ordered from bottom-most to top-most
B - FRR Backup
===============================================================================
Destination                                  Protocol         Tunnel-ID
  Lbl/SID
    NextHop                                                  Intf/Tunnel
  Lbl/SID (backup)
    NextHop   (backup)
-------------------------------------------------------------------------------
192.0.2.4/32                                 BGP              -
  524282
    192.0.2.3                                                LDP
-------------------------------------------------------------------------------
Total Entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

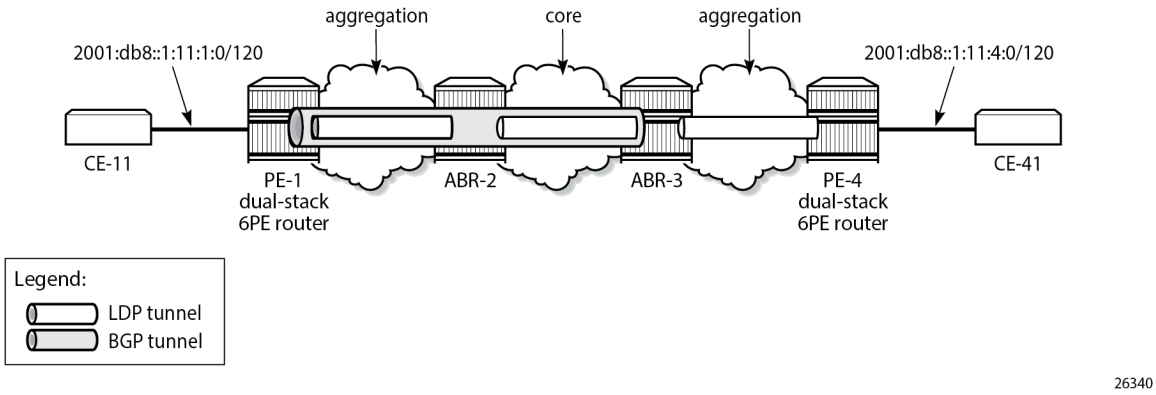
On ABR-3, no BGP labeled IPv4 route is active for prefix 192.0.2.4 because IS-IS routes are preferred to BGP routes. An LDP tunnel is used toward PE-4 in the aggregation network, as follows:

```
[/]
A:admin@ABR-3# show router fp-tunnel-table 1 192.0.2.4/32

===============================================================================
IPv4 Tunnel Table Display

Legend:
label stack is ordered from bottom-most to top-most
B - FRR Backup
===============================================================================
Destination                                  Protocol         Tunnel-ID
  Lbl/SID
    NextHop                                                  Intf/Tunnel
  Lbl/SID (backup)
    NextHop   (backup)
-------------------------------------------------------------------------------
192.0.2.4/32                                 LDP              -
  524287
    192.168.34.2                                             1/1/c1/1:1000
```

```
-------------------------------------------------------------------------------
Total Entries : 1
-------------------------------------------------------------------------------
===============================================================================
```

Figure 8: BGP labeled IPv4 tunnel for 192.0.2.4/32 using LDP tunnels shows the BGP and LDP tunnels used for 6PE next hop 192.0.2.4/32.

Figure 8: BGP labeled IPv4 tunnel for 192.0.2.4/32 using LDP tunnels



Conclusion

The 6PE next hops can be resolved to different types of MPLS tunnels, each with a different preference.

Aggregate Route Indirect Next-Hop Option

This chapter provides information about aggregate routes with indirect next-hop option.

Topics in this chapter include:

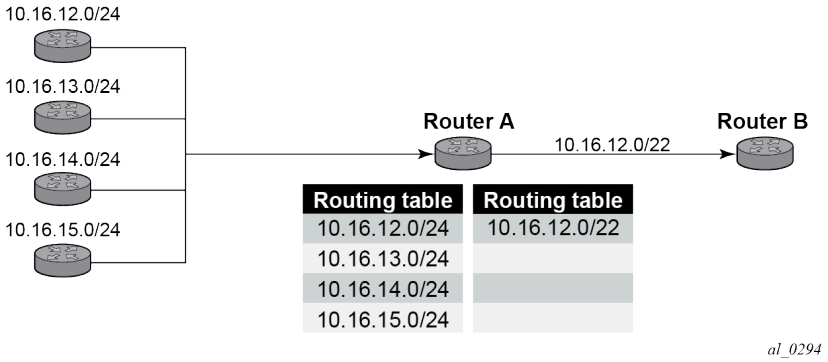
Applicability

This chapter was initially written based on SR OS Release 11.0.R1. The MD-CLI in the current edition corresponds to SR OS Release 22.10.R1.

Overview

In SR OS nodes, IPv4 and IPv6 aggregate routes can be configured. A configured aggregate route that has the best preference for the prefix is activated, and therefore, added to the routing table, when it has at least one contributing route; the aggregate route is removed from the routing table when there are no longer any contributing routes. A contributing route is any route installed in the forwarding table that is a more specific match of the aggregate. For example, the route 10.16.12.0/24 is a contributing route to the aggregate route 10.16.12.0/22, but for this same aggregate, the routes 10.16.0.0/16 and 10.0.0.0/8 are not contributing routes.

Figure 9: Aggregate routes



In Figure 9: Aggregate routes , Router A can advertise all four routes or one aggregate route. By aggregating the four routes, fewer updates are sent on the link between routers A and B, router B needs to maintain a smaller routing table resulting in better convergence and router B saves on computational resources by evaluating fewer entries in its routing table.

It is possible to configure an indirect hop for aggregate routes. The indirect next hop specifies where packets will be forwarded if they match the aggregate route, but not a more specific route in the IP forwarding table.

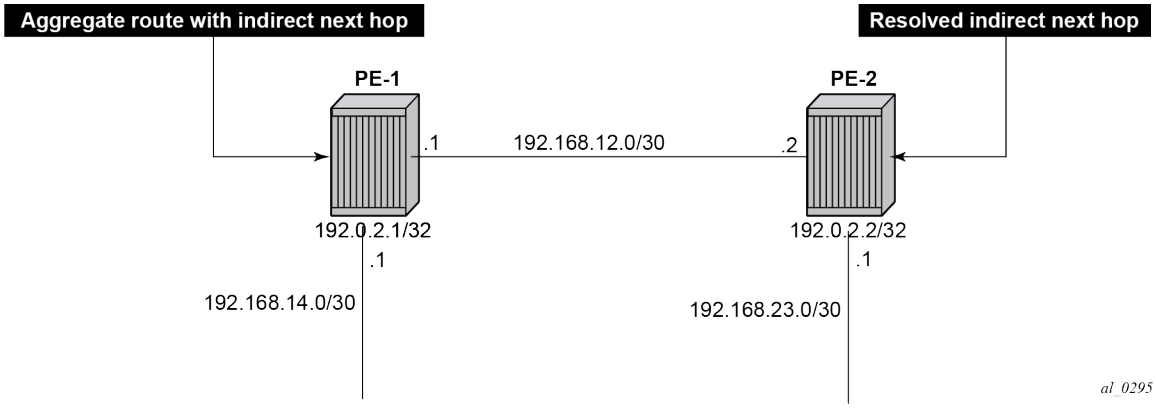
Different network operators have different requirements on how to forward a packet that matches an aggregate route but not any of the more specific routes in the forwarding table that activated the aggregate. In general, there are three different options:

1. The packet can be forwarded according to the next-most specific route, ignoring the aggregate route. This can lead to routing loops in some topologies.
2. The packet can be discarded.
3. The packet can be forwarded toward an indirect next-hop address that is configured by the operator. The indirect next-hop could be the address of a threat management server that analyzes the packets it receives for security threats. This option requires the aggregate route to be installed in the forwarding table with a resolved next-hop interface determined from a route lookup of the indirect next-hop address.

Configuration

The example topology with two PEs is shown in Figure 10: Example topology.

Figure 10: Example topology



Initial configuration

The nodes have the following basic configuration:

- cards, MDAs
- ports
- router interfaces

The router interfaces on PE-1 are configured as follows:

```
# on PE-1:
configure {
    router "Base" {
```

```
        interface "int-PE-1-PE-2" {
            port 1/1/c1/1:1000
            ipv4 {
                primary {
                    address 192.168.12.1
                    prefix-length 30
                }
            }
        }
        interface "int-PE-1-PE-4" {
            port 1/1/c2/1:1000
            ipv4 {
                primary {
                    address 192.168.14.1
                    prefix-length 30
                }
            }
        }
        interface "system" {
            ipv4 {
                primary {
                    address 192.0.2.1
                    prefix-length 32
                }
            }
        }
```

The configuration on PE-2 is similar. The IP addresses are shown in Figure 10: Example topology. In this example, static routes are configured. There is no need for an IGP, but it could be configured.

Aggregate route with indirect next hop option

This feature adds the **indirect** keyword and an associated IP address parameter to the **aggregate** command in the configuration contexts of the base router and of VPRN services.

The aggregate route configuration command in the base router context is as follows:

```
*[ex:/configure router "Base" aggregates aggregate 10.16.12.0/22]
A:admin@PE-1# ?

 Immutable fields      - indirect

 aggregator            + Enter the aggregator context
 apply-groups          - Apply a configuration group at this level
 apply-groups-exclude  - Exclude a configuration group at this level
 as-set                - Use AS_SET path segment type for the aggregate route
 community             - Community name that is added to the aggregate route
 description           - Text description
 discard-component-    - Advertise aggregate with aggregate route community set
  communities
 local-preference      - Local preference used when aggregate route is exported
 policy                - Policy name for the aggregated route
 summary-only          - Advertise the aggregate route only
 tunnel-group          - Tunnel group from which to associate the MC IPSec state

 Choice: next-hop
 blackhole            :+ Enable the blackhole context
 indirect             :- Address of the indirect next hop
```

Parameters:

- **indirect** — This indicates that the aggregate route has an indirect address. The indirect option is mutually exclusive with the black-hole option.
- <ip-address> — Installing an aggregate route with an indirect next-hop is supported for both IPv4 and IPv6 prefixes. However, if the aggregate prefix is IPv6, the indirect next-hop must be an IPv6 address and if the aggregate prefix is IPv4, the indirect next-hop must be an IPv4 address.

If an indirect next-hop is not resolved, the aggregate route will show up as black-hole.

The aggregate route 10.16.12.0/22 is configured as follows:

```
# on PE-1:
configure {
    router "Base" {
        aggregates {
            aggregate 10.16.12.0/22 {
                community ["64496:64498"]
                indirect 192.168.11.11
```

This creates an aggregate route, but there are no contributing routes that are more specific defined yet. Therefore, the aggregate route remains inactive:

```
[/]
A:admin@PE-1# show router aggregate

===============================================================================
Legend: G - generate-icmp enabled
===============================================================================
Aggregates (Router: Base)
===============================================================================
Prefix                                        Aggr IP-Address   Aggr AS
   Summary                                       AS Set          State
     NextHop                                       Community     NextHopType
-------------------------------------------------------------------------------
10.16.12.0/22                                 0.0.0.0           0
   False                                         False           Inactive
    192.168.11.11                                 64496:64498       Indirect
-------------------------------------------------------------------------------
No. of Aggregates: 1
===============================================================================
```

The inactive aggregate route does not appear in the routing table:

```
[/]
A:admin@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     00h02m31s  0
       system                                                        0
192.168.12.0/30                               Local   Local     00h02m31s  0
       int-PE-1-PE-2                                                 0
192.168.14.0/30                               Local   Local     00h02m31s  0
       int-PE-1-PE-4                                                 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
===============================================================================
```

Configure contributing routes to activate the aggregate route

The aggregate route remains inactive as long as there is no contributing route which is more specific than the aggregate route. The following contributing routes are statically configured on PE-1:

```
# on PE-1:
configure {
    router "Base" {
        static-routes {
            route 10.16.12.0/24 route-type unicast {
                next-hop "192.168.14.2" {
                    admin-state enable
                }
            }
            route 10.16.13.0/24 route-type unicast {
                next-hop "192.168.14.2" {
                    admin-state enable
                }
            }
            route 10.16.14.0/24 route-type unicast {
                next-hop "192.168.14.2" {
                    admin-state enable
                }
            }
            route 10.16.15.0/24 route-type unicast {
                next-hop "192.168.14.2" {
                    admin-state enable
                }
            }
```

As a result, the aggregate route becomes active:

```
[/]
A:admin@PE-1# show router aggregate

===============================================================================
Legend: G - generate-icmp enabled
===============================================================================
Aggregates (Router: Base)
===============================================================================
Prefix                                          Aggr IP-Address   Aggr AS
   Summary                                         AS Set          State
     NextHop                                         Community     NextHopType
-------------------------------------------------------------------------------
10.16.12.0/22                                   0.0.0.0           0
   False                                           False           Active
    192.168.11.11                                   64496:64498       Indirect
-------------------------------------------------------------------------------
No. of Aggregates: 1
===============================================================================
```

The active aggregate route is added to the route table, as well as the contributing routes:

```
[/]
```

```
A:admin@PE-1# show router route-table

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
10.16.12.0/22                                   Blackh* Aggr      00h00m29s  130
       Black Hole                                                   0
10.16.12.0/24                                   Remote  Static    00h00m29s  5
       192.168.14.2                                                 1
10.16.13.0/24                                   Remote  Static    00h00m29s  5
       192.168.14.2                                                 1
10.16.14.0/24                                   Remote  Static    00h00m29s  5
       192.168.14.2                                                 1
10.16.15.0/24                                   Remote  Static    00h00m29s  5
       192.168.14.2                                                 1
192.0.2.1/32                                    Local   Local     00h03m20s  0
       system                                                       0
192.168.12.0/30                                 Local   Local     00h03m20s  0
       int-PE-1-PE-2                                                0
192.168.14.0/30                                 Local   Local     00h03m20s  0
       int-PE-1-PE-4                                                0
-------------------------------------------------------------------------------
No. of Routes: 8
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
* indicates that the corresponding row element may have been truncated.
```

The aggregate route is black-holed because the next hop is not resolved. There is no route to 192.168.11.0/24.

Configure resolving route to indirect next hop

A static route is configured on PE-1 to the indirect next hop, as follows:

```
# on PE-1:
configure {
    router "Base" {
        static-routes {
            route 192.168.11.0/24 route-type unicast {
                next-hop "192.168.12.2" {
                    admin-state enable
                }
            }
```

In the route table, the aggregate route is no longer black-holed. The next hop for the indirect next hop is 192.168.12.2 (PE-2).

```
[/]
A:admin@PE-1# show router route-table

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
```

```
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.16.12.0/22                                 Remote  Aggr      00h00m03s  130
       192.168.12.2                                               0
10.16.12.0/24                                 Remote  Static    00h00m55s  5
       192.168.14.2                                               1
10.16.13.0/24                                 Remote  Static    00h00m55s  5
       192.168.14.2                                               1
10.16.14.0/24                                 Remote  Static    00h00m55s  5
       192.168.14.2                                               1
10.16.15.0/24                                 Remote  Static    00h00m55s  5
       192.168.14.2                                               1
192.0.2.1/32                                  Local   Local     00h03m45s  0
       system                                                     0
192.168.11.0/24                               Remote  Static    00h00m03s  5
       192.168.12.2                                               1
192.168.12.0/30                               Local   Local     00h03m45s  0
       int-PE-1-PE-2                                              0
192.168.14.0/30                               Local   Local     00h03m45s  0
       int-PE-1-PE-4                                              0
-------------------------------------------------------------------------------
No. of Routes: 9
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

In this example, PE-2 is the resolved indirect next hop and it has a route for prefix 10.16.12.0/22:

```
# on PE-2:
configure {
    router "Base" {
        static-routes {
            route 10.16.12.0/22 route-type unicast {
                next-hop "192.168.23.2" {
                    admin-state enable
                }
            }
        }
```

The route table on PE-2 looks as follows:

```
[/]
A:admin@PE-2# show router route-table

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.16.12.0/22                                 Remote  Static    00h00m00s  5
       192.168.23.2                                               1
192.0.2.2/32                                  Local   Local     00h04m03s  0
       system                                                     0
192.168.12.0/30                               Local   Local     00h04m03s  0
       int-PE-2-PE-1                                              0
192.168.23.0/30                               Local   Local     00h04m03s  0
       int-PE-2-PE-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
===============================================================================
```

Conclusion

Aggregate routes offer several advantages, the key being reduction in the routing table size and overcoming routing loops, among other things. Aggregate routes with indirect next hop option helps in faster network convergence by decreasing the number of route table changes. This example shows how to configure aggregate routes with indirect next hop option.

Bi-Directional Forwarding Detection

Topics in this chapter include:

Applicability

This chapter was originally written for SR OS Release 8.0.R4. The MD-CLI in the current edition corresponds to SR OS Release 23.3.R1.

Overview

Bi-directional forwarding detection (BFD) is a lightweight protocol that provides rapid path failure detection between two systems. It has been published as a series of RFCs: RFC 5880, RFC 5881, RFC 5882, RFC 5883, and RFC 5884.

If a system running BFD stops receiving BFD messages on an interface, it will determine that there has been a failure in the path and notify other protocols associated with the interface. BFD is useful in situations where two nodes are interconnected through either an optical dense wavelength division multiplexing (DWDM) or Ethernet network. In both cases, the physical network has numerous extra devices which are not part of the Layer 3 network and therefore, the Layer 3 nodes are incapable of detecting failures which occur in the physical network on spans to which the Layer 3 devices are not directly connected.

BFD protocol provides rapid link continuity checking between network devices, and the state of BFD can be propagated to IP routing protocols to drastically reduce convergence time in cases where a physical network error occurs in a transport network.

RFC 5880 defines two modes of operation for BFD:

- Asynchronous mode (supported) — Uses periodic BFD control messages to test the path between systems. If a number (configured as **multiplier**) of BFD hello packets are not received, the session is considered down.
- Demand mode (not supported)

In addition to the two operational modes, an echo function is defined. SR OS routers only support response sending, which is looping back received BFD messages to the original sender.

BFD is running between two peers and supported for scenarios such as:

- BFD for IS-IS
- BFD for OSPF

- BFD for PIM
- BFD for static routes
- BFD for RSVP
- BFD for I-LDP
- BFD for T-LDP
- BFD for MPLS-TP
- BFD for OSPF CE-PE adjacencies
- BFD for VRRP
- BFD for SRRP
- BFD for IPSec

Many of these BFD scenarios are described in this chapter.

Configuration

BFD packets are processed both locally on the IOM CPU and centrally on the CPM.

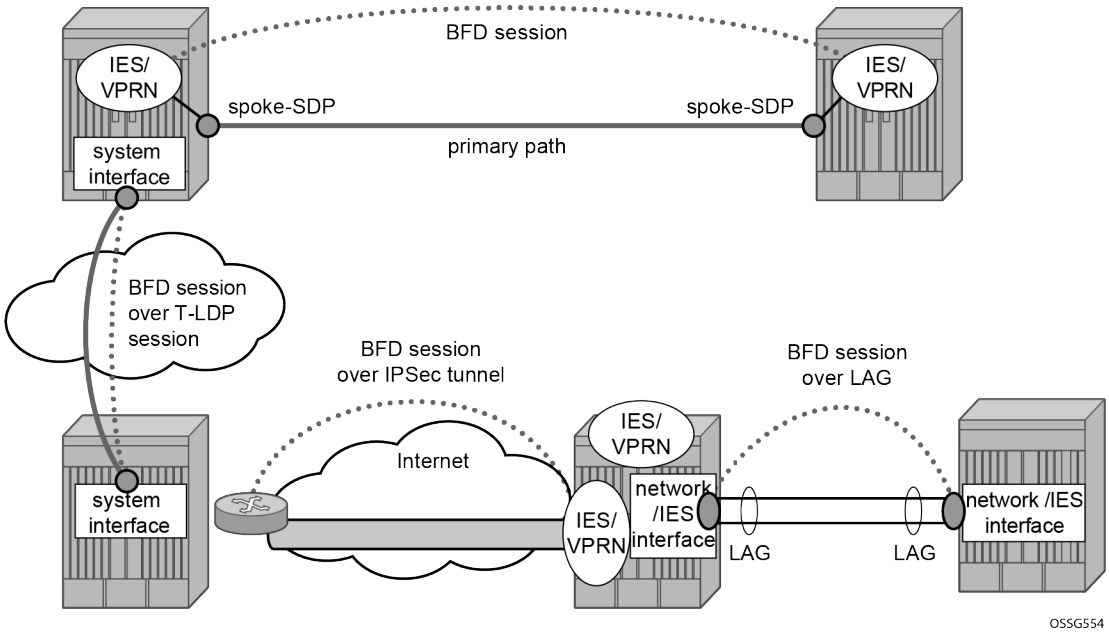
The CPM is able to centrally generate the BFD packets at a subsecond interval as low as 10 ms. The BFD state machine is implemented in software. BFD packet generation can be selectively delegated to CPM hardware as needed. This is applicable when subsecond operations or exceeding the IOM scaling limits is required.

The following applications require BFD to run centrally on the SF/CPM and a centralized session will be created independently of the type explicitly declared by the user:

- BFD for IES/VPRN over spoke SDP
- BFD for LAG and VSM interfaces
- Protocol associations using loopback and system interfaces (for example, BFD for T-LDP)
- BFD for IPSec sessions
- BFD sessions associated with multi-hop peering (BGP)

Figure 11: BFD centralized sessions shows the most relevant scenarios where centralized BFD sessions are used.

Figure 11: BFD centralized sessions



On the other end, when the two peers are directly connected, the BFD session is local by default, but the user can choose what session type (local or centralized) to implement.

As general rule, the following steps are required to configure and enable a BFD session when peers are directly connected:

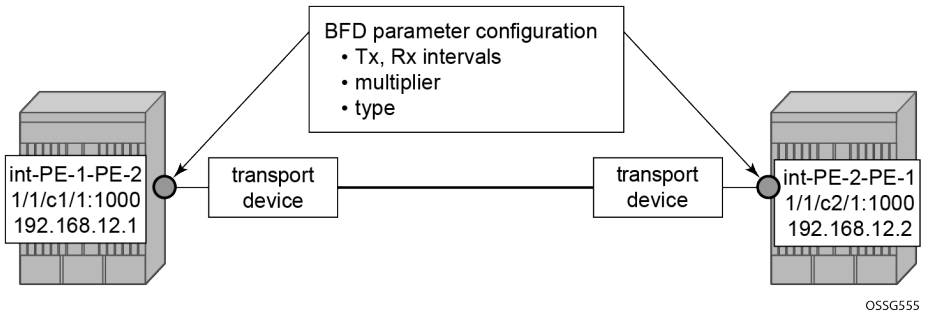
1. configure BFD parameters on the peering interfaces
2. check that the Layer 3 protocol, that is to be bound to BFD, is up and running
3. enable BFD under the Layer 3 protocol interface.

Because most of the following procedures share the same first step, it is described only once in the next section and then referred to in subsequent sections.

BFD base parameter configuration and troubleshooting

The reference topology for the generic configuration of BFD over two local peers is shown in Figure 12: BFD interface configuration.

Figure 12: BFD interface configuration



The user needs to configure base level BFD on interfaces between the peers PE-1 and PE-2.

```
# on PE-1:
configure {
    router "Base" {
        interface "int-PE-1-PE-2"
            port 1/1/c1/1:1000
            ipv4 {
                bfd {
                    admin-state enable
                }
                primary {
                    address 192.168.12.1
                    prefix-length 30
                }
            }
```

```
# on PE-2:
configure {
    router "Base" {
        interface "int-PE-2-PE-1" {
            port 1/1/c2/1:1000
            ipv4 {
                bfd {
                    admin-state enable
                }
                primary {
                    address 192.168.12.2
                    prefix-length 30
                }
            }
```

The default values for the BFD parameters are:

- transmit interval 100 ms
- receive interval 100 ms
- multiplier 3

```
*[ex:/configure router "Base" interface "int-PE-1-PE-2" ipv4 bfd]
A:admin@PE-1# info detail
    admin-state enable
    transmit-interval 100
    receive 100
    multiplier 3
 ## echo-receive
```

```
    type auto
```

The following **show** commands are used to verify the BFD configuration on the router interfaces on PE-1 and PE-2.

On PE-1:

```
[/]
A:admin@PE-1# show router bfd interface

===============================================================================
BFD Interface
===============================================================================
Interface name                    Tx Interval    Rx Interval    Multiplier
-------------------------------------------------------------------------------
int-PE-1-PE-2                     100            100            3
-------------------------------------------------------------------------------
No. of BFD Interfaces: 1
===============================================================================
```

On PE-2:

```
[/]
A:admin@PE-2# show router bfd interface

===============================================================================
BFD Interface
===============================================================================
Interface name                    Tx Interval    Rx Interval    Multiplier
-------------------------------------------------------------------------------
int-PE-2-PE-1                     100            100            3
-------------------------------------------------------------------------------
No. of BFD Interfaces: 1
===============================================================================
```

Note: BFD is an asynchronous protocol, so it is possible to configure different transmit and receive intervals on the two peers. This is because BFD transmit and receive interval values are signaled in the BFD packets while establishing the BFD session.

The configurable BFD parameters are the following:

```
*[ex:/configure router "Base" interface "int-PE-1-PE-2" ipv4 bfd]
A:admin@PE-1# ?

 admin-state           - Administrative state of BFD sessions
 echo-receive          - Minimum echo interval over this interface
 multiplier            - Number of consecutive BFD messages missed from the peer
 receive               - BFD receive interval over this interface
 transmit-interval     - BFD transmit interval over this interface
 type                  - Local termination point for the BFD session
```

By default, the BFD type is auto, but it is possible to force the BFD session to be centrally managed by the CPM hardware: **type cpm-np**.

Regarding the echo function, it is possible to set the minimum echo receive interval, in milliseconds, for the BFD session.

The base BFD configuration on the router interfaces is not sufficient for a BGP session to come up:

```
*A:PE-1# 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
-------------------------------------------------------------------------------
No Matching Entries Found
===============================================================================
```

Configuring the BFD parameters on the interface does not enable BFD sessions. BFD can be enabled afterward, for instance, in IS-IS.

Note: If a BFD session is active on an interface, it is possible to modify the BFD intervals and the multiplier on the interface, but not the BFD type. To change the BFD type, the BFD session must be disabled manually, which causes the upper layer protocols bound to it to be brought down as well.

If a BFD session is active on the interface, an attempt to modify the BFD type triggers the following error message:

```
[ex:/configure router "Base" interface "int-PE-1-PE-2" ipv4 bfd]
A:admin@PE-1# type cpm-np

*[ex:/configure router "Base" interface "int-PE-1-PE-2" ipv4 bfd]
A:admin@PE-1# commit
INFO: BFD #1001: configure router "Base" interface "int-PE-1-PE-2" - Inconsistent value - BFD
 sessions active on this interface. Cannot change BfdType on this interface
```

Forcing a centralized session in the case of directly connected peers can be useful when:

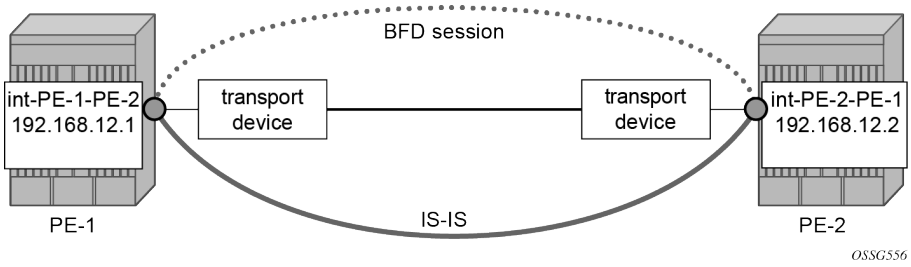
- lower Tx and Rx intervals are desired (down to 10 ms instead of 100 ms supported by local sessions)
- no more local (IOM) sessions are available
- the maximum limit of 500 packets per second per IOM has been reached

The instructions illustrated in following paragraphs are required to complete the configuration and enable BFD.

BFD for IS-IS

The goal of this section is to configure BFD on a network interlink between two SR OS nodes that are IS-IS peers. Figure 13: BFD for ISIS shows the used topology.

Figure 13: BFD for ISIS



For the base BFD configuration, see the BFD base parameter configuration and troubleshooting section.

On PE-1, BFD is applied to the IS-IS interface between PE-1 and PE-2:

```
# on PE-1:
configure {
    router "Base" {
        isis 0 {
            interface "int-PE-1-PE-2" {
                bfd-liveness {
                    ipv4 {
                    }
                }
```

When BFD is only applied on PE-1 and not on PE-2, the BFD session on PE-1 remains down, as follows:

```
[/]
A:admin@PE-1# 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
-------------------------------------------------------------------------------
int-PE-1-PE-2                                  Down           11          0
  192.168.12.2                                    3         1000        100
  isis                                          iom          N/A        N/A
  192.168.12.1
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
```

On PE-2, BFD is enabled on the interface to PE-1, as follows:

```
# on PE-2:
configure {
    router "Base" {
        isis 0 {
            interface "int-PE-2-PE-1" {
                bfd-liveness {
```

```
                    ipv4 {
                    }
                }
```

The following command verifies that the local IOM BFD session is operational between PE-1 and PE-2.

On PE-1:

```
[/]
A:admin@PE-1# 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
-------------------------------------------------------------------------------
int-PE-1-PE-2                                       Up          231        179
  192.168.12.2                                       3          100        100
  isis                                             iom          N/A        N/A
  192.168.12.1
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
```

On PE-2:

```
[/]
A:admin@PE-2# 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
-------------------------------------------------------------------------------
int-PE-2-PE-1                                       Up          152        151
  192.168.12.1                                       3          100        100
  isis                                             iom          N/A        N/A
  192.168.12.2
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
```

If the command shows that the BFD session is down, troubleshoot it by first checking that the protocol that is bound to it is up: for instance, check the IS-IS adjacency, as follows:

```
[/]
A:admin@PE-1# show router isis adjacency "int-PE-1-PE-2"
```

```
===============================================================================
Rtr Base ISIS Instance 0 Adjacency
===============================================================================
System ID               Usage State Hold Interface                    MT-ID
-------------------------------------------------------------------------------
PE-2                    L1L2  Up    22   int-PE-1-PE-2                0
-------------------------------------------------------------------------------
Adjacencies : 1
===============================================================================
```

If the IS-IS adjacency is up, then check whether a BFD resource limit has been reached (maximum number of (local or centralized) sessions or maximum number of packets per second per IOM).

If the overloaded limit is the maximum supported number of sessions, the cause is shown in log 99 (maxSessionsPerSlot).

In this case, when one of the running sessions is manually removed or goes down, then the additional configured session will come up. If the IOM limit is reached, it is possible to bring up the session by changing the session type to centralized.

To check if the IOM CPU is able to start more local BFD sessions, execute a **show router bfd session summary** command:

```
[/]
A:admin@PE-1# show router bfd session summary

============================
BFD Session Summary
============================
Termination    Session Count
----------------------------
central                    0
cpm-np                     0
iom, slot 1                1
iom, slot 2                0
iom, slot 3                0
iom, slot 4                0
iom, slot 5                0
iom, slot 6                0
Total                      1
============================
```

The **show router bfd session src <ip-address> detail** command can help debugging the BFD session. The sent and received counters are not supported for cpm-np type sessions.

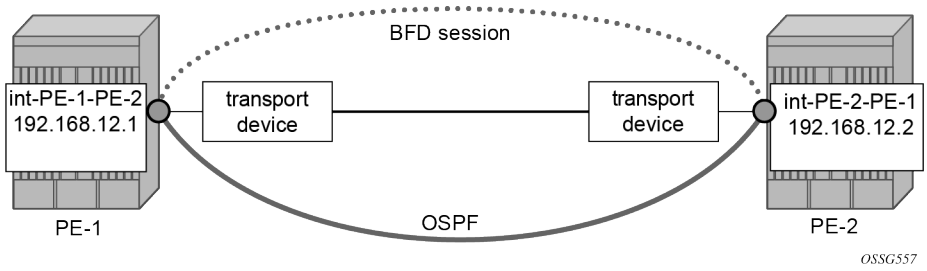
```
[/]
A:admin@PE-1# show router bfd session src 192.168.12.1 detail

===============================================================================
BFD Session
===============================================================================
Remote Address : 192.168.12.2
Local Address  : 192.168.12.1
Admin State    : Up                       Oper State       : Up
Protocols      : isis
Rx Interval    : 100                      Tx Interval      : 100
Multiplier     : 3                        Echo Interval    : 0
Recd Msgs      : 681                      Sent Msgs        : 718
Up Time        : 0d 00:00:53              Up Transitions   : 1
Last Down Time : 0d 00:00:32              Down Transitions : 0
                                          Version Mismatch : 0
```

```
Forwarding Information

Local Discr    : 1                        Local State      : Up
Local Diag     : 0 (None)                 Local Mode       : Async
Local Min Tx   : 100                      Local Mult       : 3
Last Sent      : 04/20/2023 16:17:48      Local Min Rx     : 100
Type           : iom
Remote Discr   : 1                        Remote State     : Up
Remote Diag    : 0 (None)                 Remote Mode      : Async
Remote Min Tx  : 100                      Remote Mult      : 3
Remote C-flag  : 1
Last Recv      : 04/20/2023 16:17:48      Remote Min Rx    : 100
===============================================================================
===============================================================================
```

BFD for OSPF

The goal of this section is to configure BFD on a network interlink between two SR OS nodes that are OSPF peers. Figure 14: BFD for OSPFshows the topology for this scenario.

Figure 14: BFD for OSPF



The base BFD configuration is described in the section BFD base parameter configuration and troubleshooting.

In this section, BFD is applied on the OSPF interfaces, as follows:

```
# on PE-1:
configure {
    router "Base" {
        ospf 0 {
            admin-state enable
            traffic-engineering true
            area 0.0.0.0 {
                interface "int-PE-1-PE-2" {
                    interface-type point-to-point
                    bfd-liveness {
                    }
                }
                interface "system" {
                }
            }
```

```
# on PE-2:
configure {
    router "Base" {
        ospf 0 {
            admin-state enable
```

```
            traffic-engineering true
            area 0.0.0.0 {
                interface "int-PE-2-PE-1" {
                    interface-type point-to-point
                    bfd-liveness {
                    }
                }
                interface "system" {
                }
            }
```

The following commands verify that the BFD session for OSPF is operational between PE-1 and PE-2.

On PE-1:

```
[/]
A:admin@PE-1# 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
-------------------------------------------------------------------------------
int-PE-1-PE-2                                      Up          102        101
  192.168.12.2                                      3          100        100
  ospf2                                            iom          N/A        N/A
  192.168.12.1
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
```

On PE-2:

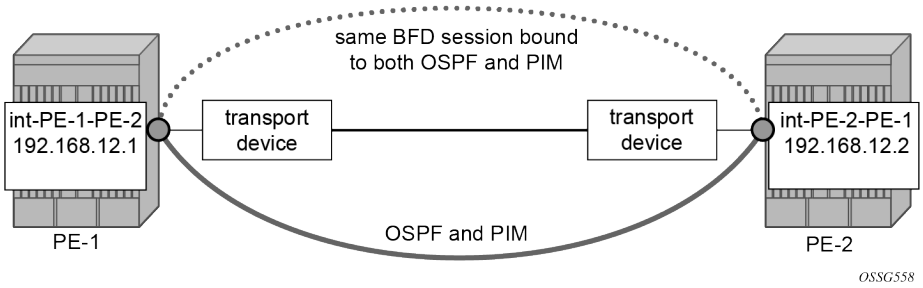
```
[/]
A:admin@PE-2# show router bfd session

===============================================================================
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
-------------------------------------------------------------------------------
int-PE-2-PE-1                                      Up           69         69
  192.168.12.1                                      3          100        100
  ospf2                                            iom          N/A        N/A
  192.168.12.2
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
```

BFD for PIM

The PIM implementation uses an interior gateway protocol (IGP) in order to determine its reverse path forwarding (RPF) tree, so the BFD configuration to support PIM requires the BFD configuration of both the IGP protocol and the PIM protocol. In this example, the IGP protocol is OSPF and that the initial configuration is as described in the section BFD for OSPF.

Figure 15: BFD for OSPF and PIM shows the topology. BFD is configured and enabled for PIM on the same interfaces that are configured with BFD for OSPF.

Figure 15: BFD for OSPF and PIM



The following commands enable BFD on the PIM interfaces on PE-1 and PE-2.

```
# on PE-1:
configure {
    router "Base" {
        pim {
            admin-state enable
            interface "int-PE-1-PE-2" {
                bfd-liveness {
                    ipv4 true
                }
            }
```

```
# on PE-2:
configure {
    router "Base" {
        pim {
            admin-state enable
            interface "int-PE-2-PE-1" {
                bfd-liveness {
                    ipv4 true
                }
            }
```

The following commands show that the BFD session is operational for OSPF and PIM between PE-1 and PE-2.

```
[/]
A:admin@PE-1# show router bfd session

===============================================================================
Legend:
  Session Id = Interface Name | LSP Name | Prefix | RSVP Sess Name | Service Id
  wp = Working path   pp = Protecting path
```

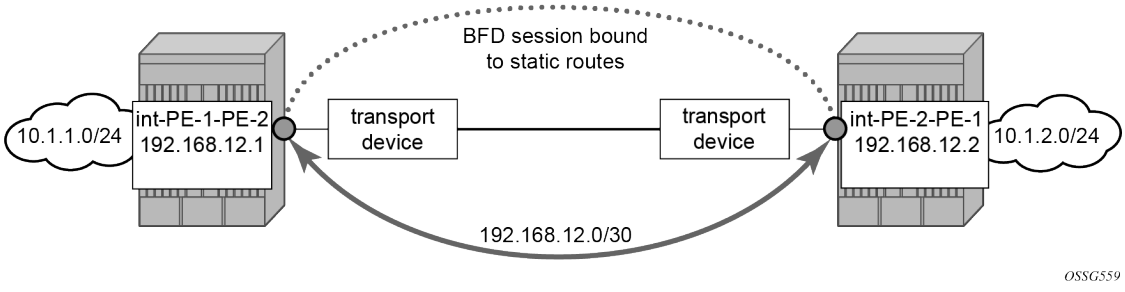
```
===============================================================================
BFD Session
===============================================================================
Session Id                                      State      Tx Pkts    Rx Pkts
  Rem Addr/Info/SdpId:VcId                    Multipl     Tx Intvl   Rx Intvl
  Protocols                                      Type     LAG Port     LAG ID
  Loc Addr                                                           LAG name
-------------------------------------------------------------------------------
int-PE-1-PE-2                                    Up          661        660
  192.168.12.2                                    3          100        100
  ospf2 pim                                     iom          N/A        N/A
  192.168.12.1
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
```

```
[/]
A:admin@PE-2# 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
-------------------------------------------------------------------------------
int-PE-2-PE-1                                    Up          637        637
  192.168.12.1                                    3          100        100
  ospf2 pim                                     iom          N/A        N/A
  192.168.12.2
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
```

BFD for static routes

In this section, BFD is applied to static routes between PE-1 and PE-2. Figure 16: BFD for static routes shows the topology.

Figure 16: BFD for static routes



The base level BFD is already configured on PE-1 and PE-2, as described in the BFD base parameter configuration and troubleshooting section.

The following commands configure static routes toward the remote networks in PE-1 and PE-2 using the BFD interfaces as next hop. BFD is enabled on the the next hop interfaces.

Note: BFD cannot be enabled if the next hop is indirect or the **black-hole** keyword is specified.

```
# on PE-1:
configure {
    router "Base" {
        static-routes {
            route 10.1.2.0/24 route-type unicast {
                next-hop "192.168.12.2" {
                    admin-state enable
                    bfd-liveness true
                }
            }
        }
```

```
# on PE-2:
configure {
    router "Base" {
        static-routes {
            route 10.1.1.0/24 route-type unicast {
                next-hop "192.168.12.1" {
                    admin-state enable
                    bfd-liveness true
                }
            }
        }
```

The following commands show the static routes populated in the routing tables on PE-1 and PE-2.

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

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
10.1.2.0/24                                     Remote  Static    00h00m04s  5
       192.168.12.2                                                    1
-------------------------------------------------------------------------------
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
===============================================================================
```

```
*A:PE-2# show router route-table protocol static

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
```

```
-------------------------------------------------------------------------------
10.1.1.0/24                                 Remote  Static    00h00m03s  5
       192.168.12.1                                                  1
-------------------------------------------------------------------------------
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 following commands show the BFD session status on PE-1 and PE-2.

```
[/]
A:admin@PE-1# 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
-------------------------------------------------------------------------------
int-PE-1-PE-2                                  Up          431        427
  192.168.12.2                                  3          100        100
  static                                       iom         N/A        N/A
  192.168.12.1
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
```

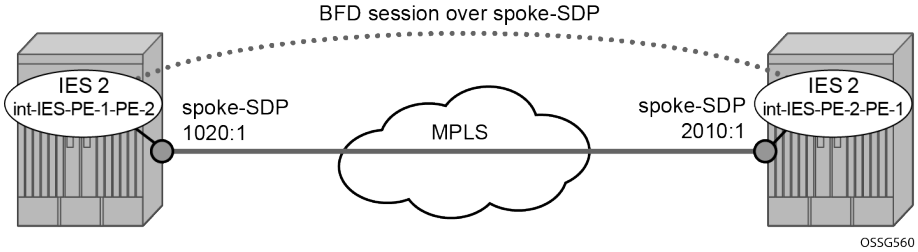
```
[/]
A:admin@PE-2# show router bfd session

===============================================================================
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
-------------------------------------------------------------------------------
int-PE-2-PE-1                                  Up          399        398
  192.168.12.1                                  3          100        100
  static                                       iom         N/A        N/A
  192.168.12.2
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
```

BFD for IES

The goal of this section is to configure BFD for an IES service over a spoke SDP.

The IES service is configured on PE-1 and PE-2, and their interfaces are connected by spoke SDPs. Figure 17: BFD for IES over spoke SDP shows the topology.

Figure 17: BFD for IES over spoke SDP



In this scenario, BFD is run between the IES interfaces independent of the SDP or LSP paths.

The following commands on PE-1 and PE-2 configure an IES service and add the IES interfaces to the OSPF area domain. BFD is not configured yet.

```
# on PE-1:
configure {
    service {
        sdp 1020 {
            admin-state enable
            delivery-type mpls
            sr-isis true
            far-end {
                ip-address 192.0.2.2
            }
        }
        ies "IES-2" {
            admin-state enable
            service-id 2
            customer "1"
            interface "int-IES-PE-1-PE-2" {
                spoke-sdp 1020:1 {
                }
                ipv4 {
                    primary {
                        address 192.168.12.5
                        prefix-length 30
                    }
                }
            }
        }
    }
    router "Base" {
        ospf 0 {
            area 0.0.0.0 {
                interface "int-IES-PE-1-PE-2"
                }
            }
```

```
        }
```

```
# on PE-2:
configure {
    service {
        sdp 2010 {
            admin-state enable
            delivery-type mpls
            sr-isis true
            far-end {
                ip-address 192.0.2.1
            }
        }
        ies "IES-2" {
            admin-state enable
            service-id 2
            customer "1"
            interface "int-IES-PE-2-PE-1" {
                spoke-sdp 2010:1 {
                }
                ipv4 {
                    primary {
                        address 192.168.12.6
                        prefix-length 30
                    }
                }
            }
        }
    }
    router "Base" {
        ospf 0 {
            area 0.0.0.0 {
                interface "int-IES-PE-2-PE-1" {
                }
            }
        }
```

The following commands verify that OSPF and the services are up on both routers.

On PE-1:

```
[/]
A:admin@PE-1# show service id 2 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 2                   Vpn Id            : 0
Service Type      : IES
MACSec enabled    : no
Name              : IES-2
Description       : (Not Specified)
Customer Id       : 1                   Creation Origin   : manual
Last Status Change: 04/20/2023 16:24:00
Last Mgmt Change  : 04/20/2023 16:23:45
Admin State       : Up                  Oper State        : Up
SAP Count         : 0                   SDP Bind Count    : 1

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                               Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
```

```
sdp:1020:1 S(192.0.2.2)                  Spok         0       8910    Up   Up
===============================================================================
```

```
[/]
A:admin@PE-1# show router ospf neighbor

===============================================================================
Rtr Base OSPFv2 Instance 0 Neighbors
===============================================================================
Interface-Name                   Rtr Id          State      Pri  RetxQ   TTL
   Area-Id
-------------------------------------------------------------------------------
int-PE-1-PE-2                    192.0.2.2       Full       1    0       39
   0.0.0.0
int-IES-PE-1-PE-2                192.0.2.2       Full       1    0       30
   0.0.0.0
-------------------------------------------------------------------------------
No. of Neighbors: 2
===============================================================================
```

On PE-2:

```
[/]
A:admin@PE-2# show service id 2 base

===============================================================================
Service Basic Information
===============================================================================
Service Id        : 2                    Vpn Id            : 0
Service Type      : IES
MACSec enabled    : no
Name              : IES-2
Description       : (Not Specified)
Customer Id       : 1                    Creation Origin   : manual
Last Status Change: 04/20/2023 16:23:59
Last Mgmt Change  : 04/20/2023 16:23:53
Admin State       : Up                   Oper State        : Up
SAP Count         : 0                    SDP Bind Count    : 1

-------------------------------------------------------------------------------
Service Access & Destination Points
-------------------------------------------------------------------------------
Identifier                                Type         AdmMTU  OprMTU  Adm  Opr
-------------------------------------------------------------------------------
sdp:2010:1 S(192.0.2.1)                   Spok         0       8910    Up   Up
===============================================================================
```

```
[/]
A:admin@PE-2# show router ospf neighbor

===============================================================================
Rtr Base OSPFv2 Instance 0 Neighbors
===============================================================================
Interface-Name                   Rtr Id          State      Pri  RetxQ   TTL
   Area-Id
-------------------------------------------------------------------------------
int-PE-2-PE-1                    192.0.2.1       Full       1    0       32
   0.0.0.0
int-IES-PE-2-PE-1                192.0.2.1       Full       1    0       34
   0.0.0.0
-------------------------------------------------------------------------------
No. of Neighbors: 2
```

```
===============================================================================
```

The following commands on PE-1 and PE-2 configure BFD on the IES interfaces and enable BFD on the OSPF interfaces.

```
# on PE-1:
configure {
    service {
        ies "IES-2" {
            interface "int-IES-PE-1-PE-2" {
                ipv4 {
                    bfd {
                        admin-state enable
                    }
                }
            }
        }
    }
    router "Base" {
        ospf 0 {
            area 0.0.0.0 {
                interface "int-IES-PE-1-PE-2" {
                    bfd-liveness {
                    }
                }
            }
```

```
# on PE-2:
configure {
    service {
        ies "IES-2" {
            interface "int-IES-PE-2-PE-1" {
                ipv4 {
                    bfd {
                        admin-state enable
                    }
                }
           }
            info
        }
    }
    router "Base" {
        ospf 0 {
            area 0.0.0.0 {
                interface "int-IES-PE-2-PE-1" {
                    bfd-liveness {
                    }
                }
            }
```

A centralized BFD session is created for BFD over spoke SDP even if a physical link exists between the two nodes. This centralized BFD session is created because the spoke SDP is terminated at the CPM. This is also the case for BFD running over LAG bundles.

The *central* type is used when BFD packets are completely generated and processed by software on the CPM. The *cpm-np* type is used when BFD packets are generated and processed with hardware assistance on the CPM. The following output shows that BFD session type is **cpm-np**.

```
[/]
A:admin@PE-1# 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
-------------------------------------------------------------------------------
int-IES-PE-1-PE-2                                   Up          N/A        N/A
  192.168.12.6                                       3         1000       1000
  ospf2                                          cpm-np          N/A        N/A
  192.168.12.5
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
```

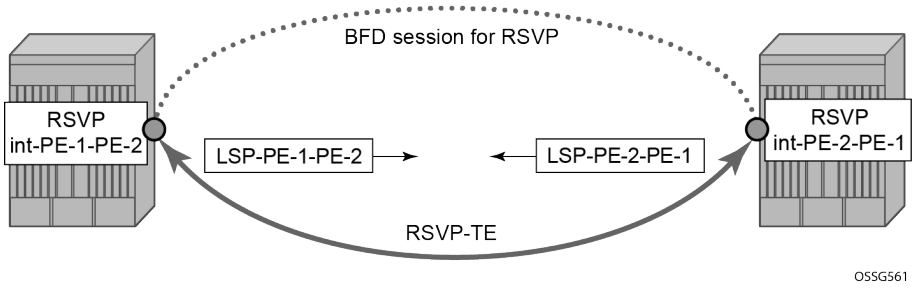
```
[/]
A:admin@PE-2# show router bfd session

===============================================================================
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
-------------------------------------------------------------------------------
int-IES-PE-2-PE-1                                   Up          N/A        N/A
  192.168.12.5                                       3         1000       1000
  ospf2                                          cpm-np          N/A        N/A
  192.168.12.6
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
```

The transmitted and received packet counters are not included in the preceding **show** commands. BFD sessions of the **cpm-np** type are handled by hardware. The hardware does not have transmitted or received packet counters. In contrast, IOM BFD sessions are handled by the CPU of the IOM, so the packets are counted. Likewise, BFD sessions of type central are handled by the CPU of the CPM and the packets are counted.

BFD for RSVP

The goal of this section is to configure BFD between two RSVP interfaces configured in two SR OS nodes. Figure 18: BFD for RSVP shows the topology for this scenario.

Figure 18: BFD for RSVP



BFD is configured on the interfaces between PE-1 and PE-2 as described in BFD base parameter configuration and troubleshooting.

The following commands on PE-1 and PE-2 configure the paths, the LSPs, and the interfaces within MPLS and RSVP.

```
# on PE-1:
configure {
    router "Base" {
        mpls {
            admin-state enable
            interface "int-PE-1-PE-2" {
                admin-state enable
            }
            interface "system" {
                admin-state enable
            }
            path "empty" {
                admin-state enable
            }
            lsp "LSP-PE-1-PE-2" {
                admin-state enable
                type p2p-rsvp
                to 192.0.2.2
                path-computation-method local-cspf
                primary "empty" {
                }
            }
        }
        rsvp {
            admin-state enable
            interface "int-PE-1-PE-2" {
            }
        }
```

```
# on PE-2:
configure {
    router "Base" {
        mpls {
            admin-state enable
            interface "int-PE-2-PE-1" {
                admin-state enable
            }
            interface "system" {
                admin-state enable
            }
            path "empty" {
                admin-state enable
```

```
            }
            lsp "LSP-PE-2-PE-1" {
                admin-state enable
                type p2p-rsvp
                to 192.0.2.1
                path-computation-method local-cspf
                primary "empty" {
                }
            }
        }
        rsvp {
            admin-state enable
            interface "int-PE-2-PE-1" {
            }
        }
```

The following command on PE-1 verifies that the RSVP sessions are up.

```
[/]
A:admin@PE-1# show router rsvp session

===============================================================================
RSVP Sessions
===============================================================================
RSVP Session Name
    From                To              Tunnel ID   LSP ID       State
-------------------------------------------------------------------------------
LSP-PE-2-PE-1::empty
192.0.2.2           192.0.2.1           1           38912        Up

LSP-PE-1-PE-2::empty
192.0.2.1           192.0.2.2           1           12800        Up

-------------------------------------------------------------------------------
Sessions : 2
===============================================================================
```

The following commands on PE-1 and PE-2 enable BFD on the RSVP interfaces.

```
# on PE-1:
configure {
    router "Base" {
        rsvp {
            interface "int-PE-1-PE-2"
                bfd-liveness true
            }
```

```
# on PE-2:
configure {
    router "Base" {
        rsvp {
            interface "int-PE-2-PE-1"
                bfd-liveness true
            }
```

The following commands verify that the BFD session is operational between PE-1 and PE-2.

On PE-1:

```
[/]
A:admin@PE-1# 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
-------------------------------------------------------------------------------
int-PE-1-PE-2                                      Up          315        284
  192.168.12.2                                      3          100        100
  rsvp                                             iom         N/A        N/A
  192.168.12.1
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
```

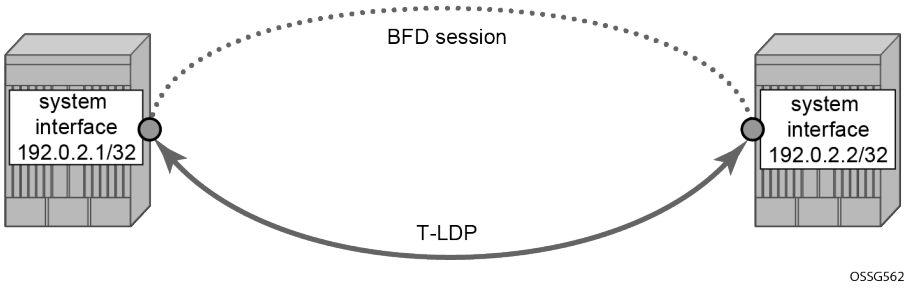
On PE-2:

```
[/]
A:admin@PE-2# 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
-------------------------------------------------------------------------------
int-PE-2-PE-1                                      Up          270        270
  192.168.12.1                                      3          100        100
  rsvp                                             iom         N/A        N/A
  192.168.12.2
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
```

BFD for T-LDP

BFD tracking of an LDP session associated with a T-LDP adjacency allows for faster detection of the liveliness of the session by registering the transport address of an LDP session with a BFD session. Figure 19: BFD for T-LDP shows the topology.

Figure 19: BFD for T-LDP



The parameters used for the BFD session are configured under the loopback interface corresponding to the LSR-ID. By default, the LSR-ID matches the system interface address.

```
# on PE-1, PE-2:
configure {
    router "Base" {
        interface "system" {
            ipv4 {
                bfd {
                    admin-state enable
                    transmit-interval 3000
                    receive 3000
                }
```

The loopback interface can be used to source BFD sessions to many peers in the network.

When using BFD over other links with the ability to reroute, such as spoke-SDPs, the interval and multiplier values configuring BFD should be set to allow sufficient time for the underlying network to reconverge before the associated BFD session expires. A general rule of thumb should be that the expiration time (interval * multiplier) is three times the convergence time for the IGP network between the two endpoints of the BFD session.

On PE-1 and PE-2, the following T-LDP session is established with BFD enabled.

```
# on PE-1:
configure {
    router "Base" {
        ldp {
            targeted-session {
                peer 192.0.2.2
                    admin-state enable
                    bfd-liveness true
                }
            }
```

```
# on PE-2:
configure {
    router "Base" {
        ldp {
            targeted-session {
                peer 192.0.2.1
                    admin-state enable
                    bfd-liveness true
                }
            }
```

By enabling BFD for a selected targeted session, the state of that session is tied to the state of the underlying BFD session between the two nodes.

The following commands on PE-1 and PE-2 verify that the T-LDP session is up.

On PE-1:

```
[/]
A:admin@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         Targeted  Established   238       237       0d 00:20:39
-------------------------------------------------------------------------------
No. of IPv4 Sessions: 1
===============================================================================
```

On PE-2:

```
[/]
A:admin@PE-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         Targeted  Established   235       237       0d 00:20:33
-------------------------------------------------------------------------------
No. of IPv4 Sessions: 1
===============================================================================
```

The following commands on PE-1 and PE-2 show that the BFD session is up.

On PE-1:

```
[/]
A:admin@PE-1# 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
  192.0.2.2                                         3         3000       3000
  ldp                                          cpm-np          N/A        N/A
  192.0.2.1
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
```

On PE-2:

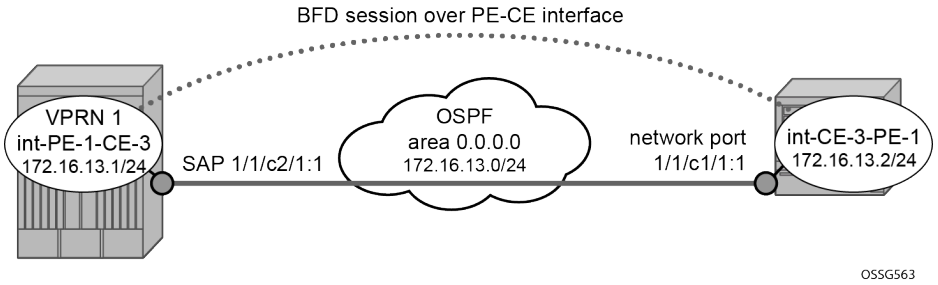
```
[/]
A:admin@PE-2# show router bfd session

===============================================================================
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
  192.0.2.1                                          3         3000       3000
  ldp                                           cpm-np          N/A        N/A
  192.0.2.2
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
```

When the T-LDP session comes up, a centralized **cpm-np** BFD session is always created even if the local interface has a direct link to the peer.

BFD for OSPF PE-CE adjacencies

BFD for OSPF PE-CE adjacencies extends BFD support to OSPF within a **vprn** context when OSPF is used as the PE-CE protocol. Figure 20: BFD for OSPF PE-CE interfaces shows the topology used in this section.

Figure 20: BFD for OSPF PE-CE interfaces



On PE-1, the following VPRN configuration includes service interface int-PE-1-CE-1 with BFD parameters.

```
# on PE-1:
configure {
    service {
        vprn "VPRN-1" {
            admin-state enable
            service-id 1
            customer "1"
            interface "int-PE-1-CE-3" {
                ipv4 {
```

```
                    bfd {
                        admin-state enable
                    }
                    primary {
                        address 172.16.13.1
                        prefix-length 24
                    }
                }
                sap 1/1/c2/1:1 {
                }
            }
            ospf 0 {
                admin-state enable
                area 0.0.0.0 {
                    interface "int-PE-1-CE-3" {
                        bfd-liveness {
                        }
                    }
                }
            }
```

On CE-3, the following configures the router interface int-CE-3-PE-1 with BFD parameters. BFD is enabled on this interfaces that is added to the OSPF area 0.0.0.0 domain.

```
# on CE-3:
configure {
    router "Base" {
        interface "int-CE-3-PE-1" {
            port 1/1/c1/1:1
            ipv4 {
                bfd {
                    admin-state enable
                }
                primary {
                    address 172.16.13.2
                    prefix-length 24
                }
            }
        }
        interface "system" {
            ipv4 {
                primary {
                    address 192.0.2.3
                    prefix-length 32
                }
            }
        }
        ospf 0 {
            admin-state enable
            area 0.0.0.0 {
                interface "int-CE-3-PE-1" {
                    bfd-liveness {
                    }
                }
            }
        }
```

The following command shows that the OSPF adjacency is up.

On PE-1:

```
[/]
A:admin@PE-1# show router 1 ospf neighbor
```

```
===============================================================================
Rtr vprn1 OSPFv2 Instance 0 Neighbors
===============================================================================
Interface-Name                   Rtr Id          State      Pri  RetxQ   TTL
   Area-Id
-------------------------------------------------------------------------------
int-PE-1-CE-3                    192.0.2.3       Full       1    0       31
   0.0.0.0
-------------------------------------------------------------------------------
No. of Neighbors: 1
===============================================================================
```

On CE-3:

```
[/]
A:admin@CE-3# show router ospf neighbor

===============================================================================
Rtr Base OSPFv2 Instance 0 Neighbors
===============================================================================
Interface-Name                   Rtr Id          State      Pri  RetxQ   TTL
   Area-Id
-------------------------------------------------------------------------------
int-CE-3-PE-1                    192.0.2.1       Full       1    0       38
   0.0.0.0
-------------------------------------------------------------------------------
No. of Neighbors: 1
===============================================================================
```

The following commands show that the BFD session is up in both PE-1 and CE-3.

```
[/]
A:admin@PE-1# show router 1 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
-------------------------------------------------------------------------------
int-PE-1-CE-3                                   Up         1788       1782
  172.16.13.2                                    3          100        100
  ospf2                                         iom         N/A        N/A
  172.16.13.1
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
```

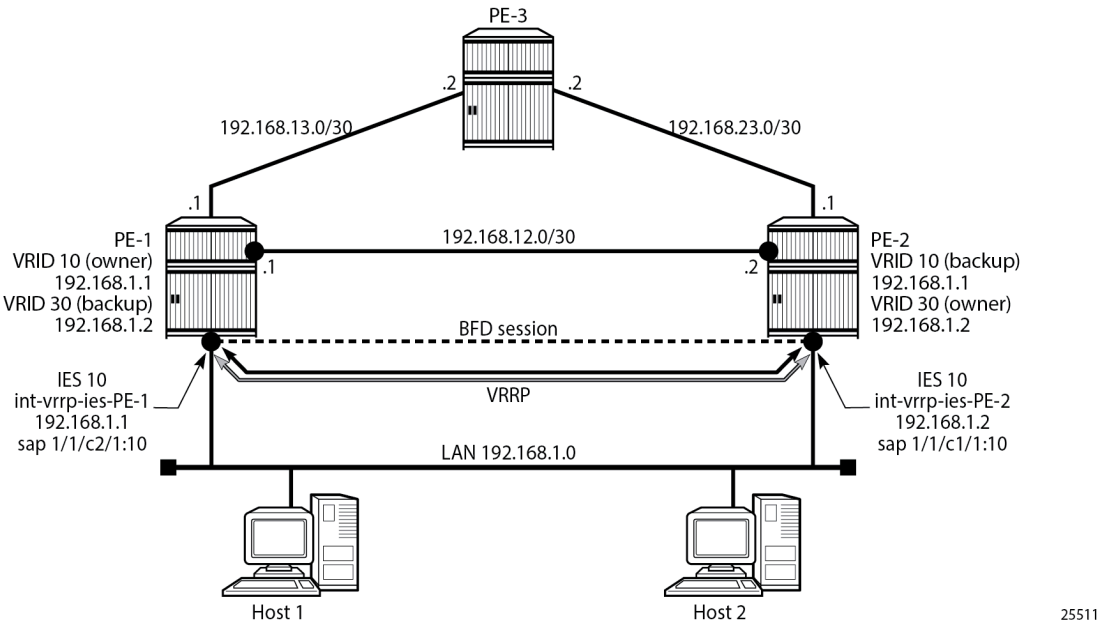
```
[/]
A:admin@CE-3# show router bfd session

===============================================================================
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
-------------------------------------------------------------------------------
int-CE-3-PE-1                                       Up         1996       1999
  172.16.13.1                                        3          100        100
  ospf2                                             iom          N/A        N/A
  172.16.13.2
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
```

BFD for VRRP

This feature assigns a BFD session to provide a heart-beat mechanism for the VRRP instance. There can be only one BFD session assigned to any VRRP instance, but there can be multiple VRRP sessions using the same BFD session. Figure 21: BFD for VRRP shows the topology for this section.

Figure 21: BFD for VRRP



Host 1 and host 2 are connected to LAN subnet 192.168.1.0/24. PE-1 and PE-2 are connected to the LAN subnet by IES or VPRN services. In the following example, IES 10 is created on PE-1 and PE-2 and BFD parameters are configured on the IES interface.

```
# on PE-1:
configure {
    service {
        ies "IES-10" {
            admin-state enable
```

```
            service-id 10
            customer "1"
            interface "int-vrrp-ies-PE-1" {
                mac 00:00:5e:00:53:01
                sap 1/1/c2/1:10 {
                }
                ipv4 {
                    bfd {
                        admin-state enable
                        multiplier 10
                    }
                    primary {
                        address 192.168.1.1
                        prefix-length 24
                    }
```

```
# on PE-2:
configure {
    service {
        ies "IES-10" {
            admin-state enable
            service-id 10
            customer "1"
            interface "int-vrrp-ies-PE-2" {
                mac 00:00:5e:00:53:02
                sap 1/1/c1/1:10 {
                }
                ipv4 {
                    bfd {
                        admin-state enable
                        multiplier 10
                    }
                    primary {
                        address 192.168.1.2
                        prefix-length 24
                    }
```

The following command on PE-1 verifies that the IES service "IES-10" is operational:

```
[/]
A:admin@PE-1# show service service-using ies

===============================================================================
Services [ies]
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
2            IES       Up   Up   1          IES-2
10           IES       Up   Up   1          IES-10
2147483648   IES       Up   Down 1          _tmnx_InternalIesService
-------------------------------------------------------------------------------
Matching Services : 3
-------------------------------------------------------------------------------
===============================================================================
```

The following command on PE-1 verifies the connectivity to the remote interface IP address 192.168.1.2:

```
[/]
A:admin@PE-1# ping 192.168.1.2 interval 0.1 output-format summary
PING 192.168.1.2 56 data bytes
!!!!!
---- 192.168.1.2 PING Statistics ----
```

```
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 3.97ms, avg = 4.03ms, max = 4.12ms, stddev = 0.068ms
```

On PE-1 and PE-2, VRRP is enabled on the IES interface that connects to the 192.168.1.0/24 subnet. In this section, the configurations are shown for the VRRP owner mode for primary but any other scenario for VRRP can be configured (non owner mode for primary). In the following example, two VRRP instances are created on the 192.168.1.0/24 subnet:

```
     VRID = 10  Owner   = PE-1
                Backup  = PE-2
                VRRP IP = 192.168.1.1
     VRID = 30  Owner   = PE-2
                Backup  = PE-1
                VRRP IP = 192.168.1.2
```

Host 1 is configured with default gateway 192.168.1.1, and host 2 is configured with default gateway 192.168.1.2.

VRRP 10 and VRRP 30 are configured on the IES interface connected to the 192.168.1.0/24 subnet. To bind the VRRP instances with a BFD session, BFD liveness is enabled on this IES interface for VRRP 10 and VRRP 30. The configuration on PE-1 is as follows:

```
# on PE-1:
configure {
    service {
        ies "IES-10" {
            admin-state enable
            service-id 10
            customer "1"
            interface "int-vrrp-ies-PE-1" {
                mac 00:00:5e:00:53:01
                sap 1/1/c2/1:10 {
                }
                ipv4 {
                    bfd {
                        admin-state enable
                        multiplier 10
                    }
                    primary {
                        address 192.168.1.1
                        prefix-length 24
                    }
                    vrrp 10 {
                        backup [192.168.1.1]
                        owner true
                        bfd-liveness {
                            dest-ip 192.168.1.2
                            service-name "IES-10"
                            interface-name "int-vrrp-ies-PE-1"
                        }
                    }
                    vrrp 30 {
                        backup [192.168.1.2]
                        ping-reply true
                        telnet-reply true
                        ssh-reply true
                        bfd-liveness {
                            dest-ip 192.168.1.2
                            service-name "IES-10"
                            interface-name "int-vrrp-ies-PE-1"
                        }
```

```
                }
            }
```

The configuration on PE-2 is as follows:

```
# on PE-2:
configure {
    service {
        ies "IES-10" {
            admin-state enable
            service-id 10
            customer "1"
            interface "int-vrrp-ies-PE-2" {
                mac 00:00:5e:00:53:02
                sap 1/1/c1/1:10 {
                }
                ipv4 {
                    bfd {
                        admin-state enable
                        multiplier 10
                    }
                    primary {
                        address 192.168.1.2
                        prefix-length 24
                    }
                    vrrp 10 {
                        backup [192.168.1.1]
                        ping-reply true
                        telnet-reply true
                        ssh-reply true
                        bfd-liveness {
                            dest-ip 192.168.1.1
                            service-name "IES-10"
                            interface-name "int-vrrp-ies-PE-2"
                        }
                    }
                    vrrp 30 {
                        backup [192.168.1.2]
                        owner true
                        bfd-liveness {
                            dest-ip 192.168.1.1
                            service-name "IES-10"
                            interface-name "int-vrrp-ies-PE-2"
                        }
                    }
                }
```

The parameters used for the BFD are set by the BFD command under the IP interface. Unlike the previous scenarios, the user can configure **bfd-liveness** for VRRP, enabling the BFD session, even if the specified interface has not been configured with BFD parameters (**ipv4>bfd>admin-state enable**).

If the BFD parameters have not been configured yet, the BFD session will be initiated only after configuring the BFD parameters (**ipv4>bfd>admin-state enable**).

```
# on PE-1:
configure {
    service {
        ies "IES-10" {
            interface "int-vrrp-ies-PE-1" {
                ipv4 {
                    bfd {
                        admin-state enable
```

```
                        # transmit-interval 100     # default
                        # receive 100               # default
                        multiplier 10
                    }
```

```
# on PE-2:
configure {
    service {
        ies "IES-10" {
            interface "int-vrrp-ies-PE-2" {
                ipv4 {
                    bfd {
                        admin-state enable
                        # transmit-interval 100     # default
                        # receive 100               # default
                        multiplier 10
                    }
```

The following command on PE-1 shows that the BFD session is up:

```
[/]
A:admin@PE-1# show router bfd session src 192.168.1.1 detail

===============================================================================
BFD Session
===============================================================================
Remote Address : 192.168.1.2
Local Address  : 192.168.1.1
Admin State    : Up                       Oper State       : Up
Protocols      : vrrp
Rx Interval    : 100                      Tx Interval      : 100
Multiplier     : 10                       Echo Interval    : 0
Recd Msgs      : 2171                     Sent Msgs        : 2185
Up Time        : 0d 00:02:48              Up Transitions   : 1
Last Down Time : 0d 00:00:23              Down Transitions : 0
                                          Version Mismatch : 0

Forwarding Information

Local Discr    : 8                        Local State      : Up
Local Diag     : 0 (None)                 Local Mode       : Async
Local Min Tx   : 100                      Local Mult       : 10
Last Sent      : 04/20/2023 16:36:33      Local Min Rx     : 100
Type           : iom
Remote Discr   : 7                        Remote State     : Up
Remote Diag    : 0 (None)                 Remote Mode      : Async
Remote Min Tx  : 100                      Remote Mult      : 10
Remote C-flag  : 1
Last Recv      : 04/20/2023 16:36:33      Remote Min Rx    : 100
===============================================================================
===============================================================================
```

This session is shared by all the VRRP instances configured between the specified interfaces.

When BFD is configured in a VRRP instance, the following command gives details of BFD related to every instance:

```
[/]
A:admin@PE-1# show router vrrp instance interface "int-vrrp-ies-PE-1"

===============================================================================
```

```
VRRP Instances for interface "int-vrrp-ies-PE-1"
===============================================================================
-------------------------------------------------------------------------------
VRID 10
-------------------------------------------------------------------------------
Owner               : Yes                 VRRP State        : Master
Primary IP of Master: 192.168.1.1 (Self)
Primary IP          : 192.168.1.1         Standby-Forwarding: Disabled
VRRP Backup Addr    : 192.168.1.1
Admin State         : Up                  Oper State        : Up
Up Time             : 04/20/2023 16:33:22 Virt MAC Addr     : 00:00:5e:00:01:0a
Auth Type           : None
Config Mesg Intvl   : 1                   In-Use Mesg Intvl : 1
Base Priority       : 255                 In-Use Priority   : 255
Init Delay          : 0                   Init Timer Expires: 0.000 sec
Creation State      : Active

-------------------------------------------------------------------------------
BFD Interface
-------------------------------------------------------------------------------
Service ID          : None
Service Name        : IES-10
Interface Name      : int-vrrp-ies-PE-1
Src IP              : 192.168.1.1
Dst IP              : 192.168.1.2
Session Oper State  : connected

-------------------------------------------------------------------------------
Master Information
-------------------------------------------------------------------------------
Primary IP of Master: 192.168.1.1 (Self)
Addr List Mismatch  : No                  Master Priority   : 255
Master Since        : 04/20/2023 16:33:22

-------------------------------------------------------------------------------
Masters Seen (Last 32)
-------------------------------------------------------------------------------
Primary IP of Master   Last Seen             Addr List Mismatch     Msg Count
-------------------------------------------------------------------------------
192.168.1.1            04/20/2023 16:33:22   No                           0

-------------------------------------------------------------------------------
Statistics
-------------------------------------------------------------------------------
Become Master       : 1                   Master Changes    : 1
Adv Sent            : 199                 Adv Received      : 0
Pri Zero Pkts Sent  : 0                   Pri Zero Pkts Rcvd: 0
Preempt Events      : 0                   Preempted Events  : 0
Mesg Intvl Discards : 0                   Mesg Intvl Errors : 0
Addr List Discards  : 0                   Addr List Errors  : 0
Auth Type Mismatch  : 0                   Auth Failures     : 0
Invalid Auth Type   : 0                   Invalid Pkt Type  : 0
IP TTL Errors       : 0                   Pkt Length Errors : 0
Total Discards      : 0

-------------------------------------------------------------------------------
VRID 30
-------------------------------------------------------------------------------
Owner               : No                  VRRP State        : Backup
Primary IP of Master: 192.168.1.2 (Other)
Primary IP          : 192.168.1.1         Standby-Forwarding: Disabled
VRRP Backup Addr    : 192.168.1.2
Admin State         : Up                  Oper State        : Up
Up Time             : 04/20/2023 16:33:22 Virt MAC Addr     : 00:00:5e:00:01:1e
```

```
Auth Type           : None
Config Mesg Intvl   : 1                   In-Use Mesg Intvl : 1
Master Inherit Intvl: No
Base Priority       : 100                 In-Use Priority   : 100
Policy ID           : n/a                 Preempt Mode      : Yes
Ping Reply          : Yes                 Telnet Reply      : Yes
Ntp Reply           : No
SSH Reply           : Yes                 Traceroute Reply  : No
Init Delay          : 0                   Init Timer Expires: 0.000 sec
Creation State      : Active

-------------------------------------------------------------------------------
BFD Interface
-------------------------------------------------------------------------------
Service ID          : None
Service Name        : IES-10
Interface Name      : int-vrrp-ies-PE-1
Src IP              : 192.168.1.1
Dst IP              : 192.168.1.2
Session Oper State  : connected

-------------------------------------------------------------------------------
Master Information
-------------------------------------------------------------------------------
Primary IP of Master: 192.168.1.2 (Other)
Addr List Mismatch  : No                  Master Priority   : 255
Master Since        : 04/20/2023 16:33:38
Master Down Interval: 3.609 sec (Expires in 2.600 sec)

-------------------------------------------------------------------------------
Masters Seen (Last 32)
-------------------------------------------------------------------------------
Primary IP of Master   Last Seen             Addr List Mismatch     Msg Count
-------------------------------------------------------------------------------
192.168.1.1            04/20/2023 16:33:25   No                           0
192.168.1.2            04/20/2023 16:36:39   No                         183

-------------------------------------------------------------------------------
Statistics
-------------------------------------------------------------------------------
Become Master       : 1                   Master Changes    : 2
Adv Sent            : 13                  Adv Received      : 183
Pri Zero Pkts Sent  : 0                   Pri Zero Pkts Rcvd: 0
Preempt Events      : 0                   Preempted Events  : 1
Mesg Intvl Discards : 0                   Mesg Intvl Errors : 0
Addr List Discards  : 0                   Addr List Errors  : 0
Auth Type Mismatch  : 0                   Auth Failures     : 0
Invalid Auth Type   : 0                   Invalid Pkt Type  : 0
IP TTL Errors       : 0                   Pkt Length Errors : 0
Total Discards      : 0

===============================================================================
```

For troubleshooting, a configuration error is introduced for VRRP 10 in service "IES-10" on PE-1. In this example, the misconfiguration is that the IES service name "IES-10" is not declared in the **bfd-enable** command for VRRP 10:

```
# on PE-1:
configure {
    service {
        ies "IES-10" {
            interface "int-vrrp-ies-PE-1" {
                ipv4 {
```

```
                    vrrp 10 {
                        bfd-liveness {
                            delete service-name
                        }
```

In this case, the BFD session between the two IP interfaces is operationally up but the command **show router vrrp instance interface <interface-name>** on PE-1 gives the following output regarding BFD for VRID 10:

```
[/]
A:admin@PE-1# show router vrrp instance interface "int-vrrp-ies-PE-1"

===============================================================================
VRRP Instances for interface "int-vrrp-ies-PE-1"
===============================================================================
-------------------------------------------------------------------------------
VRID 10
-------------------------------------------------------------------------------
Owner               : Yes                 VRRP State        : Master
Primary IP of Master: 192.168.1.1 (Self)
Primary IP          : 192.168.1.1         Standby-Forwarding: Disabled
VRRP Backup Addr    : 192.168.1.1
Admin State         : Up                  Oper State        : Up
Up Time             : 04/20/2023 16:33:22 Virt MAC Addr     : 00:00:5e:00:01:0a
Auth Type           : None
Config Mesg Intvl   : 1                   In-Use Mesg Intvl : 1
Base Priority       : 255                 In-Use Priority   : 255
Init Delay          : 0                   Init Timer Expires: 0.000 sec
Creation State      : Active

-------------------------------------------------------------------------------
BFD Interface
-------------------------------------------------------------------------------
Service ID          : None
Interface Name      : int-vrrp-ies-PE-1
Src IP              :
Dst IP              : 192.168.1.2
Session Oper State  : notConfigured

-------------------------------------------------------------------------------
---snip---
```

The session operational state and the service ID indicate that the service ID is not configured. To fix this, enable BFD with service name "IES-10" for VRRP instance 10:

```
# on PE-1:
configure {
    service {
        ies "IES-10" {
            interface "int-vrrp-ies-PE-1" {
                ipv4 {
                    vrrp 10 {
                        bfd-liveness {
                            service-name "IES-10"
                        }
```

Conclusion

BFD is a light-weight protocol which provides rapid path failure detection between two systems. BFD is useful in situations where the physical network has numerous intervening devices which are not part of the Layer 3 network.

BFD is linked to a protocol state. For a BFD session to be established, the prerequisite condition is that the protocol to which the BFD is linked must be operationally active. Once the BFD session is established, the state of the protocol to which BFD is tied to is then determined based on the BFD session's state. This means that if the BFD session goes down, the corresponding protocol will be brought down.

In this chapter, several scenarios where BFD could be implemented have been described, including the configuration, show output, and troubleshooting hints.

LFA Policies Using OSPF as IGP

This chapter provides information about LFA policies using OSPF as IGP.

Topics in this chapter include:

Applicability

This chapter was initially written for SR OS Release 12.0.R4, but the MD-CLI in the current edition corresponds to SR OS Release 23.3.R3.

Overview

Loopfree alternate (LFA) is a local control plane feature. When multiple LFAs exist, RFC 5286 chooses the LFA providing the best coverage of the failure cases. In general, this means that node LFA has preference above link LFA. In some deployments, however, this can lead to suboptimal LFA. For example, an aggregation router (typically using lower bandwidth links) protecting a core node or link (typically using high bandwidth links) is potentially undesirable.

For this reason, the operator wants to have more control in the LFA next hop selection algorithm. This is achieved by the introduction of LFA shortest path first (SPF) policies.

LFA policies can work in combination with IP fast reroute (FRR) and LDP FRR.

Implementation

The SR OS LFA policy implementation is built around the concept of **route-next-hop-policy** templates which are applied to IP interfaces. A route next hop policy template specifies criteria that influence the selection of an LFA backup next hop for either:

- a set of prefixes in a prefix list or
- a set of prefixes which resolve to a specific primary next hop

See RFC 7916 for further information. Two powerful methods which can be used as criteria inside a route next hop policy template are IP admin groups and IP shared risk link groups (SRLGs). IP admin group and IP SRLG criteria are applied before running the LFA next hop algorithm. IP admin groups and SRLGs work in a similar way as the MPLS admin groups and SRLGs.

For example, when one or more IP admin groups or SRLGs are applied to an IP interface, the same MPLS admin group and SRLG rules apply:

- IP interfaces which do not include one or more of the admin groups defined in the **include** statements are pruned before computing the LFA next hop.
- IP interfaces which belong to admin groups which have been explicitly excluded using the **exclude** statement are pruned before computing the LFA next hop.
- IP interfaces which belong to the SRLGs used by the primary next hop of a prefix are pruned before computing the LFA next hop.

For more information about MPLS admin groups, see chapter "RSVP Point-to-Point LSPs" in the *7450 ESS, 7750 SR, and 7950 XRS MPLS Advanced Configuration Guide for MD CLI*; for SRLGs, see chapter "Shared Risk Link Groups for RSVP-Based LSPs"" in the *7450 ESS, 7750 SR, and 7950 XRS MPLS Advanced Configuration Guide for MD CLI*.

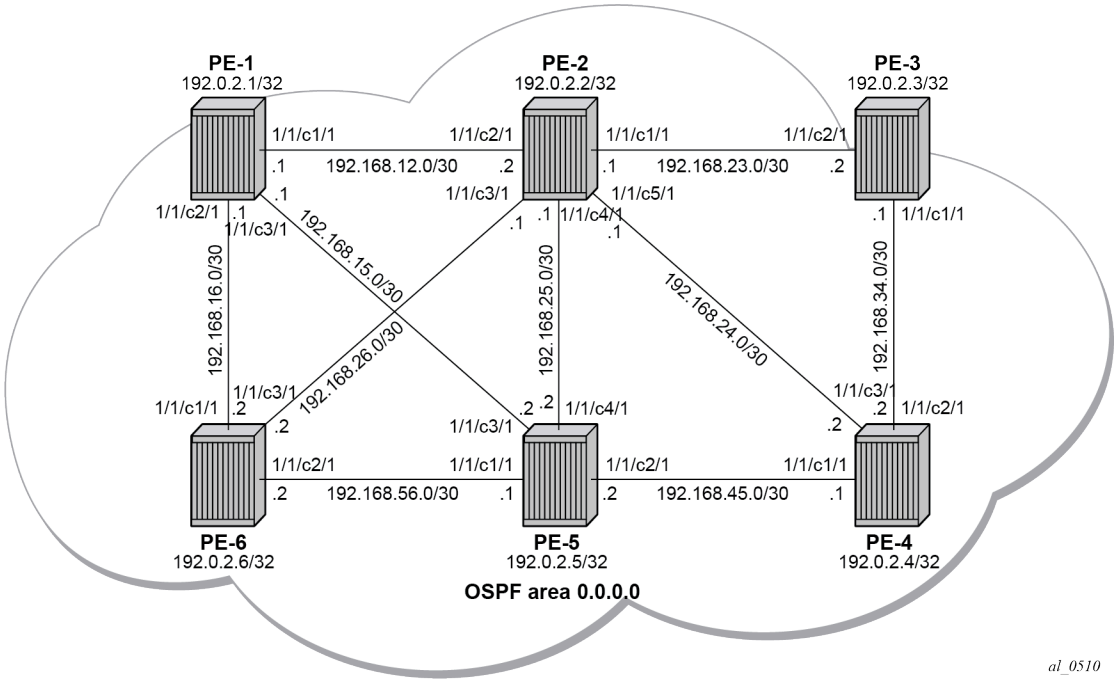
In the SR OS implementation, IP admin groups and SRLGs are locally significant, meaning they are not advertised by the IGP. Only the admin groups and SRLGs bound to an MPLS interface are advertised in TE link TLVs and sub-TLVs when the traffic engineering option is enabled in the IGP protocol. IES and VPRN interfaces do not have their attributes advertised in TE TLVs.

Other selection criteria which can be configured inside a route next hop template are protection type preference and next hop type preference. More details on these parameters are provided later in this chapter.

Configuration

Example topology shows the topology with six SR OS nodes. PE-2 will act as the point of local repair (PLR).

Figure 22: Example topology



1. Configure an IP/MPLS network with LDP FRR enabled on PE-2.

Because the focus is not on how to set up an IP/MPLS network, only summary bullets are provided.

- The system and IP interface addresses are configured according to Figure 22: Example topology.
- OSPF area 0.0.0.0 is selected as the interior gateway protocol (IGP) to distribute routing information between all PEs. All OSPF interfaces are set up as type point-to-point to avoid running the designated router/backup designated router (DR/BDR) election process. All links have an OSPF metric cost of 10, except for interface "int-PE-2-PE-5" on PE-2, which is configured with a metric of 20.
- Link LDP is enabled on all interfaces, which establishes a full mesh of LDP LSPs between all PE system interfaces. As an example, the tunnel table on PE-2 contains LDP tunnels to all the other PEs, as follows. The LDP LSP metric follows the IGP cost.

```
[/]
A:admin@PE-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   1
192.0.2.3/32          ldp       MPLS  65538     9      192.168.23.2   1
192.0.2.4/32          ldp       MPLS  65539     9      192.168.24.2   1
192.0.2.5/32          ldp       MPLS  65540     9      192.168.12.1   2
192.0.2.6/32          ldp       MPLS  65541     9      192.168.26.2   1
-------------------------------------------------------------------------------
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
===============================================================================
```

- Enable LDP FRR on PE-2. This is a two-fold configuration command: the IGP needs to be triggered to do LFA next hop computation, and FRR needs to be enabled within the **ldp** context. First, LFA is enabled in OSPF on PE-2:

```
# on PE-2:
configure {
    router "Base" {
        ospf 0 {
            loopfree-alternate {
            }

[/]
A:admin@PE-2# show router ospf status | match LFA
LFA                          : Enabled
Remote-LFA                   : Disabled
Max PQ Cost (Remote-LFA)     : 65535
Remote-LFA (node-protect)    : Disabled
TI-LFA                       : Disabled
TI-LFA (node-protect)        : Disabled
Mhp-LFA (IP-FRR)             : Disabled
Mhp-LFA (SR)                 : Disabled
```

Remote LFA and topology-independent LFA (TI-LFA) can be enabled for segment routing, but this is beyond the scope of this chapter.

Second, LDP FRR is enabled on PE-2:

```
# on PE-2:
configure {
    router "Base" {
        ldp {
            fast-reroute {
            }

[/]
A:admin@PE-2# show router ldp status | match FRR
FRR               : Enabled              Mcast Upstream FRR   : Disabled
Mcast Upst ASBR FRR: Disabled
```

Multicast upstream FRR is for multicast LDP and is beyond the scope of this chapter.

After issuing these two CLI commands, the software precomputes both a primary and a backup next hop label forwarding entry (NHLFE) for each LDP forwarding equivalence class (FEC) in the network and downloads them into the IOM/IMM. The primary NHLFE corresponds to the label of the FEC received from the primary next hop as per standard LDP resolution of the FEC prefix in the routing table manager (RTM). The backup NHLFE corresponds to the label received for the same FEC from an LFA next hop. The **show router route-table alternative** command adds an LFA flag to the associated alternative next hop for a specific destination prefix. Other useful IGP related show commands are **show router ospf lfa-coverage** and **show router ospf routes alternative detail**.

```
[/]
A:admin@PE-2# show router route-table alternative

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
      Alt-NextHop                                                 Alt-
                                                                  Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Remote  OSPF      00h02m41s  10
       192.168.12.1                                               1
       192.168.26.2 (LFA)                                         2
192.0.2.2/32                                    Local   Local     00h02m42s  0
       system                                                     0
192.0.2.3/32                                    Remote  OSPF      00h02m32s  10
       192.168.23.2                                               1
       192.168.24.2 (LFA)                                         2
192.0.2.4/32                                    Remote  OSPF      00h02m27s  10
       192.168.24.2                                               1
       192.168.23.2 (LFA)                                         2
192.0.2.5/32                                    Remote  OSPF      00h02m15s  10
       192.168.12.1                                               2
       192.168.24.2 (LFA)                                         2
192.0.2.6/32                                    Remote  OSPF      00h02m06s  10
       192.168.26.2                                               1
       192.168.12.1 (LFA)                                         2
192.168.12.0/30                                 Local   Local     00h02m42s  0
       int-PE-2-PE-1                                              0
192.168.15.0/30                                 Remote  OSPF      00h02m41s  10
       192.168.12.1                                               2
       192.168.26.2 (LFA)                                         3
192.168.16.0/30                                 Remote  OSPF      00h02m41s  10
       192.168.12.1                                               2
       192.168.26.2 (LFA)                                         3
```

```
192.168.23.0/30                                 Local   Local     00h02m42s  0
       int-PE-2-PE-3                                                  0
192.168.24.0/30                                 Local   Local     00h02m42s  0
       int-PE-2-PE-4                                                  0
192.168.25.0/30                                 Local   Local     00h02m42s  0
       int-PE-2-PE-5                                                  0
192.168.26.0/30                                 Local   Local     00h02m42s  0
       int-PE-2-PE-6                                                  0
192.168.34.0/30                                 Remote  OSPF      00h02m32s  10
       192.168.23.2                                                   2
       192.168.24.2 (LFA)                                             3
192.168.45.0/30                                 Remote  OSPF      00h02m27s  10
       192.168.24.2                                                   2
       192.168.23.2 (LFA)                                             3
192.168.56.0/30                                 Remote  OSPF      00h02m06s  10
       192.168.26.2                                                   2
       192.168.12.1 (LFA)                                             3
-------------------------------------------------------------------------------
No. of Routes: 16
Flags: n = Number of times nexthop is repeated
       Backup = BGP backup route
       LFA = Loop-Free Alternate nexthop
       S = Sticky ECMP requested
===============================================================================
```

Displaying the label forwarding information base (LFIB) on PE-2 shows the available alternate next hops that are displayed with the BU flag.

```
[/]
A:admin@PE-2# show router ldp bindings active prefixes 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
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op
IngLbl                                  EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.1/32                            Push
  --                                    524287
192.168.12.1                            1/1/c2/1:1000

192.0.2.1/32                            Push
  --                                    524286BU
192.168.26.2                            1/1/c3/1:1000

192.0.2.1/32                            Swap
524286                                  524287
```

```
192.168.12.1                            1/1/c2/1:1000

192.0.2.1/32                            Swap
524286                                  524286BU
192.168.26.2                            1/1/c3/1:1000

192.0.2.2/32                            Pop
524287                                    --
  --                                      --

192.0.2.3/32                            Push
  --                                    524287
192.168.23.2                            1/1/c1/1:1000

192.0.2.3/32                            Push
  --                                    524286BU
192.168.24.2                            1/1/c5/1:1000

192.0.2.3/32                            Swap
524285                                  524287
192.168.23.2                            1/1/c1/1:1000

192.0.2.3/32                            Swap
524285                                  524286BU
192.168.24.2                            1/1/c5/1:1000

192.0.2.4/32                            Push
  --                                    524287
192.168.24.2                            1/1/c5/1:1000

192.0.2.4/32                            Push
  --                                    524284BU
192.168.23.2                            1/1/c1/1:1000

192.0.2.4/32                            Swap
524284                                  524287
192.168.24.2                            1/1/c5/1:1000

192.0.2.4/32                            Swap
524284                                  524284BU
192.168.23.2                            1/1/c1/1:1000

192.0.2.5/32                            Push
  --                                    524283
192.168.12.1                            1/1/c2/1:1000

192.0.2.5/32                            Push
  --                                    524283BU
192.168.24.2                            1/1/c5/1:1000

192.0.2.5/32                            Swap
524283                                  524283
192.168.12.1                            1/1/c2/1:1000

192.0.2.5/32                            Swap
524283                                  524283BU
192.168.24.2                            1/1/c5/1:1000

192.0.2.6/32                            Push
  --                                    524287
192.168.26.2                            1/1/c3/1:1000

192.0.2.6/32                            Push
  --                                    524282BU
```

```
192.168.12.1                                1/1/c2/1:1000

192.0.2.6/32                                Swap
524282                                      524287
192.168.26.2                                1/1/c3/1:1000

192.0.2.6/32                                Swap
524282                                      524282BU
192.168.12.1                                1/1/c2/1:1000

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 21
===============================================================================
```

Finally, a synchronization timer is enabled between the IGP and LDP protocol when LDP FRR is enabled. From the moment that the interface for the previous primary next hop is restored, the IGP may reconverge back to that interface before LDP has completed the FEC exchange with its neighbor over that interface. This may cause LDP to de-program the LFA next hop from the FEC and blackhole the traffic. In this example, a synchronization timer of 10 seconds is configured, as follows:

```
# on all PEs:
configure {
    router "Base" {
        interface <itf-name> {
            ldp-sync-timer {
                seconds 10
            }
```

When this timer is set, on restoring a failed interface, the IGP advertises this link into the network with an infinite metric for the duration of this timer. When the failed link is restored, the LDP synchronization timer is started, and LDP adjacencies are brought up over the restored link and a label exchange is completed between the peers. After the LDP synchronization timer expires, the normal metric is advertised into the network again.

At this point, everything is in place to start creating LFA policies to influence the calculated LFA next hops.

2. Create a route next hop policy template.

   This is a mandatory step in the context of LFA policies. The route next hop template name is 32 characters at maximum. Creating a route next hop policy is done in the following way:

```
configure {
    routing-options {
        route-next-hop-policy {
            template <template name>
```

   After a **commit** of a route next hop policy template, the IGP re-evaluates the template and schedules a new LFA SPF to recompute the LFA next hop for the prefixes associated with this template.

3. Configure admin group constraints in route next hop policy.

   Admin groups are optional in the context of LFA policies. First, configure a group name and a group value for each admin group locally on the router. Admin groups are configured as follows:

```
configure {
    routing-options {
        if-attribute {
            admin-group <group-name> {
                value <number>
```

```
            }
```

Second, configure the admin group membership of the IP interfaces (network, IES, or VPRN), as follows. Maximum 32 admin groups can be assigned to an IP interface in one command. The configured IP admin group membership applies to all levels or areas the interface is participating in.

```
configure {
    router "Base" {
        interface <itf-name> {
            if-attribute {
                admin-group ["group-name-1" "group-name-2" ... (up to 32 max)]

configure {
    service {
        vprn <svc-name> {
            interface <itf-name> {
                if-attribute {
                    admin-group ["group-name-1" "group-name-2" ... (up to 32 max)]

configure {
    service {
        ies <svc-name> {
            interface <itf-name> {
                if-attribute {
                    admin-group ["group-name-1" "group-name-2" ... (up to 32 max)]
```

Third, add the IP admin group constraints to the route next hop policy template one by one. The **include-group** statement instructs the LFA SPF selection algorithm to select a subset of LFA next hops among the links which belong to one or more of the specified admin groups. A link which does not belong to any of the admin groups is excluded. The **preference** option is used to provide a relative preference for the admin group selection. A lower preference value means that LFA SPF will first attempt to select an LFA backup next hop which is a member of the corresponding admin group. If none is found, then the admin group with the next higher preference value is evaluated. If no preference value is configured, then it is the least preferred with a default preference value of 255.

When evaluating multiple **include-group** statements having the same preference, any link which belongs to one or more of the included admin groups can be selected as an LFA next hop. There is no relative preference based on how many of those included admin groups the link is a member of.

The **exclude-group** command simply prunes all links belonging to the specified admin group before making the LFA backup next hop selection for a prefix. If the same group name is part of both include and exclude statements, the exclude statement takes precedence. In other words, the exclude statement can be viewed as having an implicit preference value of 0.

Configure the admin group constraints in the route next hop policy template with the following command:

```
configure {
    routing-options {
        route-next-hop-policy {
            template <template-name> {
                exclude-group <ip-admin-group-name>
                include-group <ip-admin-group-name> {
                    preference <preference>
                }
```

4. Configure SRLG constraints in route next hop policy.

SRLG constraints are optional in the context of LFA policies. First, configure a group name and group value of each SRLG group locally on the router. The penalty weight controls the likelihood of paths with links sharing SRLG values with a primary path being used by a bypass or detour LSP. The higher the penalty weight, the less desirable it is to use the link with an SRLG. SRLG constraints are configured as follows:

```
configure {
    routing-options {
        if-attribute {
             srlg-group <group-name> {
                 value <group-value>
                 penalty-weight <penalty-weight>        # default: 0
             }
```

Second, configure the SRLG group membership of the IP interfaces (network, IES, or VPRN), as follows. One SRLG group can be applied to an IP interface in the **srlg-group** command but the command can be applied multiple times. The configured IP SRLG group membership is applied in all levels or areas the interface is participating in.

```
configure {
    router "Base" {
        interface <itf-name> {
            if-attribute {
                srlg-group <group-name>

configure {
    service {
        vprn <svc-name> {
            interface <itf-name> {
                if-attribute {
                    srlg-group <group-name>

configure {
    service {
        ies <svc-name> {
            interface <itf-name> {
                if-attribute {
                    srlg-group <group-name>
```

Third, add IP SRLG group constraints to the route next hop policy template, as follows. When this command is applied to a prefix, the LFA SPF attempts to select an LFA next hop which uses an outgoing interface that does not participate in any of the SRLGs of the outgoing interface used by the primary next hop.

```
configure {
    routing-options {
        route-next-hop-policy {
            template <template-name> {
                srlg true
            }
```

5. Configure the protection type in route next hop policy.

   This is an optional step in the context of LFA policies. With the following command, the user can also select if link protection or node protection is preferred for IP prefixes and LDP FEC prefixes protected

by a backup LFA next hop. By default, node protection is chosen. The implementation falls back to link protection if no LFA next hop is found for node protection.

```
configure {
    routing-options {
        route-next-hop-policy {
            template <template-name> {
                protection-type {link|node}
            }
```

6. Configure the next hop preference type in route next hop policy.

   This is an optional step in the context of LFA policies. With the following command, the user can also select if tunnel backup next hop or IP backup next hop is preferred for IP prefixes and LDP FEC prefixes protected by a backup LFA next hop. By default, IP backup next hop is chosen. The implementation falls back to the other type (tunnel) if no LFA next hop of the preferred type is found.

```
configure {
    routing-options {
        route-next-hop-policy {
            template <template-name> {
                nh-type {ip|tunnel}
            }
```

7. Apply the route next hop policy template to an IP interface.

   When the route next hop policy is applied to an IP interface with one of the following commands, all prefixes using this interface as primary next hop take the selection criteria specified in Step 3, Step 4, Step 5, and Step 6 into account.

```
configure {
    router "Base" {
        ospf <ospf-instance> {
            area <area-id> {
                interface <itf-name> {
                    loopfree-alternate {
                        policy-map {
                            route-nh-template <template-name>

configure {
    router "Base" {
        ospf3 <ospf-instance> {
            area <area-id> {
                interface <itf-name> {
                    loopfree-alternate {
                        policy-map {
                            route-nh-template <template-name>

configure {
    service {
        vprn <svc-name> {
            ospf <ospf-instance> {
                area <area-id> {
                    interface <itf-name> {
                        loopfree-alternate {
                            policy-map {
                                route-nh-template <template-name>

configure {
    service {
        vprn <svc-name> {
```

```
            ospf3 <ospf-instance> {
                area <area-id> {
                    interface <itf-name> {
                        loopfree-alternate {
                            policy-map {
                                route-nh-template <template-name>
```

LFA policy examples

All the following examples focus on providing another LFA next hop for LDP FEC prefix 192.0.2.1/32 and 192.0.2.6/32 (the system IP addresses of PE-1 and PE-6), with PE-2 being the PLR.

See Figure 22: Example topology for the example topology.

The default LFA next hop (without policy) for LDP FEC prefix 192.0.2.1/32 is 192.168.26.2 on PE-6, as follows:

```
[/]
A:admin@PE-2# show router ldp bindings active prefixes prefix 192.0.2.1/32

===============================================================================
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
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op
IngLbl                                  EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.1/32                            Push
  --                                    524287
192.168.12.1                            1/1/c2/1:1000

192.0.2.1/32                            Push
  --                                    524285BU
192.168.26.2                            1/1/c3/1:1000

192.0.2.1/32                            Swap
524286                                  524287
192.168.12.1                            1/1/c2/1:1000

192.0.2.1/32                            Swap
524286                                  524285BU
192.168.26.2                            1/1/c3/1:1000

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 4
```

```
===============================================================================
```

The default LFA next hop for LDP FEC prefix 192.0.2.6/32 is 192.168.12.1 on PE-1, as follows:

```
[/]
A:admin@PE-2# show router ldp bindings active prefixes prefix 192.0.2.6/32

===============================================================================
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
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op
IngLbl                                  EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.6/32                            Push
  --                                    524287
192.168.26.2                            1/1/c3/1:1000

192.0.2.6/32                            Push
  --                                    524282BU
192.168.12.1                            1/1/c2/1:1000

192.0.2.6/32                            Swap
524282                                  524287
192.168.26.2                            1/1/c3/1:1000

192.0.2.6/32                            Swap
524282                                  524282BU
192.168.12.1                            1/1/c2/1:1000

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 4
===============================================================================
```

This default LFA next hop can be changed by adding specific selection criteria inside a route next hop policy template.

Example 1: LFA policy with admin group constraint

The objective is to force the LFA next hop for both LDP FEC prefixes to use the path between PE-2 and PE-5.

Define admin group "red" with value 1 and apply it to the IP interfaces "int-PE-2-PE-1" and "int-PE-2-PE-6":

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

```
configure {
    routing-options {
        if-attribute {
            admin-group "red" {
                value 1
            }
        }
    }
    router "Base" {
        interface "int-PE-2-PE-1" {
            if-attribute {
                admin-group ["red"]
            }
        }
        interface "int-PE-2-PE-6" {
            if-attribute {
                admin-group ["red"]
            }
        }
```

Define a route next hop policy template "LFA_NH_exclRed", which excludes IP admin group "red".

```
# on PE-2:
configure {
    routing-options {
        route-next-hop-policy {
            template "LFA_NH_exclRed" {
                exclude-group "red" { }
            }
```

Apply the policy to the OSPF interfaces toward PE-1 and PE-6:

```
# on PE-2:
configure {
    router "Base" {
        ospf 0 {
            area 0.0.0.0 {
                interface "int-PE-2-PE-1" {
                    loopfree-alternate {
                        policy-map {
                            route-nh-template "LFA_NH_exclRed"
                        }
                    }
                }
                interface "int-PE-2-PE-6" {
                    loopfree-alternate {
                        policy-map {
                            route-nh-template "LFA_NH_exclRed"
                        }
                    }
                }
            }
```

From the moment that the route next hop policy template "LFA_NH_exclRed" is applied to the OSPF interfaces toward PE-1 and PE-6, the LFA next hops for both LDP FEC prefixes change. They now both point to the IP interface from PE-2 to PE-5 as LFA backup next hop:

```
[/]
A:admin@PE-2# show router ldp bindings active prefixes prefix 192.0.2.1/32

===============================================================================
```

```
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
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op
IngLbl                                  EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.1/32                            Push
  --                                    524287
192.168.12.1                            1/1/c2/1:1000

192.0.2.1/32                            Push
  --                                    524286BU
192.168.25.2                            1/1/c4/1:1000

192.0.2.1/32                            Swap
524286                                  524287
192.168.12.1                            1/1/c2/1:1000

192.0.2.1/32                            Swap
524286                                  524286BU
192.168.25.2                            1/1/c4/1:1000

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 4
===============================================================================
```

```
[/]
A:admin@PE-2# show router ldp bindings active prefixes prefix 192.0.2.6/32

===============================================================================
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
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op
```

```
IngLbl                                    EgrLbl
EgrNextHop                                EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.6/32                              Push
  --                                      524287
192.168.26.2                              1/1/c3/1:1000

192.0.2.6/32                              Push
  --                                      524282BU
192.168.25.2                              1/1/c4/1:1000

192.0.2.6/32                              Swap
524282                                    524287
192.168.26.2                              1/1/c3/1:1000

192.0.2.6/32                              Swap
524282                                    524282BU
192.168.25.2                              1/1/c4/1:1000

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 4
===============================================================================
```

Example 2: LFA policy with SRLG constraint

The objective is to force the LFA next hop for both LDP FEC prefixes to use the path from PE-2 to PE-5.

Define SRLG group "blue" with value 2 and apply it to the IP interfaces "int-PE-2-PE-1" and "int-PE-2-PE-6".

```
# on PE-2:
configure {
    routing-options {
        if-attribute {
            srlg-group "blue" {
                value 2
            }
        }
    }
    router "Base" {
        interface "int-PE-2-PE-1" {
            if-attribute {
                srlg-group "blue" { }
            }
        }
        interface "int-PE-2-PE-6" {
            if-attribute {
                srlg-group "blue" { }
            }
        }
```

Define a route next hop policy template "LFA_NH_SRLG", where SRLG is enabled, as follows:

```
# on PE-2:
configure {
    routing-options {
        route-next-hop-policy {
            template "LFA_NH_SRLG" {
                srlg true
            }
```

Apply the policy to the OSPF interface toward PE-1 and PE-6:

```
# on PE-2:
configure {
    router "Base" {
        ospf 0 {
            area 0.0.0.0 {
                interface "int-PE-2-PE-1" {
                    loopfree-alternate {
                        policy-map {
                            route-nh-template "LFA_NH_SRLG"
                        }
                    }
                }
                interface "int-PE-2-PE-6" {
                    loopfree-alternate {
                        policy-map {
                            route-nh-template "LFA_NH_SRLG"
                        }
                    }
                }
            }
```

Only one LFA policy mapping is allowed on an OSPF interface at a time. The new LFA policy mapping replaces the previous one.

The LFA next hops for both LDP FEC prefixes will both point now to the interface from PE-2 to PE-5 as LFA backup next hop, as follows:

```
[/]
A:admin@PE-2# show router ldp bindings active prefixes prefix 192.0.2.1/32

===============================================================================
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
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op
IngLbl                                  EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.1/32                            Push
  --                                    524287
192.168.12.1                            1/1/c2/1:1000

192.0.2.1/32                            Push
  --                                    524286BU
192.168.25.2                            1/1/c4/1:1000

192.0.2.1/32                            Swap
```

```
524286                                  524287
192.168.12.1                            1/1/c2/1:1000

192.0.2.1/32                            Swap
524286                                  524286BU
192.168.25.2                            1/1/c4/1:1000

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 4
===============================================================================
```

```
[/]
A:admin@PE-2# show router ldp bindings active prefixes prefix 192.0.2.6/32

===============================================================================
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
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op
IngLbl                                  EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.6/32                            Push
  --                                    524287
192.168.26.2                            1/1/c3/1:1000

192.0.2.6/32                            Push
  --                                    524282BU
192.168.25.2                            1/1/c4/1:1000

192.0.2.6/32                            Swap
524282                                  524287
192.168.26.2                            1/1/c3/1:1000

192.0.2.6/32                            Swap
524282                                  524282BU
192.168.25.2                            1/1/c4/1:1000

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 4
===============================================================================
```

The LFA policy mapping is removed from the OSPF interfaces as follows:

```
# on PE-2:
configure {
    router "Base" {
        ospf 0 {
            area 0.0.0.0 {
```

```
                interface "int-PE-2-PE-1" {
                    delete loopfree-alternate
                }
                interface "int-PE-2-PE-6" {
                    delete loopfree-alternate
                }
            }
```

Example 3: LFA policy with next hop type constraint

The objective is to force the LFA next hop for IP prefix 192.0.2.6/32 to use an RSVP tunnel.

Enable IP FRR as follows:

```
# on PE-2:
configure {
    routing-options {
        ip-fast-reroute true
```

Set up an RSVP LSP tunnel toward 192.0.2.6 with a strict MPLS path going over PE-2 to PE-4 to PE-5 to PE-6.

Note:
Because an RSVP LSP is set up between PE-2 and PE-6, MPLS and RSVP protocols need to be enabled on all the corresponding IP interfaces along the MPLS path.

```
# on PE-2:
configure {
    router "Base" {
        mpls {
            interface "int-PE-2-PE-4" {
            }
            path "path-PE-2-PE-4-PE-5-PE-6" {
                admin-state enable
                hop 10 {
                    ip-address 192.168.24.2
                    type strict
                }
                hop 20 {
                    ip-address 192.168.45.2
                    type strict
                }
                hop 30 {
                    ip-address 192.168.56.2
                    type strict
                }
            }
            lsp "LSP-PE-2-PE-6-strict" {
                admin-state enable
                type p2p-rsvp
                to 192.0.2.6
                primary "path-PE-2-PE-4-PE-5-PE-6" {
                }
            }
```

Enable IGP shortcut with resolution filter RSVP within the IGP on PE-2 and indicate that the newly created RSVP LSP is a possible shortcut candidate for LFA backup next hop only.

```
# on PE-2:
configure {
    router "Base" {
        ospf 0 {
            igp-shortcut {
                admin-state enable
                tunnel-next-hop {
                    family ipv4 {
                        resolution filter
                        resolution-filter {
                            rsvp true
                        }
                    }
                }
            }
        }
        mpls {
            lsp "LSP-PE-2-PE-6-strict" {
                igp-shortcut {
                    lfa-type lfa-only
                }
            }
        }
```

The following tunnel table on PE-2 for prefix 192.0.2.6 shows that an LDP LSP and an RSVP LSP are available toward PE-6:

```
[/]
A:admin@PE-2# show router tunnel-table 192.0.2.6

===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination           Owner     Encap TunnelId  Pref   Nexthop        Metric
   Color
-------------------------------------------------------------------------------
192.0.2.6/32          rsvp      MPLS  1         7      192.168.24.2   16777215
192.0.2.6/32 [L]      ldp       MPLS  65541     9      192.168.26.2   1
-------------------------------------------------------------------------------
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 RSVP tunnel with tunnel ID 1 corresponds to the RSVP LSP "LSP-PE-2-PE-6-strict", as follows:

```
[/]
A:admin@PE-2# show router mpls lsp

===============================================================================
MPLS LSPs (Originating)
===============================================================================
LSP Name                                          Tun     Fastfail  Adm  Opr
  To                                              Id      Config
-------------------------------------------------------------------------------
LSP-PE-2-PE-6-strict                              1       No        Up   Up
  192.0.2.6
-------------------------------------------------------------------------------
```

```
LSPs : 1
===============================================================================
```

By default, the preferred next hop type is IP, not tunnel. Therefore, the RSVP tunnel will not be used for the LFA backup, as follows:

```
[/]
A:admin@PE-2# show router route-table alternative 192.0.2.6/32

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
      Alt-NextHop                                                 Alt-
                                                                  Metric
-------------------------------------------------------------------------------
192.0.2.6/32                                    Remote  OSPF      00h00m22s  10
       192.168.26.2                                                 1
       192.168.12.1 (LFA)                                           2
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       Backup = BGP backup route
       LFA = Loop-Free Alternate nexthop
       S = Sticky ECMP requested
===============================================================================
```

Define a route next hop policy template "LFA_NH_Tunnel", where the next hop type is set to tunnel.

```
# on PE-2:
configure {
    routing-options {
        route-next-hop-policy {
            template "LFA_NH_Tunnel" {
                nh-type tunnel
            }
```

Apply the route next hop policy template to the OSPF interface toward PE-6, as follows:

```
# on PE-2:
configure {
    router "Base" {
        ospf 0 {
            area 0.0.0.0 {
                interface "int-PE-2-PE-6" {
                    loopfree-alternate {
                        policy-map {
                            route-nh-template "LFA_NH_Tunnel"
                        }
                    }
```

The LFA next hop uses the RSVP tunnel. The reference to the RSVP tunnel ID 1 in the following show output corresponds with the tunnel ID shown in the preceding **show router tunnel-table 192.0.2.6** output:

```
[/]
A:admin@PE-2# show router route-table alternative 192.0.2.6/32

===============================================================================
Route Table (Router: Base)
```

```
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
      Alt-NextHop                                                 Alt-
                                                                  Metric
-------------------------------------------------------------------------------
192.0.2.6/32                                  Remote  OSPF      00h00m38s  10
       192.168.26.2                                                 1
       192.0.2.6 (LFA) (tunneled:RSVP:1)                            65535
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       Backup = BGP backup route
       LFA = Loop-Free Alternate nexthop
       S = Sticky ECMP requested
===============================================================================
```

The following command shows the FIB next hop summary:

```
[/]
A:admin@PE-2# show router fib 1 nh-table-usage

===============================================================================
FIB Next-Hop Summary
===============================================================================
IPv4/IPv6                     Active                  Available
-------------------------------------------------------------------------------
IP Next-Hop                   9                       65535
Tunnel Next-Hop               1                       993279
ECMP Next-Hop                 0                       512000
ECMP Tunnel Next-Hop          0                       261120
===============================================================================
```

Example 4: Exclude prefix from LFA computation

The objective is to force no LFA next hop for LDP FEC prefix 192.0.2.1/32 where PE-2 is the PLR.

The IP FRR and LDP FRR implementation in SR OS allows to exclude an IGP interface, IGP area (OSPF), or IGP level (IS-IS) from the LFA SPF computation. The user can also exclude specific prefixes from the LFA SPF by using prefix lists and policy statements, which is configured as follows:

```
# on PE-2:
configure {
    policy-options {
        prefix-list "lo0-PE-1" {
            prefix 192.0.2.1/32 type exact {
            }
        }
        policy-statement "LFA_Exclude_PE-1" {
            entry 10 {
                from {
                    prefix-list ["lo0-PE-1"]
                }
                action {
                    action-type accept
                }
            }
        }
```

The configured policy statement is applied to the IGP protocol, as follows:

```
# on PE-2:
configure {
    router "Base" {
        ospf 0 {
            loopfree-alternate {
                exclude {
                    prefix-policy ["LFA_Exclude_PE-1"]
                }
            }
```

From the moment that it is applied, the existing LFA next hop entries for LDP FEC prefix 192.0.2.5/32 disappear instantly (compare with the preceding example 1):

```
[/]
A:admin@PE-2# show router ldp bindings active prefixes prefix 192.0.2.1/32

===============================================================================
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
        (S) - Static           (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
        (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
        (C) - FEC resolved with class-based-forwarding
===============================================================================
LDP IPv4 Prefix Bindings (Active)
===============================================================================
Prefix                                  Op
IngLbl                                  EgrLbl
EgrNextHop                              EgrIf/LspId
-------------------------------------------------------------------------------
192.0.2.1/32                            Push
  --                                    524287
192.168.12.1                            1/1/c2/1:1000

192.0.2.1/32                            Swap
524286                                  524287
192.168.12.1                            1/1/c2/1:1000

-------------------------------------------------------------------------------
No. of IPv4 Prefix Active Bindings: 2
===============================================================================
```

Conclusion

In production MPLS networks where IP FRR and/or LDP FRR are deployed, it is possible that the existing calculated LFA next hops are not always taking the most optimal or desirable paths.

With LFA policies, operators have better control on the way in which LFA backup next hops are computed.

Different selection criteria can be part of the route next hop policy: IP admin groups, IP SRLG groups, protection type preference, and next hop type preference.

PBR/PBF Redundancy

This chapter provides information about policy-based routing and policy-based forwarding redundancy.

Topics in this chapter include:

Applicability

This chapter was initially written based on SR OS Release 14.0.R7, but the MD-CLI in the current edition corresponds to SR OS Release 23.7.R1. Secondary actions in IPv4, IPv6, and MAC access control list (ACL) filter policies are supported in SR OS Release 14.0.R1, and later.

Overview

PBR and PBF

Policy-based routing (PBR) and policy-based forwarding (PBF) are used to make forwarding decisions based on filter policies defined by the network administrator. PBR is L3 traffic steering, whereas PBF is L2 traffic steering. For ordinary routing, the destination IP address is looked up in the routing table; for ordinary forwarding in a VPLS, the destination MAC address is looked up in the forwarding database (FDB). However, with PBR, routing decisions are based on IP filters that use more criteria, such as source and destination IP address, port number, DSCP value, and so on. Packets can take paths that differ from the next hop path specified by the routing table. PBF forwarding decisions can be made based on IP filters, but also on MAC filters that use criteria such as source and destination MAC address, inner and outer VLAN tag, dot1p priority, and so on.

The benefits of PBR/PBF are the following:

- The forwarding decision can be based on multiple attributes of a packet, not only its destination address
- Different QoS treatment can be provided, based on additional criteria
- Cost saving: time-sensitive traffic can be sent over higher-speed links at a higher cost, while bulk file transfers are sent over lower-speed links at a lower cost
- Load sharing: traffic can be load balanced across multiple and unequal paths

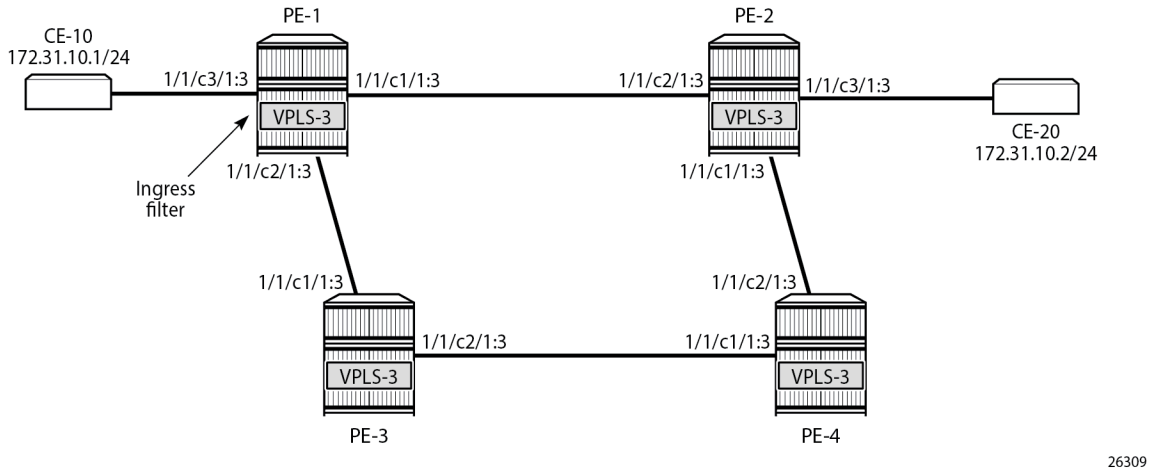
In most situations, PBR/PBF works on inbound unicast packets; therefore, a filter is applied at the ingress of access or network interfaces. In this chapter, examples will be shown for IPv4 filters and MAC filters applied on SAP ingress. IPv6 filters are also supported, but the examples in this chapter are based on IPv4. Filters are also supported on the egress, but that is beyond the scope of this chapter.

An IPv4 filter contains one or more entries, which can be configured with the following command:

```
[ex:/configure filter ip-filter "IP-1"]
A:admin@PE-1# entry 10 ?

 entry

 Immutable fields      - egress-pbr

 action                + Enable the action context
 apply-groups          - Apply a configuration group at this level
 apply-groups-exclude  - Exclude a configuration group at this level
 description           - Text description
 egress-pbr            - PBR that has an effect when this filter is applied on egress
 filter-sample         - Sample matching traffic if IP interface is set to cflowd ACL mode
 interface-sample      - Sample matching traffic if IP interface is set to cflowd interface
 mode
 log                   - Log that is used for packets matching this entry
 match                 + Enter the match context
 pbr-down-action-      - Action when PBR or PBF target for this entry is not available
  override
 sample-profile        - Cflowd sample profile ID to match packets
 sticky-dest           - Time before action with available PBR or PBF destination and highest
                         priority
```

Figure 23: PBF in the "VPLS-3" service on PE-1 shows the example topology with the "VPLS-3" service configured on the PEs. PBF is applied in the "VPLS-3" service on PE-1.

Figure 23: PBF in the "VPLS-3" service on PE-1



The following configuration creates an IPv4 filter that forwards all packets matching the source and destination IPv4 addresses, 172.31.10.1/24 and 172.31.10.2/24 respectively, to SAP 1/1/c1/1:3. When SAP 1/1/c1/1:3 is operationally down, the default behavior is to drop the packet. Not every IPv4/v6 filter needs to have match criteria defined, but in this case, only packets with the configured IPv4 SA and IPv4 DA are affected, whereas the other packets are forwarded per the FDB in the "VPLS-3" service on PE-1.

```
configure {
    filter {
        ip-filter "IP-1" {
            filter-id 1
            entry 10 {
```

```
                match {
                    src-ip {
                        address 172.31.10.1
                        mask 255.255.255.0
                    }
                    dst-ip {
                        address 172.31.10.2
                        mask 255.255.255.0
                    }
                }
                action {
                    forward {
                        sap {
                            vpls "VPLS-3"
                            sap-id 1/1/c1/1:3
                        }
                    }
                }
            }
```

In a similar way, an entry in a MAC filter can be configured with the following command:

```
[ex:/configure filter mac-filter "MAC-2" entry 10]
A:admin@PE-1# ?

 action                + Enable the action context
 apply-groups          - Apply a configuration group at this level
 apply-groups-exclude  - Exclude a configuration group at this level
 description           - Text description
 log                   - Log that is used for packets matching this entry
 match                 + Enter the match context
 pbr-down-action-      - Action when PBR or PBF target for this entry is not available
  override
 sticky-dest           - Time before action with available PBR or PBF destination and highest
                         priority
```

The following MAC filter forwards all frames with source MAC SA 00:00:5e:00:53:01 to SAP 1/1/c1/1:3:

```
configure {
    filter {
        mac-filter "MAC-2" {
            filter-id 2
            entry 10 {
                match {
                    src-mac {
                        address 00:00:5e:00:53:01
                    }
                }
                action {
                    forward {
                        sap {
                            vpls "VPLS-3"
                            sap-id 1/1/c1/1:3
                        }
                    }
                }
            }
```

Instead of defining a specific MAC address, a range of MAC addresses can be defined using a mask. The default mask is all 1s, ff:ff:ff:ff:ff:ff (not shown), which corresponds to an exact match of the configured MAC address.

When the primary SAP 1/1/c1/1:3 is down, the default action is drop. However, PBR/PBF redundancy can be configured, as described in the following section.

PBR/PBF redundancy

PBR/PBF redundancy is supported for MAC filters, IPv4 filters, and IPv6 filters. Within each entry in the IP/MAC filter, a secondary action can be configured; for example, for entry 10 in IPv4 filter "IP-1", as follows:

```
configure {
    filter {
        ip-filter "IP-1" {
            filter-id 1
            entry 10 {
                match {
                    src-ip {
                        address 172.31.10.1
                        mask 255.255.255.0
                    }
                    dst-ip {
                        address 172.31.10.2
                        mask 255.255.255.0
                    }
                }
                action {
                    forward {
                        sap {
                            vpls "VPLS-3"
                            sap-id 1/1/c1/1:3
                        }
                    }
                    secondary {
                        forward {
                            sap {
                                vpls "VPLS-3"
                                sap-id 1/1/c2/1:3
                            }
                        }
                    }
                }
            }
```

The IPv4 filter is applied on the ingress of SAP 1/1/c3/1:3 in the "VPLS-3" service on PE-1. This IPv4 filter only affects packets with IPv4 SA 172.31.10.1/24 and IPv4 DA 172.31.10.2/24. When the primary action SAP 1/1/c1/1:3 is operationally up, the primary action is executed; when SAP 1/1/c1/1:3 is operationally down, the secondary action is executed, until SAP 1/1/c1/1:3 is operationally up again. When both SAPs are down, the default behavior is to drop the packet.

When the primary action SAP 1/1/c1/1:3 is operationally up (PBR Target Status: Up), the primary action is executed (Downloaded Action: Primary), as follows:

```
[/]
A:admin@PE-1# show filter ip "IP-1"

===============================================================================
IP Filter
===============================================================================
Filter Id           : 1                             Applied        : Yes
Scope               : Template                      Def. Action    : Drop
Type                : Normal
```

```
Shared Policer      : Off
System filter       : Unchained
Radius Ins Pt       : n/a
CrCtl. Ins Pt       : n/a
RadSh. Ins Pt       : n/a
PccRl. Ins Pt       : n/a
Entries             : 1
Description         : (Not Specified)
Filter Name         : IP-1
-------------------------------------------------------------------------------
Filter Match Criteria : IP
-------------------------------------------------------------------------------
Entry               : 10
Description         : (Not Specified)
Log Id              : n/a
Src. IP             : 172.31.10.1/24
Src. Port           : n/a
Dest. IP            : 172.31.10.2/24
Dest. Port          : n/a
Protocol            : Undefined
Dscp                : Undefined
ICMP Type           : Undefined                    ICMP Code      : Undefined
Fragment            : Off                          Src Route Opt  : Off
Sampling            : Off                          Int. Sampling  : On
IP-Option           : 0/0                          Multiple Option: Off
Tcp-flag            : (Not Specified)
Option-pres         : Off
Egress PBR          : Disabled
Primary Action      : Forward (SAP)
  Next Hop          : 1/1/c1/1:3
  Service Id        : 3
  PBR Target Status : Up
Secondary Action    : Forward (SAP)
  Next Hop          : 1/1/c2/1:3
  Service Id        : 3
  PBR Target Status : Up
PBR Down Action     : Drop (entry-default)
Downloaded Action   : Primary
Dest. Stickiness    : None                         Hold Remain    : 0
Ing. Matches        : 205 pkts (21730 bytes)
Egr. Matches        : 0 pkts

===============================================================================
```

When the primary action SAP 1/1/c1/1:3 is operationally down, the secondary action is executed. When SAP 1/1/c1/1:3 is down, packets are forwarded to secondary action SAP 1/1/c2/1:3 instead. However, when the primary action SAP 1/1/c1/1:3 is operationally up again, the primary action is executed. This revertive behavior can be disabled by configuring stickiness in the filter entry, as follows:

```
[ex:/configure filter ip-filter "IP-1" entry 10]
A:admin@PE-1# sticky-dest ?

 sticky-dest (<number> | <keyword>)
 <number>  - <0..65535>      - seconds
 <keyword> - no-hold-time-up - seconds

    Time before action with available PBR or PBF destination and highest priority
```

When both the primary action SAP 1/1/c1/1:3 and the secondary action SAP 1/1/c2/1:3 are down, the default action is drop, unless the **pbr-down-action-override <filter-action>** parameter is configured. When the configured filter action is **forward**, the packets can be forwarded to another object in the service

that is up, for example, to another SAP or to an SDP binding, per the packet's destination address. This means that in a VPLS (PBF), the MAC DA is looked up in the FDB; in a VPRN (PBR), the IP DA is looked up in the routing table. The configuration of the **pbr-down-action-override** parameter is as follows. No specific SAPs or SDP bindings need to be defined.

```
[ex:/configure filter ip-filter "IP-1" entry 10]
A:admin@PE-1# pbr-down-action-override ?

 pbr-down-action-override <keyword>
 <keyword> - (drop|forward|filter-default-action)

    Action when PBR or PBF target for this entry is not available
```

In the example, the filter "IP-1" contains two actions that both forward packets to a SAP, but the PBR/PBF target can also be an SDP binding or—for PBR—a next-hop IP address in a VPRN. Table 1: Primary and secondary forwarding actions shows the allowed primary and secondary forwarding action combinations within a filter entry.

Table 1: Primary and secondary forwarding actions

primary forwarding action	secondary forwarding action
sap <sap-id>	sap <sap-id>
sap <sap-id>	sdp <sdp-id:vc-id>
sdp <sdp-id:vc-id>	sdp <sdp-id:vc-id>
sdp <sdp-id:vc-id>	sap <sap-id>
next-hop <ipv4/ipv6-address> router <router-instance>	next-hop <ipv4-ipv6-address> router <router-instance>
next-hop indirect <ipv4/ipv6-address> router <router-instance>	next-hop indirect <ipv4/ipv6-address> router <router-instance>

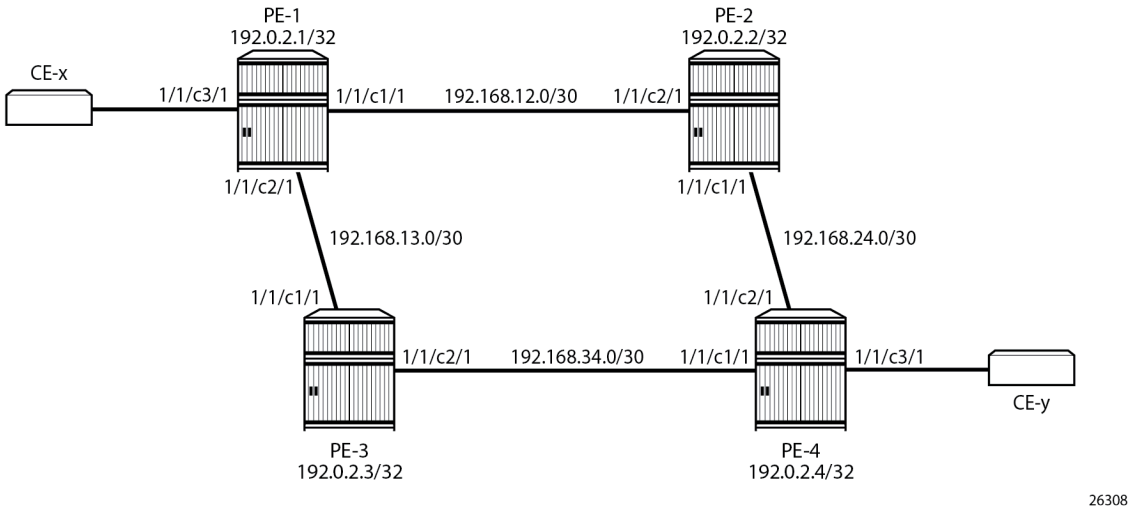
Configuration

In this section, the following examples are described:

- PBF in a VPLS using an IPv4 filter
- PBF in a VPLS using a MAC filter
- PBR in a VPRN using an IPv4 filter

Figure 24: Example topology shows the example topology with four PEs and two CEs.

Figure 24: Example topology



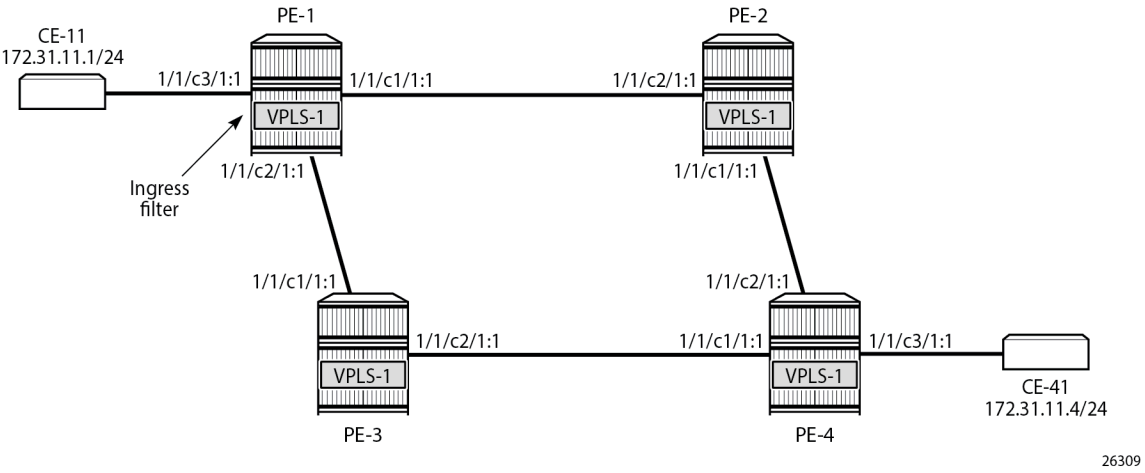
The initial configuration is as follows:

- Cards, MDAs, ports (all ports are in hybrid mode with dot1q encapsulation)
- Router interfaces
- IS-IS as IGP between the PEs (alternatively, OSPF could be configured as IGP)
- LDP between the PEs
- The CEs are emulated using a VPRN on PE-1 or PE-4 with a hairpin to loop the traffic back to the PE.

PBF in a VPLS using an IP filter

Figure 25: PBF in the "VPLS-1" service on PE-1 shows the example topology with the "VPLS-1" service configured on the four PEs. CE-11 is connected with the "VPLS-1" service on PE-1 and CE-14 with the "VPLS-1" service on PE-4. PBF is applied in the "VPLS-1" service on PE-1.

Figure 25: PBF in the "VPLS-1" service on PE-1



The configuration is shown for PE-1. The following cases are described in this section:

1. Initial situation: primary action is executed.
2. Primary action SAP 1/1/c1/1:1 is disabled. The secondary action in the entry in the IPv4 filter is executed.
3. Both primary and secondary action SAPs 1/1/c1/1:1 and 1/1/c2/1:1 are disabled. The default action is drop.
4. Both primary and secondary action SAPs 1/1/c1/1:1 and 1/1/c2/1:1 are disabled. The **pbr-down-action-override** parameter is configured with action *forward*.
5. The secondary action SAP 1/1/c2/1:1 is re-enabled. The secondary action is executed.
6. The primary action SAP 1/1/c1/1:1 is re-enabled. The primary action is executed.
7. Stickiness is configured with a hold timer of, for example, 120 seconds. At timer expiry, stickiness takes effect. If SAP 1/1/c1/1:1 is up at timer expiry, the primary action is programmed; otherwise, if SAP 1/1/c2/1:1 is up, the secondary action is programmed.
8. Stickiness is configured without a hold timer and takes effect immediately.

Configure the "VPLS-1" service with IPv4 filter on SAP ingress

IPv4 filter 10 has one entry with primary action to forward to SAP 1/1/c1/1:1 and secondary action to forward to SAP 1/1/c2/1:1. No match criteria are defined. When all action forward SAPs are operationally down, the default action is drop. No stickiness is configured.

```
# on PE-1:
configure {
    filter {
        ip-filter "IP-10" {
            filter-id 10
            entry 10 {
                action {
                    forward {
                        sap {
                            vpls "VPLS-1"
```

```
                        sap-id 1/1/c1/1:1
                    }
                }
                secondary {
                    forward {
                        sap {
                            vpls "VPLS-1"
                            sap-id 1/1/c2/1:1
                        }
                    }
                }
            }
        }
```

The "VPLS-1" service on PE-1 is configured with three SAPs and two spoke-SDPs, as follows. IPv4 filter "IP-10" is configured on the ingress of SAP 1/1/c3/1:1 and applies to traffic originating from CE-11.

```
configure {
    service {
        sdp 12 {
            admin-state enable
            delivery-type mpls
            ldp true
            far-end {
                ip-address 192.0.2.2
            }
        }
        sdp 13 {
            admin-state enable
            delivery-type mpls
            ldp true
            far-end {
                ip-address 192.0.2.3
            }
        }
        vpls "VPLS-1" {
            admin-state enable
            service-id 1
            customer "1"
            spoke-sdp 12:1 {
            }
            spoke-sdp 13:1 {
            }
            sap 1/1/c1/1:1 {
            }
            sap 1/1/c2/1:1 {
            }
            sap 1/1/c3/1:1 {
                ingress {
                    filter {
                        ip "IP-10"
                    }
                }
            }
        }
```

When all SAPs are up, all packets from CE-11 enter SAP 1/1/c3/1:1 and are forwarded to primary action SAP 1/1/c1/1:1. No other traffic is sent and the number of packets received or sent on port 1/1/c1/1 will only slightly exceed the number of packets sent on the SAP, because of signaling between the PEs for

IS-IS and LDP. The port statistics are cleared for ports 1/1/c1/1 through 1/1/c3/1 on PE-1. CE-11 sends a series of ICMP echo requests and, afterward, the port statistics on PE-1 are verified.

```
[/]
A:admin@PE-1# ping 172.31.11.4 router-instance "CE-11" source-address 172.31.11.1
                                          interval 0.01 count 200 output-format summary
PING 172.31.11.4 56 data bytes
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!!!!!!!!!!!!!!!!!!!!
---- 172.31.11.4 PING Statistics ----
200 packets transmitted, 200 packets received, 0.00% packet loss
round-trip min = 2.20ms, avg = 2.80ms, max = 11.1ms, stddev = 0.966ms
```

```
[/]
A:admin@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                                 203                         21545
                                         205                         21743
===============================================================================

[/]
A:admin@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                                   3                           315
                                           1                           129
===============================================================================

[/]
A:admin@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                                 200                         21200
                                         200                         21200
===============================================================================
```

All traffic is forwarded from ingress SAP 1/1/c3/1:1 to SAP 1/1/c1/1:1 and the reply messages from SAP 1/1/c1/1:1 to SAP 1/1/c3/1:1. No packets are forwarded via SAP 1/1/c2/1:1.

When the primary action SAP 1/1/c1/1:1 is operationally up, the primary action is executed, as follows:

```
[/]
A:admin@PE-1# show filter ip "IP-10"

===============================================================================
```

```
IP Filter
===============================================================================
Filter Id           : 10                            Applied        : Yes
Scope               : Template                      Def. Action    : Drop
---snip---

-------------------------------------------------------------------------------
Filter Match Criteria : IP
-------------------------------------------------------------------------------
Entry               : 10
---snip---

Primary Action      : Forward (SAP)
  Next Hop          : 1/1/c1/1:1
  Service Id        : 1
  PBR Target Status : Up
Secondary Action    : Forward (SAP)
  Next Hop          : 1/1/c2/1:1
  Service Id        : 1
  PBR Target Status : Up
PBR Down Action     : Drop (entry-default)
Downloaded Action   : Primary
Dest. Stickiness    : None                          Hold Remain    : 0
Ing. Matches        : 203 pkts (21518 bytes)
Egr. Matches        : 0 pkts

===============================================================================
```

Primary action PBR target down

The primary action SAP 1/1/c1/1:1 is disabled. Therefore, the primary action cannot be executed, and the secondary action is executed instead. When CE-11 sends ICMP echo requests, all packets are forwarded to SAP 1/1/c2/1:1.

```
# Disable SAP 1/1/c1/1:1 in the "VPLS-1" service on PE-1:
configure {
    service {
        vpls "VPLS-1" {
            sap 1/1/c1/1:1
                admin-state disable

[/]
A:admin@PE-1# show filter ip "IP-10"

===============================================================================
IP Filter
===============================================================================
Filter Id           : 10                            Applied        : Yes
Scope               : Template                      Def. Action    : Drop
---snip---

Entry               : 10
---snip---

Primary Action      : Forward (SAP)
  Next Hop          : 1/1/c1/1:1
  Service Id        : 1
  PBR Target Status : Down
Secondary Action    : Forward (SAP)
  Next Hop          : 1/1/c2/1:1
  Service Id        : 1
```

```
  PBR Target Status : Up
PBR Down Action     : Drop (entry-default)
Downloaded Action   : Secondary
Dest. Stickiness    : None                          Hold Remain    : 0
Ing. Matches        : 403 pkts (42718 bytes)
Egr. Matches        : 0 pkts

===============================================================================
```

Secondary action PBR target down

The secondary action SAP 1/1/c2/1:1 is disabled, as follows:

```
# Disable SAP 1/1/c2/1:1 in the "VPLS-1" service on PE-1:
configure {
    service {
        vpls "VPLS-1" {
            sap 1/1/c2/1:1
                admin-state disable
```

Both SAP 1/1/c1/1:1 and SAP 1/1/c2/1:1 are disabled. Neither the primary nor the secondary action in entry 10 of IPv4 filter 10 can be executed. Therefore, the default action is executed, which is drop; see the following output (PBR Down Action: Drop (entry-default)).

```
[/]
A:admin@PE-1# show filter ip "IP-10"

===============================================================================
IP Filter
===============================================================================
Filter Id           : 10                            Applied        : Yes
Scope               : Template                      Def. Action    : Drop
---snip---

Entry               : 10
---snip---

Primary Action      : Forward (SAP)
  Next Hop          : 1/1/c1/1:1
  Service Id        : 1
  PBR Target Status : Down
Secondary Action    : Forward (SAP)
  Next Hop          : 1/1/c2/1:1
  Service Id        : 1
  PBR Target Status : Down
PBR Down Action     : Drop (entry-default)
Downloaded Action   : Primary
Dest. Stickiness    : None                          Hold Remain    : 0
Ing. Matches        : 403 pkts (42718 bytes)
Egr. Matches        : 0 pkts

===============================================================================
```

When CE-11 sends ICMP echo requests, they are all dropped.

```
[/]
A:admin@PE-1# ping 172.31.11.4 router-instance "CE-11" source-address 172.31.11.1
                                          interval 0.01 count 50 output-format summary
PING 172.31.11.4 56 data bytes
```

```
.................................................
---- 172.31.11.4 PING Statistics ----
50 packets transmitted, 0 packets received, 100% packet loss
```

PBR down action override

Both SAPs remain disabled. The default PBR down action is drop, but that can be overruled by configuring the **pbr-down-action-override** parameter, as follows:

```
# on PE-1:
configure {
    filter {
        ip-filter "IP-10" {
            entry 10 {
                pbr-down-action-override forward
```

With this configuration added in entry 10 of the "IP-10" filter, the PBR down action will be forward. No specific next hop needs to be defined. The forwarding is based on the destination address. When CE-11 sends ICMP echo requests to CE-41, the traffic is forwarded, as follows:

```
[/]
A:admin@PE-1# ping 172.31.11.4 router-instance "CE-11" source-address 172.31.11.1
                                          interval 0.01 count 200 output-format summary
PING 172.31.11.4 56 data bytes
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
---snip---
---- 172.31.11.4 PING Statistics ----
200 packets transmitted, 200 packets received, 1 duplicate
round-trip min = 2.29ms, avg = 2.94ms, max = 12.0ms, stddev = 0.752ms
```

The statistics in the detailed output for spoke-SDP 12:1 in the "VPLS-1" service shows that these packets have been sent over this spoke-SDP. It is possible that spoke-SDP 13:1 in the "VPLS-1" service is used instead.

```
[/]
A:admin@PE-1# show service id 1 sdp 12:1 detail | match Statistics post-lines 5
Statistics            :
I. Fwd. Pkts.      : 203                      I. Dro. Pkts.     : 0
I. Fwd. Octs.      : 19818                    I. Dro. Octs.     : 0
E. Fwd. Pkts.      : 207                      E. Fwd. Octets    : 20020
-------------------------------------------------------------------------------
```

The PBR down action for entry 10 in IPv4 filter 10 is forward, as defined by the **pbr-down-action-override** parameter, and the PBR downloaded action is forward, as follows:

```
[/]
A:admin@PE-1# show filter ip "IP-10"

===============================================================================
IP Filter
===============================================================================
Filter Id           : 10                           Applied        : Yes
Scope               : Template                     Def. Action    : Drop
---snip---

Entry               : 10
---snip---
```

```
Primary Action      : Forward (SAP)
  Next Hop          : 1/1/c1/1:1
  Service Id        : 1
  PBR Target Status : Down
Secondary Action    : Forward (SAP)
  Next Hop          : 1/1/c2/1:1
  Service Id        : 1
  PBR Target Status : Down
PBR Down Action     : Forward (pbr-down-action-override)
Downloaded Action   : Forward
Dest. Stickiness    : None                          Hold Remain    : 0
Ing. Matches        : 653 pkts (69218 bytes)
Egr. Matches        : 0 pkts

===============================================================================
```

Secondary action up - revertive behavior

The primary action SAP 1/1/c1/1:1 remains disabled, whereas secondary action SAP 1/1/c2/1:1 is re-enabled, as follows:

```
# on PE-1:
configure {
    service {
        vpls "VPLS-1" {
            sap 1/1/c2/1:1 {
                admin-state enable
```

The secondary action in entry 10 of IPv4 filter 10 is executed (Downloaded Action: Secondary), as follows:

```
[/]
A:admin@PE-1# show filter ip "IP-10"

===============================================================================
IP Filter
===============================================================================
Filter Id           : 10                            Applied        : Yes
Scope               : Template                      Def. Action    : Drop
---snip---

Entry               : 10
---snip---

Primary Action      : Forward (SAP)
  Next Hop          : 1/1/c1/1:1
  Service Id        : 1
  PBR Target Status : Down
Secondary Action    : Forward (SAP)
  Next Hop          : 1/1/c2/1:1
  Service Id        : 1
  PBR Target Status : Up
PBR Down Action     : Forward (pbr-down-action-override)
Downloaded Action   : Secondary
Dest. Stickiness    : None                          Hold Remain    : 0
Ing. Matches        : 853 pkts (90418 bytes)
Egr. Matches        : 0 pkts

===============================================================================
```

Primary action up - revertive behavior

As well as the secondary action SAP, also the primary action SAP 1/1/c1/1:1 is re-enabled, as follows:

```
# on PE-1:
configure {
    service {
        vpls "VPLS-1" {
            sap 1/1/c1/1:1 {
                admin-state enable
```

The default PBR/PBF behavior is revertive; therefore, the primary action is executed: the packets are forwarded to SAP 1/1/c1/1:1, as follows:

```
[/]
A:admin@PE-1# show filter ip "IP-10"

===============================================================================
IP Filter
===============================================================================
Filter Id           : 10                          Applied        : Yes
Scope               : Template                    Def. Action    : Drop
---snip---

Entry               : 10
---snip---

Primary Action      : Forward (SAP)
  Next Hop          : 1/1/c1/1:1
  Service Id        : 1
  PBR Target Status : Up
Secondary Action    : Forward (SAP)
  Next Hop          : 1/1/c2/1:1
  Service Id        : 1
  PBR Target Status : Up
PBR Down Action     : Forward (pbr-down-action-override)
Downloaded Action   : Primary
Dest. Stickiness    : None                        Hold Remain    : 0
Ing. Matches        : 1053 pkts (111618 bytes)
Egr. Matches        : 0 pkts

===============================================================================
```

Stickiness in IP filter with hold timer

When the primary action SAP becomes up, traffic will be forwarded to this SAP instantaneously, unless stickiness applies. Stickiness can be defined on the IPv4/v6 filter entry level to override this revertive behavior. The following command enables stickiness at timer expiry with a hold remain timer of—in this case—120 seconds for entry 10 in IPv4 filter 10:

```
# on PE-1:
configure {
    filter {
        ip-filter "IP-10" {
            entry 10 {
                sticky-dest 120
```

The hold remain timer starts counting down when stickiness is configured and at least one PBR target is up. If the primary action SAP 1/1/c1/1:1 remains operationally up for the configured 120 seconds, the primary action will be active, and at timer expiry, stickiness applies. However, if SAP 1/1/c1/1:1 goes down and then up again before timer expiry, the secondary action remains active until the hold remain timer expires, as shown in the following example.

The hold remain timer has not expired. The primary action SAP 1/1/c1/1:1 is disabled, so the secondary action is active, as follows. The hold remain timer keeps counting down.

```
# on PE-1:
configure exclusive
    service {
        vpls "VPLS-1" {
            sap 1/1/c1/1:1 {
                admin-state disable
```

```
[/]
A:admin@PE-1# show filter ip "IP-10"

===============================================================================
IP Filter
===============================================================================
Filter Id           : 10                          Applied        : Yes
Scope               : Template                    Def. Action    : Drop
---snip---

Entry               : 10
---snip---

Primary Action      : Forward (SAP)
  Next Hop          : 1/1/c1/1:1
  Service Id        : 1
  PBR Target Status : Down
Secondary Action    : Forward (SAP)
  Next Hop          : 1/1/c2/1:1
  Service Id        : 1
  PBR Target Status : Up
PBR Down Action     : Forward (pbr-down-action-override)
Downloaded Action   : Secondary
Dest. Stickiness    : 120                         Hold Remain    : 91
Ing. Matches        : 1253 pkts (132818 bytes)
Egr. Matches        : 0 pkts

===============================================================================
```

The primary action SAP 1/1/c1/1:1 is restored and the secondary action is active until the hold remain timer expires, as follows:

```
# on PE-1:
configure {
    service {
        vpls "VPLS-1" {
            sap 1/1/c1/1:1 {
                admin-state enable


[/]
A:admin@PE-1# show filter ip "IP-10"

===============================================================================
IP Filter
```

```
===============================================================================
Filter Id           : 10                          Applied        : Yes
Scope               : Template                    Def. Action    : Drop
---snip---

Entry               : 10
---snip---

Primary Action      : Forward (SAP)
  Next Hop          : 1/1/c1/1:1
  Service Id        : 1
  PBR Target Status : Up
Secondary Action    : Forward (SAP)
  Next Hop          : 1/1/c2/1:1
  Service Id        : 1
  PBR Target Status : Up
PBR Down Action     : Forward (pbr-down-action-override)
Downloaded Action   : Secondary
Dest. Stickiness    : 120                         Hold Remain    : 55
Ing. Matches        : 1453 pkts (154018 bytes)
Egr. Matches        : 0 pkts

===============================================================================
```

In the preceding output, the secondary action is active and the hold remain time is 55 seconds. When the hold remain timer expires and the primary action SAP 1/1/c1/1:1 is up, the primary action is activated again and stickiness applies from then onward, as follows:

```
[/]
A:admin@PE-1# show filter ip "IP-10"

===============================================================================
IP Filter
===============================================================================
Filter Id           : 10                          Applied        : Yes
Scope               : Template                    Def. Action    : Drop
---snip---

Primary Action      : Forward (SAP)
  Next Hop          : 1/1/c1/1:1
  Service Id        : 1
  PBR Target Status : Up
Secondary Action    : Forward (SAP)
  Next Hop          : 1/1/c2/1:1
  Service Id        : 1
  PBR Target Status : Up
PBR Down Action     : Forward (pbr-down-action-override)
Downloaded Action   : Primary
Dest. Stickiness    : 120                         Hold Remain    : 0
Ing. Matches        : 1453 pkts (154018 bytes)
Egr. Matches        : 0 pkts

===============================================================================
```

The hold remain timer stays at zero. When the primary action cannot be activated, the secondary action is activated and will remain activated even when the primary action SAP 1/1/c1/1:1 is up again. However, when the secondary action SAP 1/1/c2/1:1 is down, the primary action can be activated again.

The hold remain timer starts counting down when it is first configured, or reconfigured with a different value, and at least one of the PBR/PBF targets is up. The hold remain timer also starts counting down after both the primary and the secondary PBR/PBF targets have been down, for example, after a reboot, and at

least one of them transitions to the up status. The secondary action might be available first, even though the primary action is preferred. This situation is automatically resolved when the timer expires: the primary action will be activated if available when the hold remain timer expires.

Force primary action

Stickiness can be enabled without any delay, as follows:

```
# on PE-1:
configure exclusive
    filter {
        ip-filter "IP-10" {
            entry 10 {
                sticky-dest no-hold-time-up
```

Initially, the primary action was executed, but when the primary action SAP 1/1/c1/1:1 is disabled, the secondary action is executed, as follows:

```
# on PE-1:
configure {
    service {
        vpls "VPLS-1" {
            sap 1/1/c1/1:1
                admin-state disable
```

```
[/]
A:admin@PE-1# show filter ip "IP-10"

===============================================================================
IP Filter
===============================================================================
Filter Id           : 10                            Applied        : Yes
Scope               : Template                      Def. Action    : Drop
---snip---

Entry               : 10
---snip---

Primary Action      : Forward (SAP)
  Next Hop          : 1/1/c1/1:1
  Service Id        : 1
  PBR Target Status : Down
Secondary Action    : Forward (SAP)
  Next Hop          : 1/1/c2/1:1
  Service Id        : 1
  PBR Target Status : Up
PBR Down Action     : Forward (pbr-down-action-override)
Downloaded Action   : Secondary
Dest. Stickiness    : 0                             Hold Remain    : 0
Ing. Matches        : 1653 pkts (175218 bytes)
Egr. Matches        : 0 pkts

===============================================================================
```

The secondary action is active and will remain active as long as the secondary action SAP 1/1/c2/1:1 is up. The hold remain timer is not enabled (== value 0). When the primary action SAP 1/1/c1/1:1 is operationally up again, the secondary action remains active, as follows:

```
# on PE-1:
configure {
    service {
        vpls "VPLS-1" {
            sap 1/1/c1/1:1
                admin-state enable


[/]
A:admin@PE-1# show filter ip "IP-10"

===============================================================================
IP Filter
===============================================================================
Filter Id           : 10                            Applied        : Yes
Scope               : Template                      Def. Action    : Drop
---snip---

Entry               : 10
---snip---

Primary Action      : Forward (SAP)
  Next Hop          : 1/1/c1/1:1
  Service Id        : 1
  PBR Target Status : Up
Secondary Action    : Forward (SAP)
  Next Hop          : 1/1/c2/1:1
  Service Id        : 1
  PBR Target Status : Up
PBR Down Action     : Forward (pbr-down-action-override)
Downloaded Action   : Secondary
Dest. Stickiness    : 0                             Hold Remain    : 0
Ing. Matches        : 1853 pkts (196418 bytes)
Egr. Matches        : 0 pkts

===============================================================================
```

The following **tools** command forces activation of the primary action in entry 10 of the "IP-10" filter:

```
[/]
A:admin@PE-1# tools perform filter ip-filter 10 entry 10 activate-primary-action
```

The result is that the primary action is executed again, as shown in the following output:

```
[/]
A:admin@PE-1# show filter ip "IP-10"

===============================================================================
IP Filter
===============================================================================
Filter Id           : 10                            Applied        : Yes
Scope               : Template                      Def. Action    : Drop
---snip---

Entry               : 10
---ping---

Primary Action      : Forward (SAP)
```

```
  Next Hop          : 1/1/c1/1:1
  Service Id        : 1
  PBR Target Status : Up
Secondary Action    : Forward (SAP)
  Next Hop          : 1/1/c2/1:1
  Service Id        : 1
  PBR Target Status : Up
PBR Down Action     : Forward (pbr-down-action-override)
Downloaded Action   : Primary
Dest. Stickiness    : 0                             Hold Remain    : 0
Ing. Matches        : 2053 pkts (217618 bytes)
Egr. Matches        : 0 pkts

===============================================================================
```

This **tools** command can also be used in combination with a running sticky-destination hold remain timer. In that case, the hold remain timer will stop counting down and the primary action immediately reverts.

PBF in a VPLS using a MAC filter

PBF in a VPLS can use a MAC filter instead of an IPv4 filter, but not both. The following MAC filter is defined on PE-1:

```
configure exclusive
    filter {
        mac-filter "MAC-20" {
            filter-id 20
            entry 10 {
                pbr-down-action-override forward
                sticky-dest no-hold-time-up
                match {
                    src-mac {
                        address 00:00:5e:00:53:11
                    }
                }
                action {
                    forward {
                        sap {
                            vpls "VPLS-1"
                            sap-id 1/1/c1/1:1
                        }
                    }
                    secondary {
                        forward {
                            sap {
                                vpls "VPLS-1"
                                sap-id 1/1/c2/1:1
                            }
                        }
                    }
                }
            }
```

MAC filter "MAC-20" cannot be applied next to IPv4 filter "IP-10" on the ingress direction of SAP 1/1/c3/1:1 in the "VPLS-1" service; therefore, an error message is raised, as follows:

```
[ex:/configure service vpls "VPLS-1" sap 1/1/c3/1:1 ingress filter]
A:admin@PE-1# mac "MAC-20"

*[ex:/configure service vpls "VPLS-1" sap 1/1/c3/1:1 ingress filter]
```

```
A:admin@PE-1# info
    mac "MAC-20"
    ip "IP-10"

*[ex:/configure service vpls "VPLS-1" sap 1/1/c3/1:1 ingress filter]
A:admin@PE-1# commit
MINOR: SVCMGR #12: configure service vpls "VPLS-1" sap 1/1/c3/1:1 ingress filter mac
 - Inconsistent Value error - another filter is already configured
```

The filter that was applied must be removed first, then the MAC filter can be applied, as follows:

```
# on PE-1:
configure {
    service {
        vpls "VPLS-1" {
            sap 1/1/c3/1:1 {
                ingress {
                    delete filter            # remove filter
                    filter {
                        mac "MAC-20"
                    }
```

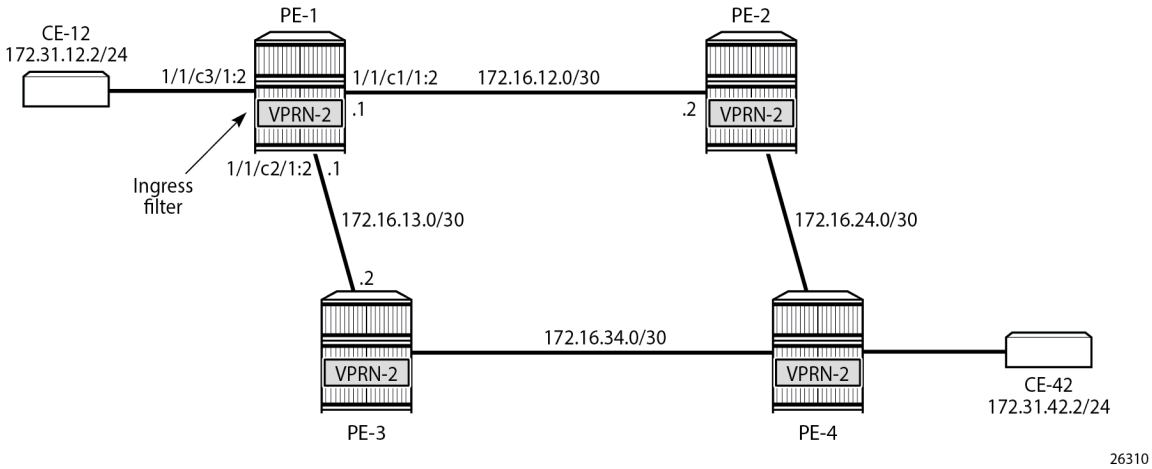
When all SAPs in the "VPLS-1" service are up, the primary action is activated, as follows:

```
[/]
A:admin@PE-1# show filter mac "MAC-20"

===============================================================================
Mac Filter
===============================================================================
Filter Id           : 20                           Applied        : Yes
Scope               : Template                     Def. Action    : Drop
Entries             : 1                            Type           : normal
Description         : (Not Specified)
Filter Name         : MAC-20
-------------------------------------------------------------------------------
Filter Match Criteria : Mac
-------------------------------------------------------------------------------
Entry               : 10                           FrameType      : Ethernet
Description         : (Not Specified)
Log Id              : n/a
Src Mac             : 00:00:5e:00:53:11 ff:ff:ff:ff:ff:ff
Dest Mac            : Undefined
Dot1p               : Undefined                    Ethertype      : Undefined
DSAP                : Undefined                    SSAP           : Undefined
Snap-pid            : Undefined                    ESnap-oui-zero : Undefined
Primary Action      : Forward (SAP)
  Next Hop          : 1/1/c1/1:1
  Service Id        : 1
  PBR Target Status : Up
Secondary Action    : Forward (SAP)
  Next Hop          : 1/1/c2/1:1
  Service Id        : 1
  PBR Target Status : Up
PBR Down Action     : Forward (pbr-down-action-override)
Downloaded Action   : Primary
Dest. Stickiness    : 0                            Hold Remain    : 0
Ing. Matches        : 200 pkts (21200 bytes)
Egr. Matches        : 0 pkts

===============================================================================
```

PBR in a VPRN using an IP filter

Figure 26: PBR in a VPRN shows the example topology used with the "VPRN-2" service configured on each PE and the CEs configured as another VPRN service on PE-1 and PE-4.

Figure 26: PBR in a VPRN



The following IPv4 filter is configured on PE-1:

```
configure {
    filter {
        ip-filter "IP-30" {
            filter-id 30
            entry 10 {
                action {
                    forward {
                        next-hop {
                            nh-ip-vrf {
                                router-instance "VPRN-2"
                                address 172.16.12.2
                            }
                        }
                    }
                    secondary {
                        forward {
                            next-hop {
                                nh-ip-vrf {
                                    router-instance "VPRN-2"
                                    address 172.16.13.2
                                }
                            }
                        }
                    }
                }
            }
```

The "VPRN-2" service in PE-1 has the "IP-30" filter applied to SAP 1/1/c3/1:2 toward CE-12:

```
configure {
```

```
    service {
        vprn "VPRN-2" {
            admin-state enable
            service-id 2
            customer "1"
            bgp-ipvpn {
                mpls {
                    admin-state enable
                    route-distinguisher "64496:2"
                }
            }
            interface "int-VPRN-2-PE-1-CE-12" {
                ipv4 {
                    primary {
                        address 172.31.12.1
                        prefix-length 24
                    }
                }
                sap 1/1/c3/1:2 {
                    ingress {
                        filter {
                            ip "IP-30"
                        }
                    }
                }
            }
            interface "int-VPRN-2-PE-1-PE-2" {
                ipv4 {
                    primary {
                        address 172.16.12.1
                        prefix-length 30
                    }
                }
                sap 1/1/c1/1:2 {
                }
            }
            interface "int-VPRN-2-PE-1-PE-3" {
                ipv4 {
                    primary {
                        address 172.16.13.1
                        prefix-length 30
                    }
                }
                sap 1/1/c2/1:2 {
                }
            }
        }
```

The configuration of the "VPRN-2" service on the remaining PEs is similar, except that static route entries are configured for subnets 172.31.12.0/24 (toward CE-12) and 172.31.42.0/24 (toward CE-42). No filters are applied to the "VPRN-2" service on the other nodes.

The primary action forwards packets from CE-12 to next-hop 172.16.12.2, which is an interface in the "VPRN-2" service on PE-2; the secondary action forwards to next-hop 172.16.13.2, an interface in the "VPRN-2" service on PE-3. When all interfaces are up, the primary action is executed and traffic from CE-12 to CE-42 is forwarded from the "VPRN-2" router on PE-1 to the "VPRN-2" router on PE-2 (next hop 172.16.12.2), as follows:

```
[/]
A:admin@PE-1# show filter ip "IP-30"

===============================================================================
```

```
IP Filter
===============================================================================
Filter Id           : 30                          Applied        : Yes
Scope               : Template                    Def. Action    : Drop
---snip---

Primary Action      : Forward (Next Hop VRF)
  Next Hop          : 172.16.12.2
  Router            : 2
  PBR Target Status : Up
  Extended Action   : None             # optional DSCP remarking for PBR
Secondary Action    : Forward (Next Hop VRF)
  Next Hop          : 172.16.13.2
  Router            : 2
  PBR Target Status : Up
  Extended Action   : None
PBR Down Action     : Drop (entry-default)
Downloaded Action   : Primary
Dest. Stickiness    : None                        Hold Remain    : 0
Ing. Matches        : 201 pkts (21306 bytes)
Egr. Matches        : 0 pkts

===============================================================================
```

The output includes an additional line per action: both the primary and the secondary action in PBR can have DSCP remarking as extended action, but that is not configured in this example. It can be configured using the following command; for example, for the primary action, as follows:

```
*[ex:/configure filter ip-filter "IP-30" entry 10 action remark]
A:admin@PE-1# dscp ?

 dscp <keyword>
 <keyword> - (be|cp1|cp2|cp3|cp4|cp5|cp6|cp7|cs1|cp9|af11|cp11|af12|cp13|af13|cp15|cs2|cp17|
af21|cp19|af22|
              cp21|af23|cp23|cs3|cp25|af31|cp27|af32|cp29|af33|cp31|cs4|cp33|af41|cp35|af42|
cp37|af43|cp39|
              cs5|cp41|cp42|cp43|cp44|cp45|ef|cp47|nc1|cp49|cp50|cp51|cp52|cp53|cp54|cp55|nc2|
cp57|cp58|
              cp59|cp60|cp61|cp62|cp63)

 'dscp' is: mandatory

    Destination SAP
```

When the primary action cannot be activated, the secondary action is activated, as follows:

```
# on PE-1:
configure {
    service {
        vprn "VPRN-2" {
            interface "int-VPRN-2-PE-1-PE-2" {
                sap 1/1/c1/1:2 {
                    admin-state disable
```

```
*A:PE-1# show filter ip "IP-30"

===============================================================================
IP Filter
===============================================================================
Filter Id           : 30                          Applied        : Yes
```

```
Scope               : Template                    Def. Action    : Drop
---snip---

Entry               : 10
---snip---

Primary Action      : Forward (Next Hop VRF)
  Next Hop          : 172.16.12.2
  Router            : 2
  PBR Target Status : Down
  Extended Action   : None
Secondary Action    : Forward (Next Hop VRF)
  Next Hop          : 172.16.13.2
  Router            : 2
  PBR Target Status : Up
  Extended Action   : None
PBR Down Action     : Drop (entry-default)
Downloaded Action   : Secondary
Dest. Stickiness    : None                        Hold Remain    : 0
Ing. Matches        : 201 pkts (21306 bytes)
Egr. Matches        : 0 pkts

===============================================================================
```

When both PBR targets are down, the default action is drop, because the IPv4 filter does not have the **pbr-down-action-override** parameter configured. Stickiness is not enabled in this filter. The configuration of the IPv4/v6 filters is similar for PBR and PBF.

In the preceding PBR example, the primary and secondary next-hop router is the same VRF "VPRN-2", but it can be any mix of VRFs, such as primary next-hop router 100 and secondary next-hop router 200.

PBR can also steer traffic to the base routing instance; for example, with the following IP filter:

```
configure {
    filter {
        ip-filter "IP-40" {
            filter-id 40
            entry 10 {
                action {
                    forward {
                        next-hop {
                            nh-ip-vrf {
                                router-instance "Base"
                                address 192.0.2.2
                            }
                        }
                    }
                    secondary {
                        forward {
                            next-hop {
                                nh-ip-vrf {
                                    router-instance "Base"
                                    address 192.0.2.3
                                }
                            }
                        }
                    }
                }
            }
```

Conclusion

Operators can define two targets for L2 and L3 traffic steering (PBF and PBR): primary and secondary. The primary target is used when both targets are up; the secondary target is used when the primary is down. However, when stickiness is enabled, it is possible that the secondary action is executed, even when the primary action PBR target reverts to up. When both targets are down, the default action is drop, unless the **pbr-down-action-override** parameter is configured. Both 1+1 redundancy and N+1 redundancy are supported.

Rate Limit Filter Action

This chapter provides information about Rate Limit Filter Action.

Topics in this chapter include:

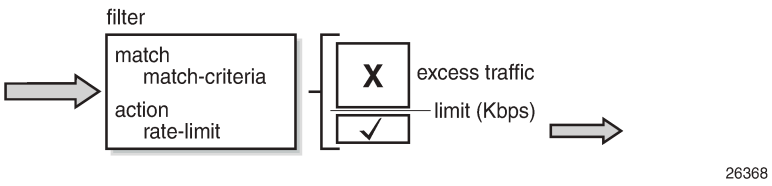
Applicability

This chapter is applicable to SR OS routers and is based on SR OS Release 24.3.R1.

Overview

Filter-based rate limiting can be used by operators for security reasons to protect their network resources or mitigate DDoS attacks; see Figure 27: Filter Based Rate Limiting.

Figure 27: Filter Based Rate Limiting



SR OS supports filter-based rate limiting on ingress (SR OS Release 14.0.R1) and on egress (SR OS Release 14.0.R4) for IPv4, IPv6, and MAC filter policies

The rate-limit value is configurable in kilobits per second and applicable to traffic matching the filter condition. Packets matching the filter condition are dropped when the traffic rate is above the configured policer rate value and forwarded when the traffic rate is below the configured policer rate value.

QoS Interaction

On ingress, if the MAC or IPv4/IPv6 filter action indicates that traffic must be rate limited, this traffic is redirected to a rate-limiting filter policer before delivery to the switching fabric. Traffic not matching the MAC or IP filter will pass through the regular packet processing chain, and can be limited through SAP-ingress policies. Control traffic that is extracted to the CPM is not rate limited. Rate-limiting filter policies can coexist with the cflowd, log, and mirror features.

On egress, control and data traffic matching an egress rate-limiting filter policy bypasses egress QoS policing, but the usual egress QoS queueing still applies.

Rate-Limiting with Single or Multiple FlexPaths

Filter-based rate limiting can be applied to Layer 2 and Layer 3 services, and is supported on following items, including but not limited to:

- SAPs
- Network interface
- Spoke-SDPs
- group interfaces
- ESM subscribers

Filter-based rate limiting can also be used when the underlying infrastructure uses link aggregation.

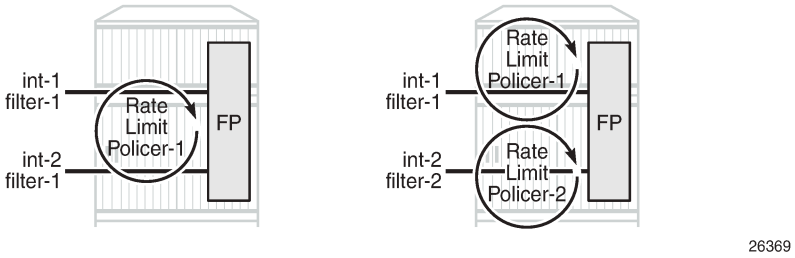
If multiple interfaces use the same rate-limiting filter policy on the same FP, the system will allocate a single rate-limiter resource to the FP; a common aggregate rate limit is applied to those interfaces.

If multiple interfaces use the same rate-limiting filter policy on different FPs, the system will allocate a rate-limiter resource for each FP; an independent rate limit applies to each FP.

The example to the left in Figure 28: Rate Limit Filters and FlexPaths has two interfaces with the same filter applied, and terminated on the same FP. Therefore, there is only one policer, and the aggregate traffic is topped at the rate defined in the filter. The example to the right has two interfaces with different filters, again terminated on the same FP. Because the interfaces have distinct filters, two different rate-limiting policers are created, which could (but not necessarily) define the same rate.

The actual packet length is used for the rate limit, not factoring in the encapsulation.

Figure 28: Rate Limit Filters and FlexPaths



Use caution when applying filter-based rate limiting to SAPs on group interfaces, because group interfaces can host many ESM subscribers, which could defeat per-subscriber and per-ESM host rate limiting.

Syntax

The following syntax defines an IPv4/IPv6 filter or a MAC filter with rate-limiting action:

```
# on PE-1:
[ex:/configure filter]
A:admin@PE-1#         info
    ip-filter | ipv6-filter | mac-filter "<filter-name>" {
```

```
        description "<filter-description>"
        default-action accept | forward | drop
        filter-id <filter-id>
        entry <entry-id> {
            match {
                ** match criteria, e.g.: IP/Port/MAC **
            }
            action {
                accept
                rate-limit {
                    pir <value-Kbps>
                }
            }
        }
    }
```

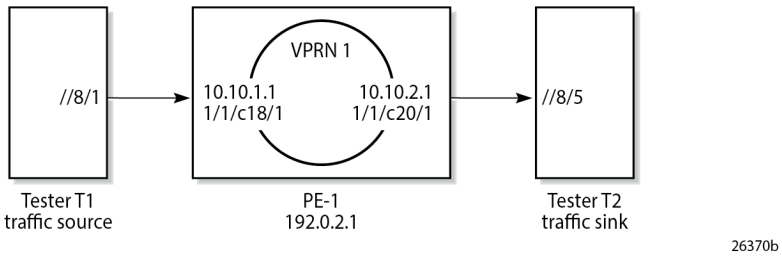
All regular IP and MAC match criteria are supported with the **action rate-limit**.

Configuration

Figure 29: Example Configuration shows the example configuration. Traffic is sourced on Tester T1, port //8/1, passes through VPRN 1, and is received on port //8/5 of Tester T2.

Ingress IPv4 filtering applies at the ingress SAP in VPRN 1. Ingress IPv6 filtering and ingress MAC filtering are similar to ingress IPv4 filtering and are not shown in this chapter.

Figure 29: Example Configuration



The configuration of VPRN 1 on PE-1 is as follows:

```
# on PE-1:
configure {
    service {
        vprn "VPRN 1" {
            admin-state enable
            description "rate limit action for ip filter"
            service-id 1
            customer "1"
            bgp-ipvpn {
                mpls {
                    admin-state enable
                    route-distinguisher "65536:1"
                }
            }
            interface "int-TST-receiver" {
                ipv4 {
                    primary {
                        address 10.10.2.1
```

```
                        prefix-length 24
                    }
                }
                sap 1/1/c20/1 { }
            }
            interface "int-TST-source" {
                ipv4 {
                    primary {
                        address 10.10.1.1
                        prefix-length 24
                    }
                }
                sap 1/1/c18/1 {
                    ingress {
                        filter {
                            ip "ip-filter-1M"
                        }
                    }
                }
            }
        }
```

The filter configuration is as follows:

```
# on PE-1:
configure {
    filter {
        ip-filter "ip-filter-1M" {
            description "IP filter test for rate limit action"
            default-action accept
            filter-id 3
            entry 10 {
                match {
                    src-ip {
                        address 10.10.1.2/32
                    }
                    dst-ip {
                        address 10.10.2.2/32
                    }
                }
                action {
                    accept
                    rate-limit {
                        pir 1024    # in Kbps ; 1024000/8/128 = 1000 packets/s
                    }
                }
            }
        }
```

A stream of UDP packets with a fixed size of 128 bytes is sent out of Tester T1 at a rate of 500 packets/s, accounting for a data rate of 500 x 128 x 8 = 512Kbit/s. At this rate, all packets pass through because the actual rate is lower than the rate-limit 1024Kbit/s, as follows:

```
[/]
A:admin@PE-1# monitor filter ip "ip-filter-1M" entry 10 rate repeat 6 interval 11

===============================================================================
Monitor statistics for IP filter 3 entry 10
===============================================================================
-------------------------------------------------------------------------------
At time t = 0 sec (Base Statistics)
-------------------------------------------------------------------------------
```

```
Ing. Matches        : 2 pkts (256 bytes)
Egr. Matches        : 0 pkts
Ing. Rate-limiter
  Offered           : 0 pkts
  Forwarded         : 0 pkts
  Dropped           : 0 pkts
Egr. Rate-limiter
  Offered           : 0 pkts
  Forwarded         : 0 pkts
  Dropped           : 0 pkts

-------------------------------------------------------------------------------
At time t = 11 sec (Mode: Rate)
-------------------------------------------------------------------------------
Ing. Matches        : 500 pkts (63988 bytes)
Egr. Matches        : 0 pkts
Ing. Rate-limiter
  Offered           : 500 pkts (64012 bytes)
  Forwarded         : 500 pkts (64012 bytes)
  Dropped           : 0 pkts
Egr. Rate-limiter
  Offered           : 0 pkts
  Forwarded         : 0 pkts
  Dropped           : 0 pkts

---snip---
```

Note: In mode **rate**, pkts means pkts/s

Increasing the actual rate to 1500 packets/s without changing the frame size corresponds to a data rate of 1500 x 128 x 8 = 1536Kbit/s, so part of the traffic is dropped as 1536Kbit/s > 1024Kbit/s, as follows:

```
[/]
A:admin@PE-1# monitor filter ip "ip-filter-1M" entry 10 rate repeat 6 interval 11

===============================================================================
Monitor statistics for IP filter 3 entry 10
===============================================================================
-------------------------------------------------------------------------------
At time t = 0 sec (Base Statistics)
-------------------------------------------------------------------------------
Ing. Matches        : 7 pkts (896 bytes)
Egr. Matches        : 0 pkts
Ing. Rate-limiter
  Offered           : 0 pkts
  Forwarded         : 0 pkts
  Dropped           : 0 pkts
Egr. Rate-limiter
  Offered           : 0 pkts
  Forwarded         : 0 pkts
  Dropped           : 0 pkts

-------------------------------------------------------------------------------
At time t = 11 sec (Mode: Rate)
-------------------------------------------------------------------------------
Ing. Matches        : 1500 pkts (191965 bytes)
Egr. Matches        : 0 pkts
Ing. Rate-limiter
  Offered           : 1500 pkts (192047 bytes)
  Forwarded         : 996 pkts (127523 bytes)
  Dropped           : 504 pkts (64524 bytes)
```

```
Egr. Rate-limiter
  Offered           : 0 pkts
  Forwarded         : 0 pkts
  Dropped           : 0 pkts

---snip---
```

Note: In mode **rate**, pkts means pkts/s

When sending traffic at a rate of 500 packets/s with a 128 bytes packet-size and monitoring at entry-point SAP 1/1/c18/1 over 11 s intervals, 500 packets/s should be received on interface int-TST-source, accounting for 500 x 128 = 64000 octets/s. The output shows:

```
[/]
A:admin@PE-1# monitor service id "VPRN 1" sap "1/1/c18/1" rate repeat 6 interval 11

===============================================================================
Monitor statistics for Service 1 SAP 1/1/c18/1
===============================================================================
-------------------------------------------------------------------------------
At time t = 0 sec (Base Statistics)
-------------------------------------------------------------------------------
---snip---

-------------------------------------------------------------------------------
At time t = 11 sec (Mode: Rate)
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
At time t = 22 sec (Mode: Rate)
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
At time t = 33 sec (Mode: Rate)
-------------------------------------------------------------------------------

-------------------------------------------------------------------------------
Sap Aggregate Stats
-------------------------------------------------------------------------------
                        Packets                 Octets
Ingress
Aggregate Offered     : 0                       0
Aggregate Forwarded   : 0                       0
Aggregate Dropped     : 0                       0

Egress
Aggregate Forwarded   : 0                       0
Aggregate Dropped     : 0                       0
-------------------------------------------------------------------------------
Sap Statistics
-------------------------------------------------------------------------------
Last Cleared Time     : 04/12/2024 18:53:07

                        Packets                 Octets                  % Port
                                                                        Util.
CPM Ingress           : 0                       0                       0.00

Forwarding Engine Stats
Dropped               : 0                       0                       0.00
Received Valid        : 455                     58193                   ~0.00
Off. HiPrio           : 0                       0                       0.00
```

```
Off. LowPrio          : 0                       0                       0.00
Off. Uncolor          : 0                       0                       0.00
Off. Managed          : 0                       0                       0.00

---snip---
```

Note: In mode **rate**, Packets means Packets/s and Octets means Octets/s

Note: There may be an error in the computation: 455 = <rate> / 11 x 10, should be 500 = <rate>

When sending traffic at a rate of 1500 packets/s with a 128 bytes packet-size and monitoring at exit-point SAP 1/1/c20/1 over 11 s intervals, only 1000 packets/s are sent out of interface int-TST-receiver, accounting for 128000 octets/s. The output shows:

```
[/]
A:admin@PE-1# monitor service id "VPRN 1" sap "1/1/c20/1" rate repeat 6 interval 11

===============================================================================
Monitor statistics for Service 1 SAP 1/1/c20/1
===============================================================================
-------------------------------------------------------------------------------
At time t = 0 sec (Base Statistics)
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
At time t = 11 sec (Mode: Rate)
-------------------------------------------------------------------------------

-------------------------------------------------------------------------------
Sap Aggregate Stats
-------------------------------------------------------------------------------
                        Packets                 Octets
Ingress
Aggregate Offered     : 0                       0
Aggregate Forwarded   : 0                       0
Aggregate Dropped     : 0                       0

Egress
Aggregate Forwarded   : 996                     127488
Aggregate Dropped     : 0                       0
-------------------------------------------------------------------------------
Sap Statistics
-------------------------------------------------------------------------------
Last Cleared Time     : 04/12/2024 18:58:38

                        Packets                 Octets                  % Port
                                                                        Util.
CPM Ingress           : 0                       0                       0.00

Forwarding Engine Stats
Dropped               : 0                       0                       0.00
Received Valid        : 0                       0                       0.00
Off. HiPrio           : 0                       0                       0.00
Off. LowPrio          : 0                       0                       0.00
Off. Uncolor          : 0                       0                       0.00
Off. Managed          : 0                       0                       0.00

Queueing Stats(Ingress QoS Policy 1)
Dro. HiPrio           : 0                       0                       0.00
Dro. LowPrio          : 0                       0                       0.00
```

```
For. InProf           : 0                     0                     0.00
For. OutProf          : 0                     0                     0.00

Queueing Stats(Egress QoS Policy 1)
Dro. In/InplusProf    : 0                     0                     0.00
Dro. Out/ExcProf      : 0                     0                     0.00
For. In/InplusProf    : 996                   127488                ~0.00
For. Out/ExcProf      : 0                     0                     0.00
-------------------------------------------------------------------------------
Sap per Queue Stats
-------------------------------------------------------------------------------
                        Packets               Octets                % Port
                                                                    Util.

Ingress Queue 1 (Unicast) (Priority)
Off. HiPrio           : 0                     0                     0.00
Off. LowPrio          : 0                     0                     0.00
Dro. HiPrio           : 0                     0                     0.00
Dro. LowPrio          : 0                     0                     0.00
For. InProf           : 0                     0                     0.00
For. OutProf          : 0                     0                     0.00

Ingress Queue 11 (Multipoint) (Priority)
Off. Combined         : 0                     0                     0.00
Off. Managed          : 0                     0                     0.00
Dro. HiPrio           : 0                     0                     0.00
Dro. LowPrio          : 0                     0                     0.00
For. InProf           : 0                     0                     0.00
For. OutProf          : 0                     0                     0.00

Egress Queue 1
For. In/InplusProf    : 996                   127488                ~0.00
For. Out/ExcProf      : 0                     0                     0.00
Dro. In/InplusProf    : 0                     0                     0.00
Dro. Out/ExcProf      : 0                     0                     0.00

---snip---
```

Note: In mode **rate**, Packets means Packets/s and Octets means Octets/s

Other commands to verify the rate limiting operation within a counting period are:

- **clear service statistics id "<service-name>" counters**, or **clear service statistics sap "<sap-id>" all** or **clear service statistics sap "<sap-id>" counters**, followed by: **show service id "<service-name>" sap "<sap-id>" base**, **show service id "<service-name>" sap "<sap-id>" stats** and **show service id "<service-name>" sap "<sap-id>" sap-stats** after the end of the counting period
- **clear filter ip|ipv6|mac "<filter-name>"**, followed by: **show filter ip|ipv6|mac "<filter-name>" counters [detail]** after the end of the counting period

They show absolute values, no rates.

Conclusion

Rate-limiting filter actions can be used by network operators for security purposes to protect network resources and can also be used to mitigate DDoS attacks.

Weighted ECMP for 6PE over RSVP-TE LSPs

This chapter provides information about Weighted Equal Cost Multipath (ECMP) for IPv6 Provider Edge (6PE) routers over Resource Reservation Protocol with Traffic Engineering (RSVP-TE) Label Switched Paths (LSPs).

Topics in this chapter include:

Applicability

The information and configuration in this chapter are based on SR OS Release 23.3.R2. Weighted ECMP for 6PE routers over RSVP-TE LSPs is supported in SR OS Release 15.0.R6, and later.

Chapter *Weighted ECMP for VPRN over RSVP-TE and SR-TE LSPs* in the *7450 ESS, 7750 SR, and 7950 XRS Layer 3 Services Advanced Configuration Guide for MD CLI* is recommended reading.

Overview

Equal Load Balancing

In this chapter, ECMP refers to spraying traffic flows over multiple RSVP-TE LSPs within an ECMP set. ECMP spraying consists of hashing the relevant fields in the packet header and selecting the tunnel next-hop based on the modulo operation of the output of the hash and the number of RSVP-TE LSPs present in the ECMP set. The maximum number of RSVP-TE LSPs in the ECMP set is defined by the **ecmp** command.

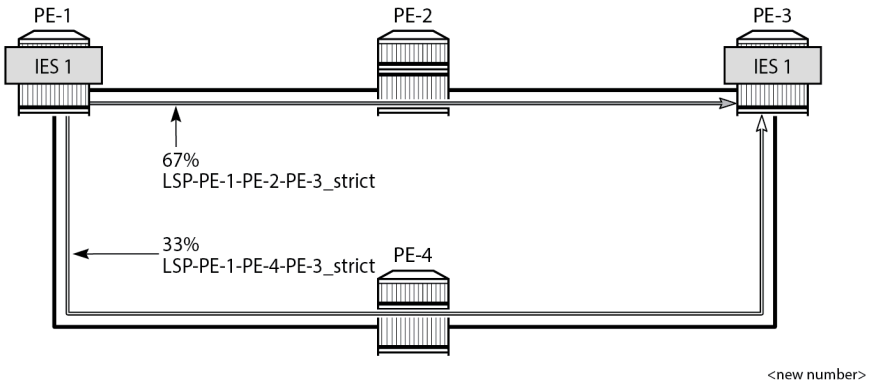
Only RSVP-TE LSPs with the same lowest LSP metric can be part of the ECMP set. If the number of such RSVP-TE LSPs exceeds the maximum number of RSVP-TE LSPs allowed in the ECMP set as defined by the **ecmp** command, the RSVP-TE LSPs with the lowest tunnel IDs are selected first. By default, all RSVP-TE LSPs in the ECMP set have the same weight, and traffic flows are spread evenly over all RSVP-TE LSPs in the ECMP set, regardless of the bandwidth of the active path in the RSVP-TE LSPs. By default, ECMP is enabled and set to 1.

Unequal Load Balancing

Weighted ECMP sprays traffic flows over RSVP-TE LSPs proportionally to the **load-balancing-weight** <*weight*> value configured on each RSVP-TE LSP in the ECMP set. Figure 30: Weighted ECMP in AS

64496 shows that PE-1 forwards two thirds of the traffic flows on LSP-PE-1-PE-2-PE-3_strict with weight 2 and one third on LSP-PE-1-PE-4-PE-3_strict with weight 1.

Figure 30: Weighted ECMP in AS 64496



The LSP load balancing weight can be configured in an LSP template or on an RSVP-TE LSP. By default, the load balancing weight equals zero, in which case regular ECMP applies.

Weighted load balancing can be performed only when all the next-hops are associated with the same neighbor and all the RSVP-TE LSPs are configured with a non-zero load balancing weight. If one or more RSVP-TE LSPs in the ECMP set toward a specific next-hop do not have a load balancing weight configured, regular ECMP spraying is used.

The following command is used to configure the weight in an LSP template:

```
configure {
    router "Base" {
        mpls {
            lsp-template "LSPtemplate1" {
                load-balancing-weight ?

 load-balancing-weight <number>
 <number> - <1..4294967295>

    Load balancing weight for an MPLS LSP template

    Warning: Modifying this element toggles
    'configure router "Base" mpls lsp-template "LSPtemplate1" admin-state' automatically
    for the new value to take effect.
```

The following command is used to configure the weight on an LSP (for example on "LSP-PE-1-PE-2-PE-3_strict"):

```
configure {
    router "Base" {
        mpls {
            lsp "LSP-PE-1-PE-2-PE-3_strict" {
                load-balancing-weight ?

 load-balancing-weight <number>
 <number> - <1..4294967295>

    Load balancing weight for an MPLS LSP
```

The LSP load balancing weight on LSP-PE-1-PE-2-PE-3_strict is configured with a value of 2, as follows:

```
configure {
    router "Base" {
        mpls {
            path "path-PE-1-PE-2-PE-3_strict" {
                admin-state enable
                hop 10 {
                    ip-address 192.168.12.2
                    type strict
                }
                hop 20 {
                    ip-address 192.168.23.2
                    type strict
                }
            }
            lsp "LSP-PE-1-PE-2-PE-3_strict" {
                admin-state enable
                type p2p-rsvp
                to 192.0.2.3
                path-computation-method local-cspf
                metric 100
                load-balancing-weight 2
                primary "path-PE-1-PE-2-PE-3_strict" {
                }
            }
```

Weighted ECMP for 6PE over RSVP-TE LSPs is enabled in the **bgp next-hop-resolution** context as follows:

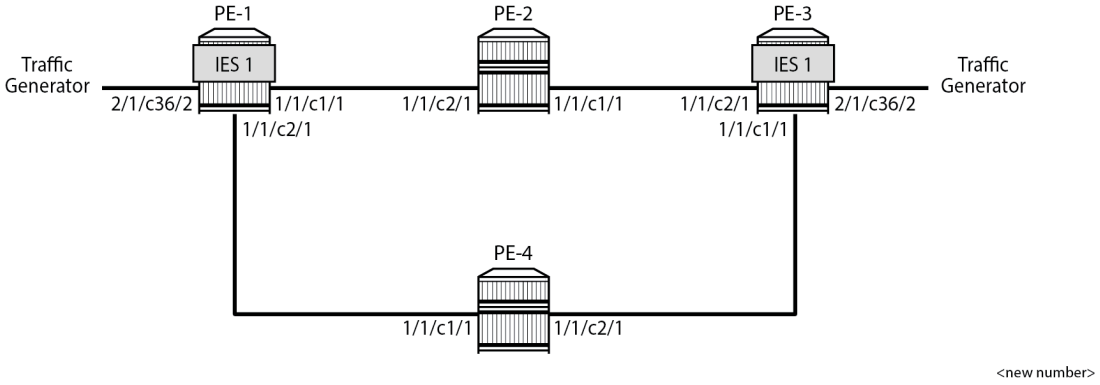
```
configure {
    router "Base" {
        bgp {
            next-hop-resolution {
                weighted-ecmp true
```

The **weighted-ecmp** option controls load balancing to the same next-hop only.

Configuration

Figure 31: Example Topology for 6PE over RSVP-TE LSPs shows the example topology with four PEs. IES 1 is configured on PE-1 and PE-3. A traffic generator is connected to IES 1 SAP 2/1/c36/2 on PE-1 and IES 1 SAP 2/1/c36/2 on PE-3. The traffic generator generates multiple IPv6 traffic flows with random IP addresses and TCP/UDP port numbers. As a result, these flows are sprayed over different MPLS LSPs between PE-1 and PE-3.

Figure 31: Example Topology for 6PE over RSVP-TE LSPs



Initial Configuration

The initial configuration on the PEs includes the following:

- Cards, MDAs, ports
- Router interfaces
- IS-IS as IGP (alternatively, OSPF can be used) with traffic engineering enabled
- MPLS and RSVP enabled on all router interfaces
- MPLS paths with strict hops from PE-1 to PE-3 and the other way around: one via PE-2 and the other via PE-4. The LSP via PE-2 gets a load balancing weight of 2, whereas the LSP via PE-4 gets a load balancing weight of 1. Both LSPs have the same metric.

The initial configuration on PE-1 is as follows.

```
configure {
    router "Base" {
        interface "int-PE-1-PE-2" {
            port 1/1/c1/1
            ipv4 {
                primary {
                    address 192.168.12.1
                    prefix-length 30
                }
            }
        }
        interface "int-PE-1-PE-4" {
            port 1/1/c2/1
            ipv4 {
                primary {
                    address 192.168.14.1
                    prefix-length 30
                }
            }
        }
        interface "system" {
            ipv4 {
                primary {
                    address 192.0.2.1
```

```
                prefix-length 32
            }
        }
    }
    isis 0 {
        admin-state enable
        area-address [49.0001]
        traffic-engineering true
        interface "system" {
        }
        interface "int-PE-1-PE-2" {
            interface-type point-to-point
        }
        interface "int-PE-1-PE-4" {
            interface-type point-to-point
        }
    }
    mpls {
        admin-state enable
        interface "int-PE-1-PE-2" {
        }
        interface "int-PE-1-PE-4" {
        }
        path "path-PE-1-PE-2-PE-3_strict" {
            admin-state enable
            hop 10 {
                ip-address 192.168.12.2
                type strict
            }
            hop 20 {
                ip-address 192.168.23.2
                type strict
            }
        }
        path "path-PE-1-PE-4-PE-3_strict" {
            admin-state enable
            hop 10 {
                ip-address 192.168.14.2
                type strict
            }
            hop 20 {
                ip-address 192.168.34.1
                type strict
            }
        }
        lsp "LSP-PE-1-PE-2-PE-3_strict" {
            admin-state enable
            type p2p-rsvp
            to 192.0.2.3
            path-computation-method local-cspf
            metric 100
            load-balancing-weight 2
            primary "path-PE-1-PE-2-PE-3_strict" {
            }
        }
        lsp "LSP-PE-1-PE-4-PE-3_strict" {
            admin-state enable
            type p2p-rsvp
            to 192.0.2.3
            path-computation-method local-cspf
            metric 100
            load-balancing-weight 1
            primary "path-PE-1-PE-4-PE-3_strict" {
            }
```

```
            }
        }
        rsvp {
            admin-state enable
            interface "int-PE-1-PE-2" {
            }
            interface "int-PE-1-PE-4" {
            }
        }
```

The configuration on PE-3 is similar.

With the preceding configuration, MPLS and RSVP are enabled on all interfaces, including the system interface, which is added automatically.

Weighted ECMP for 6PE over RSVP-TE LSPs

BGP is configured for the label-IPv6 address family and the next-hop resolution is set to RSVP; see the 6PE Next-Hop Resolution chapter.

In this example, the traffic generator sends IPv6 traffic to the SAP in IES 1. The IPv6 packets are tunneled through the IPv4 network between PE-1 and PE-3. The service configuration on PE-1 is as follows:

```
configure {
    service {
        ies "IES-1" {
            admin-state enable
            service-id 1
            customer "1"
            description "6PE-1"
            interface "int-PE-1-STC" {
                sap 2/1/c36/2 {
                }
                ipv6 {
                    address 2001:db8::11:1 {
                        prefix-length 120
                    }
```

The configuration on PE-3 is similar.

On PE-1, the following BGP configuration defines next-hop resolution with weighted ECMP and the resolution filter only allows RSVP-TE LSPs. BGP is configured for the label-IPv6 address family and BGP multipath is configured in the **bgp** context.

```
configure {
    router "Base" {
        autonomous-system 64496
        bgp {
            ibgp-multipath true
            split-horizon true
            next-hop-resolution {
                weighted-ecmp true
                labeled-routes {
                    transport-tunnel {
                        family label-ipv6 {
                            resolution-filter {
                                ldp false
                                rsvp true
                            }
                        }
```

```
                    }
                }
            }
            group "iBGP" {
                path-mtu-discovery true
                peer-as 64496
                export {
                    policy ["export-6PE-1"]
                }
            }
            neighbor 192.0.2.3 {
                group "iBGP"
                family {
                    label-ipv6 true
                }
            }
        }
```

The configuration on PE-3 is similar.

On PE-1 and PE-3, the following export policy is configured:

```
configure {
    policy-options {
        policy-statement "export-6PE-1" {
            entry 10 {
                from {
                    protocol {
                        name [direct]
                    }
                }
                action {
                    action-type accept
                }
            }
            default-action {
                action-type reject
            }
        }
```

The following command enables ECMP in the base router.

```
configure {
    router "Base" {
        ecmp 2
```

On PE-1, the route table in the base router shows that the remote prefix 2001:db8::33:0/120 has flag [2], meaning that the next-hop 192.0.2.3 occurs twice for this prefix, as follows:

```
[/]
A:admin@PE-1# show router route-table 2001:db8::33:0/120

===============================================================================
IPv6 Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
2001:db8::33:0/120 [2]                        Remote  BGP_LABEL 00h01m10s  170
       192.0.2.3 (tunneled:RSVP:2)                                  100
2001:db8::33:0/120 [2]                        Remote  BGP_LABEL 00h01m10s  170
       192.0.2.3 (tunneled:RSVP:4)                                  100
```

```
-------------------------------------------------------------------------------
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 shows a similar route with flag [2] for prefix 2001:db8::11:0/120.

On PE-1, the following detailed route table info (using keyword **extensive**) for prefix 2001:db8::33:0/120 shows that RSVP-TE tunnel 2 and RSVP-TE tunnel 4 are used to reach the next-hop 192.0.2.3. Both RSVP-TE tunnels have metric 100, but the weight of RSVP-TE tunnel 2 is twice as much as the weight of RSVP tunnel 4, so the load on RSVP-TE LSP 2 is twice as high as the load on RSVP LSP 4.

```
[/]
A:admin@PE-1# show router route-table 2001:db8::33:0/120 extensive

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix             : 2001:db8::33:0/120
  Protocol              : BGP_LABEL
  Age                   : 00h01m10s
  Preference            : 170
  Indirect Next-Hop     : 192.0.2.3
    Label               : 2
    QoS                 : Priority=n/c, FC=n/c
    Source-Class        : 0
    Dest-Class          : 0
    ECMP-Weight         : N/A
    Resolving Next-Hop  : 192.0.2.3 (RSVP tunnel:2)
      Metric            : 100
      ECMP-Weight       : 2
    Resolving Next-Hop  : 192.0.2.3 (RSVP tunnel:4)
      Metric            : 100
      ECMP-Weight       : 1
-------------------------------------------------------------------------------
No. of Destinations: 1
===============================================================================
```

The following tunnel table on PE-1 shows that RSVP-TE tunnel 2 has PE-2 as next-hop (192.168.12.2) and RSVP-TE tunnel 4 has next-hop PE-4 (192.168.14.2):

```
[/]
A:admin@PE-1# show router tunnel-table

===============================================================================
IPv4 Tunnel Table (Router: Base)
===============================================================================
Destination           Owner     Encap TunnelId  Pref   Nexthop        Metric
   Color
-------------------------------------------------------------------------------
192.0.2.3/32          rsvp      MPLS  2         7      192.168.12.2   100
192.0.2.3/32          rsvp      MPLS  4         7      192.168.14.2   100
---snip---
-------------------------------------------------------------------------------
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
```

```
===============================================================================
```

Traffic Verification

The traffic generator sends IPv6 traffic flows to SAP 2/1/c36/2 of IES 1 on PE-1. The packets are tunneled over the available RSVP-TE LSPs present in the ECMP set. The traffic is load balanced unevenly: two thirds of the traffic flows is tunneled via PE-2 (port 1/1/c1/1) while one third of the traffic flows is tunneled via PE-4 (port 1/1/c2/1). The load on the ports is as follows:

```
[/]
A:admin@PE-1# monitor port 1/1/c1/1 rate interval 3 repeat 3

===============================================================================
Monitor statistics for Port 1/1/c1/1
===============================================================================
                                                    Input                 Output
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
At time t = 6 sec (Mode: Rate)
-------------------------------------------------------------------------------
Octets                                               101                444150
Packets                                                1                   431
Errors                                                 0                     0
Bits                                                 808               3553200
Utilization (% of port capacity)                   ~0.00                  0.03
---snip---
===============================================================================

[/]
A:admin@PE-1# monitor port 1/1/c2/1 rate interval 3 repeat 3

===============================================================================
Monitor statistics for Port 1/1/c2/1
===============================================================================
                                                    Input                 Output
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
At time t = 6 sec (Mode: Rate)
-------------------------------------------------------------------------------
Octets                                               226                186190
Packets                                                2                   182
Errors                                                 0                     0
Bits                                                1808               1489520
Utilization (% of port capacity)                   ~0.00                  0.01
---snip---
===============================================================================

[/]
A:admin@PE-1# monitor port 2/1/c36/2 rate interval 3 repeat 3

===============================================================================
Monitor statistics for Port 2/1/c36/2
===============================================================================
                                                    Input                 Output
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
At time t = 6 sec (Mode: Rate)
```

```
-------------------------------------------------------------------------------
Octets                                          602112                     0
Packets                                            588                     0
Errors                                               0                     0
Bits                                           4816896                     0
Utilization (% of port capacity)                  0.04                  0.00
---snip---
===============================================================================
```

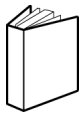
This can also be verified as follows:

```
[/]
A:admin@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                                47                           4863
                                     14730                       15157578
===============================================================================

[/]
A:admin@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                                44                           4423
                                      6266                        6424681
===============================================================================

[/]
A:admin@PE-1# show port 2/1/c36/2 statistics

===============================================================================
Port Statistics on Slot 2
===============================================================================
Port                        Ingress Packets                 Ingress Octets
Id                           Egress Packets                  Egress Octets
-------------------------------------------------------------------------------
2/1/c36/2                           20904                       21405696
                                        0                              0
===============================================================================
```

Conclusion

Operators can control how 6PE traffic is load balanced unequally over multiple RSVP-TE LSPs by defining a load balancing weight value on each LSP.

Customer document and product support



Customer documentation

Customer documentation welcome page

Technical support

Product support portal

Documentation feedback

Customer documentation feedback