7450 Ethernet Service Switch
7750 Service Router
7950 Extensible Routing System
Virtualized Service Router

Releases up to 24.7.R2

Router Configuration Advanced Configuration Guide for Classic CLI

3HE 20802 AAAB TQZZA
Edition: 01
October 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

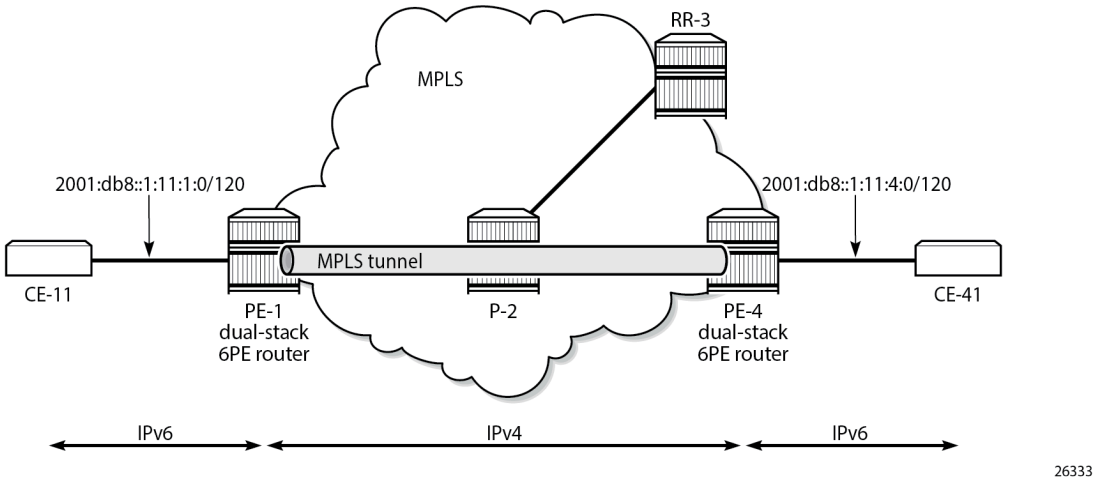
This chapter was initially written based on SR OS Release 14.0.R7, but the CLI in the current edition corresponds to SR OS Release 23.7.R1.

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:

```
#A:PE-1>config>router>bgp>next-hop-res>lbl-routes>transport-tunn>family# resolution ?
  - resolution {any|filter|disabled}
```

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, a subset of protocols is required, and LDP is automatically added to the protocol list in the resolution filter. The following example shows that when one tries to create a resolution filter that includes the BGP protocol only, the resolution filter includes LDP and BGP. The first info command shows that initially no resolution filter had been defined.

```
*A:PE-1>config>router>bgp>next-hop-res>lbl-routes>transport-tunn>family>res-filter# info
----------------------------------------------
----------------------------------------------
*A:PE-1>config>router>bgp>next-hop-res>lbl-routes>transport-tunn>family>res-filter# info detail
----------------------------------------------
                        ldp
                        no rsvp
                        no sr-isis
                        no sr-ospf
                        no sr-ospf3
                        no bgp
                        no sr-te
                        no udp
                        no sr-policy
                        no rib-api
                        no mpls-fwd-policy
----------------------------------------------
*A:PE-1>config>router>bgp>next-hop-res>lbl-routes>transport-tunn>family>res-filter$ bgp
*A:PE-1>config>router>bgp>next-hop-res>lbl-routes>transport-tunn>family>res-filter$ info
----------------------------------------------
                        ldp
                        bgp
----------------------------------------------
```

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

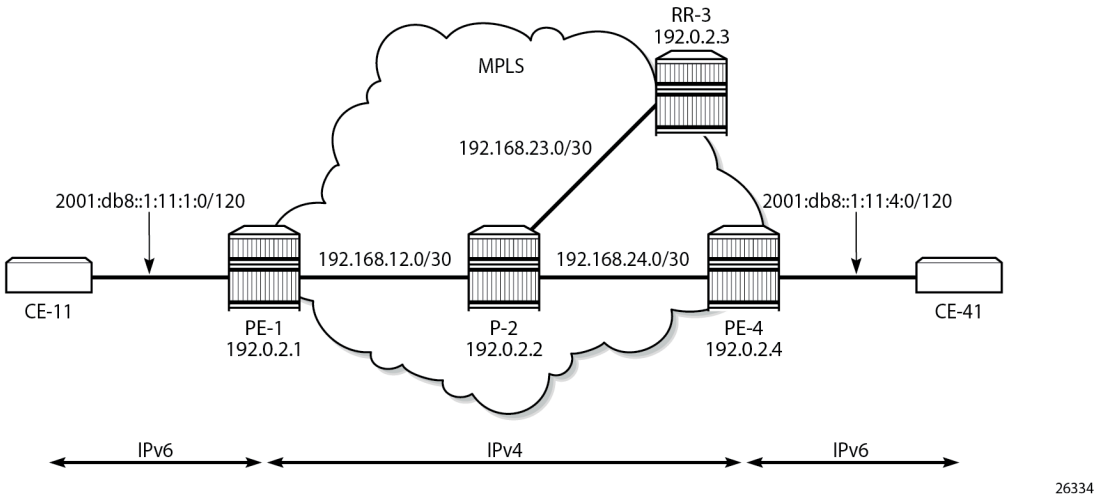
It is possible to explicitly exclude LDP from the list, as follows:

```
*A:PE-1>config>router>bgp>next-hop-res>lbl-routes>transport-tunn>family>res-filter# no ldp
*A:PE-1>config>router>bgp>next-hop-res>lbl-routes>transport-tunn>family>res-filter# info
----------------------------------------------
                        no ldp
                        bgp
----------------------------------------------
```

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
            next-hop-resolution
                labeled-routes
                    transport-tunnel
                        family label-ipv6
                            resolution-filter
```

```
                                ldp     # default
                                rsvp
                                sr-isis
                            exit
                            resolution filter
                        exit
                    exit
                exit
                group "IBGP"
                    export "export-6pe"
                    peer-as 64496
                    neighbor 192.0.2.3
                        family label-ipv6
                    exit
                exit
            exit
```

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
    router Base
        policy-options
            begin
            policy-statement "export-6pe"
                entry 10
                    from
                        protocol direct
                    exit
                    action accept
                    exit
                exit
                default-action drop
                exit
            exit
            commit
```

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
            group "IBGP"
                cluster 192.0.2.3
                peer-as 64496
                neighbor 192.0.2.1
                    family label-ipv6
                exit
                neighbor 192.0.2.4
                    family label-ipv6
                exit
            exit
```

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 1 name "IES-1" customer 1 create
            description "6PE"
            interface "int-PE-1-CE-11" create
                ipv6
                    address 2001:db8::1:11:1:1/120
                exit
                sap 1/1/c3/1:1 create
                exit
            exit
            no shutdown
        exit
```

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 tunnel, which is active in the routing table, is established between the PEs, as follows:

```
*A:PE-1# show router route-table ipv6

===============================================================================
IPv6 Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
2001:db8::1:11:1:0/120                        Local   Local     00h02m38s  0
       int-PE-1-CE-11                                             0
2001:db8::1:11:4:0/120                        Remote  BGP_LABEL 00h01m59s  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:PE-1# ping router 11 2001:db8::1:11:4:41 source 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=3.65ms.
64 bytes from 2001:db8::1:11:4:41 icmp_seq=2 hlim=62 time=8.41ms.
64 bytes from 2001:db8::1:11:4:41 icmp_seq=3 hlim=62 time=3.03ms.
64 bytes from 2001:db8::1:11:4:41 icmp_seq=4 hlim=62 time=3.09ms.
64 bytes from 2001:db8::1:11:4:41 icmp_seq=5 hlim=62 time=1.71ms.

---- 2001:db8::1:11:4:41 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 1.71ms, avg = 3.98ms, max = 8.41ms, stddev = 2.31ms
```

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: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 00h02m20s  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: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: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
```

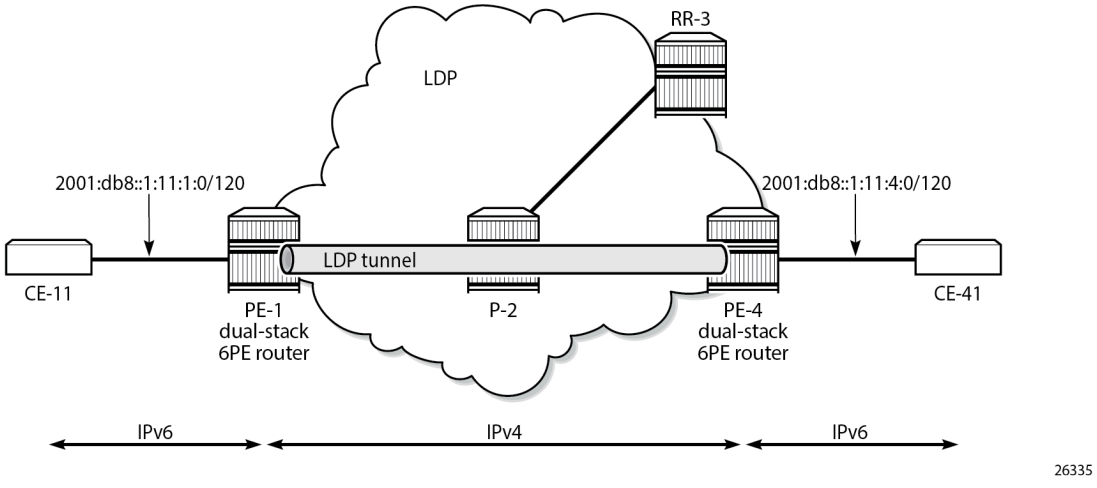
```
  Lbl/SID (backup)
    NextHop   (backup)
-------------------------------------------------------------------------------
192.0.2.4/32                                LDP              -
  524285
    192.168.12.2                                          1/1/c1/1:1000
-------------------------------------------------------------------------------
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: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                   : 00h03m39s
  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"
                no shutdown
            exit
            lsp "LSP-PE-1-PE-4"
                to 192.0.2.4
                primary "empty"
                exit
                no shutdown
            exit
```

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: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               -
  524285
    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
-------------------------------------------------------------------------------
===============================================================================
```

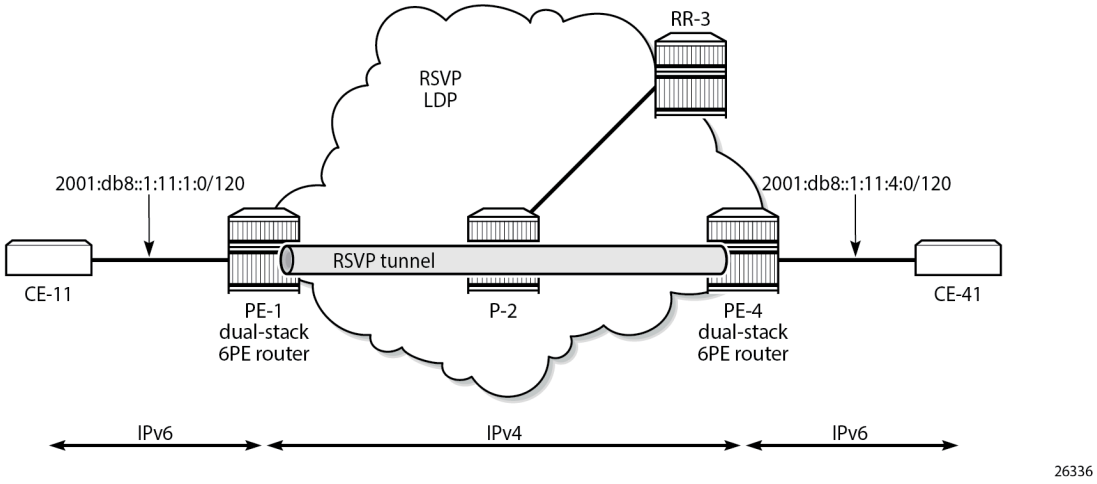
For 6PE next hop resolution, RSVP-TE tunnels are preferred to any other tunnel type in the tunnel table, so the BGP next hop 192.0.2.4 will be resolved to an RSVP-TE tunnel, as follows:

```
*A: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                   : 00h00m47s
  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
        exit
        isis 0
            advertise-router-capability area
            interface "system"
                ipv4-node-sid index 1
            exit
            segment-routing
                prefix-sid-range start-label 20000 max-index 99
```

```
                no shutdown
            exit
        exit
```

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

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

```
*A: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               -
  524285
    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: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                   : 00h02m13s
  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 put in a shutdown state, as follows:

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

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: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               -
  524285
    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: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                   : 00h00m33s
  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:

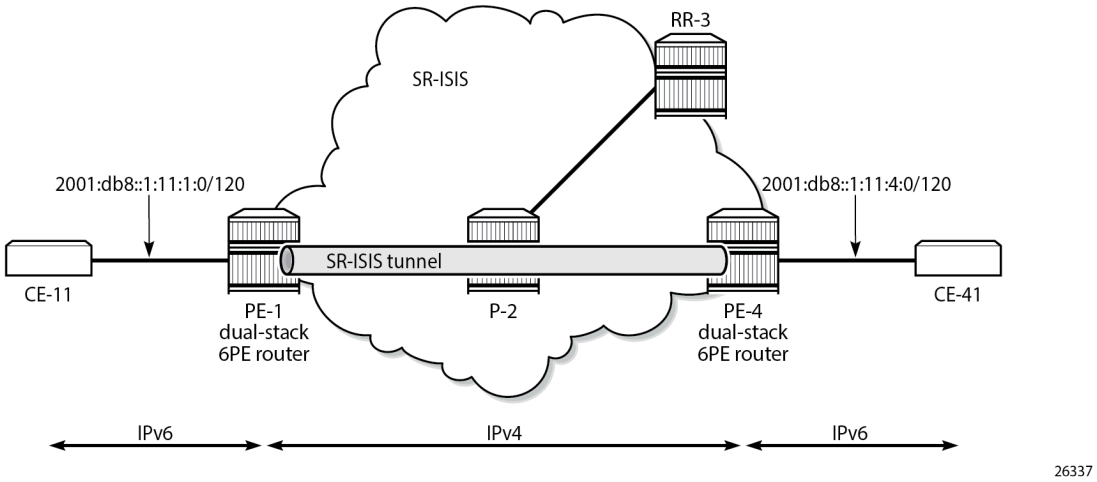
```
# on PE-1:
configure
    router Base
        ldp
            interface-parameters
                interface "int-PE-1-P-2"
                    shutdown


*A: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: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                   : 00h00m48s
  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



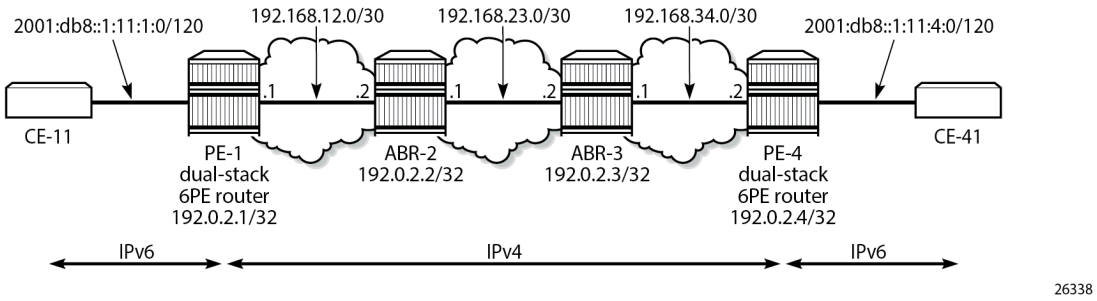
6PE next-hop resolution to a BGP IPv4 tunnel

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 *7450 ESS, 7750 SR, and 7950 XRS MPLS Advanced Configuration Guide for Classic 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

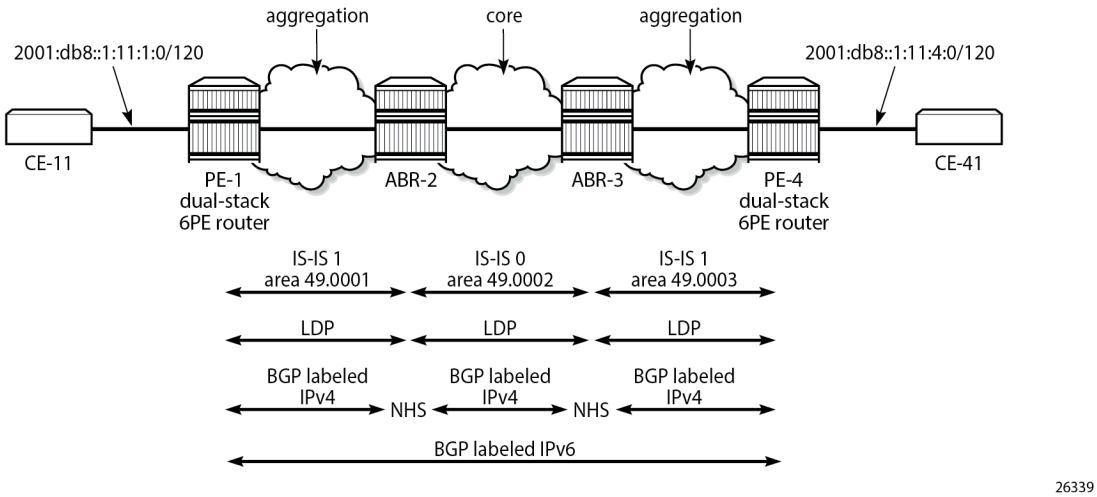


Different IS-IS instances are configured: IS-IS instance 0 is configured in the core, whereas IS-IS instance 1 is configured in the aggregation networks. On the area border routers (ABRs) ABR-2 and ABR-3, two instances of IS-IS are configured: IS-IS instance 0 for the core and IS-IS instance 1 for the 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: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"
            address 192.168.23.1/30
            port 1/1/c3/1:1000
            no shutdown
        exit
        interface "int-ABR-2-PE-1"
            address 192.168.12.2/30
            port 1/1/c2/1:1000
            no shutdown
        exit
        interface "system"
            address 192.0.2.2/32
            no shutdown
        exit
        isis 0
            level-capability level-2
            area-id 49.0002
            interface "system"
                no shutdown
            exit
            interface "int-ABR-2-ABR-3"
                interface-type point-to-point
                no shutdown
            exit
            no shutdown
        exit
        isis 1
            level-capability level-2
            area-id 49.0001
            interface "system"
                no shutdown
            exit
```

```
            interface "int-ABR-2-PE-1"
                interface-type point-to-point
                no shutdown
            exit
            no shutdown
        exit
        ldp
            interface-parameters
                interface "int-ABR-2-PE-1" dual-stack
                    ipv4
                        no shutdown
                    exit
                    no shutdown
                exit
                interface "int-ABR-2-ABR-3" dual-stack
                    ipv4
                        no shutdown
                    exit
                    no shutdown
                exit
            exit
        exit
```

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
            next-hop-resolution
                labeled-routes
                    transport-tunnel
                        family label-ipv6
                            resolution-filter
                                bgp
                            exit
                            resolution filter
                        exit
                    exit
                exit
            exit
            group "IBGPv4"
                export "export-sys"
                peer-as 64496
                neighbor 192.0.2.2
                    family label-ipv4
                exit
            exit
            group "IBGPv6"
                export "export-6pe"
                peer-as 64496
```

```
                neighbor 192.0.2.4
                    family label-ipv6
                exit
            exit
            no shutdown
        exit
```

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
    router Base
        policy-options
            begin
            prefix-list "system"
                prefix 192.0.2.0/24 longer
            exit
            policy-statement "export-sys"
                entry 10
                    from
                        protocol direct
                        prefix-list "system"
                    exit
                    action accept
                    exit
                exit
                default-action drop
                exit
            exit
            commit
```

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

```
# on PE-1, PE-4:
configure
    router Base
        policy-options
            begin
            policy-statement "export-6pe"
                entry 10
                    from
                        protocol direct
                    exit
                    action accept
                    exit
                exit
                default-action drop
                exit
            exit
            commit
```

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
            split-horizon
            group "IBGPv4-agg"
                next-hop-self
                cluster 192.0.2.2
                peer-as 64496
                neighbor 192.0.2.1
                    family label-ipv4
                exit
            exit
            group "IBGPv4-core"
                next-hop-self
                peer-as 64496
                neighbor 192.0.2.3
                    family label-ipv4
                exit
            exit
            no shutdown
        exit
```

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: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 00h03m02s 1/1/1 (Lbl-IPv4)
                           13    0
192.0.2.4
Def. Inst       64496       8    0 00h01m31s 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: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                   : 00h01m18s
  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: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:ABR-2# show router fp-tunnel-table 1 192.0.2.4/32

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

Legend:
```

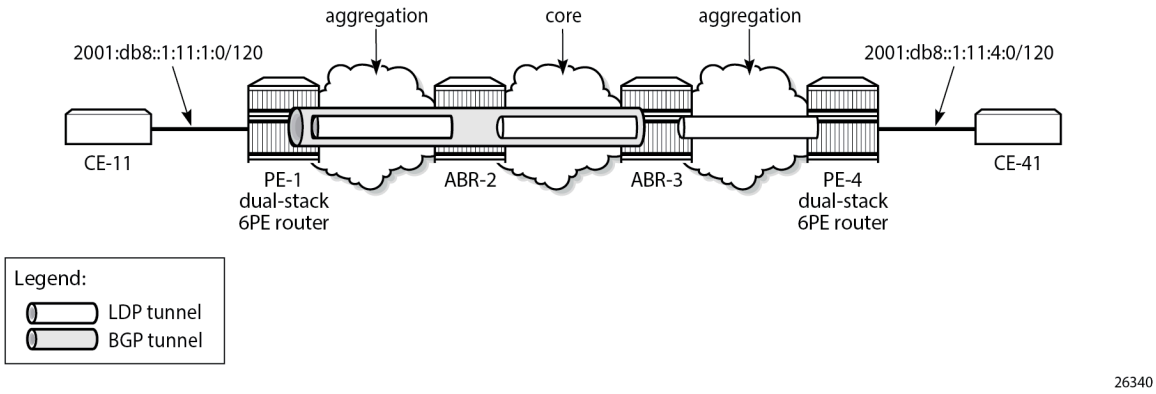
```
label stack is ordered from bottom-most to top-most
B - FRR Backup
===============================================================================
Destination                                Protocol         Tunnel-ID
  Lbl/SID
    NextHop                                                 Intf/Tunnel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

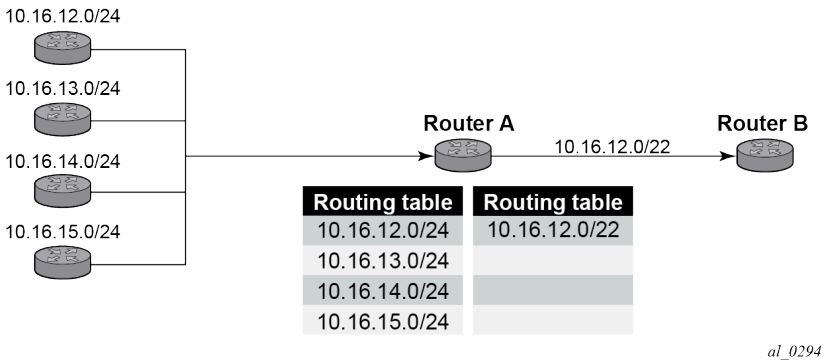
Applicability

This chapter was initially written based on SR OS Release 11.0.R1. The 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.

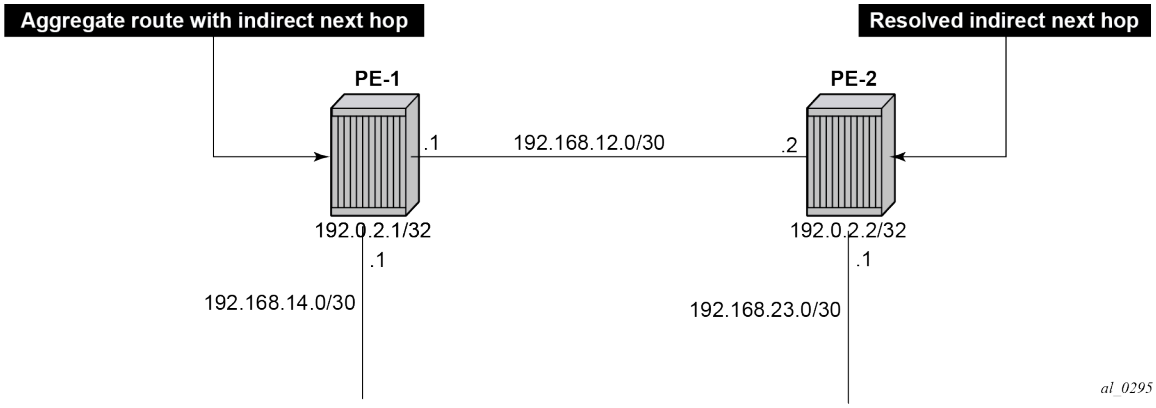
Different network operators have different requirements on how to forward a packet that matches an aggregate route but not any of the more specific routes in the forwarding table that activated the aggregate. In general, there are three different options:

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"
            address 192.168.12.1/30
            port 1/1/c1/1:1000
        exit
        interface "int-PE-1-PE-4"
            address 192.168.14.1/30
            port 1/1/c2/1:1000
        exit
        interface "system"
            address 192.0.2.1/32
        exit
```

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 commands are as follows:

```
configure [ router | service vprn <vprn-id> ] aggregate ?
  - no aggregate <ip-prefix/ip-prefix-length>
  - aggregate <ip-prefix/ip-prefix-length> [summary-only] [as-set] [aggregator
    <as-number:ip-address>] [discard-component-communities] [black-hole [generate-icmp]]
    [community <comm-id1> [<comm-id2> <comm-id3> .. up to 12]] [description
    <description>] [local-preference <local-preference>] [tunnel-group <tunnel-group-id>]
    [policy <policy-name>]
  - aggregate <ip-prefix/ip-prefix-length> [summary-only] [as-set] [aggregator
    <as-number:ip-address>] [discard-component-communities] [indirect <ip-address>]
    [community <comm-id1> [<comm-id2> <comm-id3> .. up to 12]] [description
    <description>] [local-preference <local-preference>] [tunnel-group <tunnel-group-id>]
    [policy <policy-name>]

---snip---
```

Parameters:

- **indirect** — This indicates that the aggregate route has an indirect address. The indirect option is mutually exclusive with the black-hole option. To change the next-hop type of an aggregate route (for example, from black-hole to indirect) the route must be deleted and then re-added with the new next-hop type (however, other configuration attributes can generally be changed dynamically).
- <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
        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: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: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     00h18m35s  0
       system                                                        0
192.168.12.0/30                               Local   Local     00h18m35s  0
       int-PE-1-PE-2                                                 0
192.168.14.0/30                               Local   Local     00h18m35s  0
       int-PE-1-PE-4                                                 0
-------------------------------------------------------------------------------
No. of Routes: 3
```

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-route-entry 10.16.12.0/24
            next-hop 192.168.14.2
                no shutdown
            exit
        exit
        static-route-entry 10.16.13.0/24
            next-hop 192.168.14.2
                no shutdown
            exit
        exit
```

```
        static-route-entry 10.16.14.0/24
            next-hop 192.168.14.2
                no shutdown
            exit
        exit
        static-route-entry 10.16.15.0/24
            next-hop 192.168.14.2
                no shutdown
            exit
        exit
```

As a result, the aggregate route becomes active:

```
*A: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: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      00h00m00s  130
       Black Hole                                                 0
10.16.12.0/24                                 Remote  Static    00h00m00s  5
       192.168.14.2                                               1
10.16.13.0/24                                 Remote  Static    00h00m00s  5
       192.168.14.2                                               1
10.16.14.0/24                                 Remote  Static    00h00m00s  5
       192.168.14.2                                               1
10.16.15.0/24                                 Remote  Static    00h00m00s  5
       192.168.14.2                                               1
192.0.2.1/32                                  Local   Local     00h19m40s  0
       system                                                     0
192.168.12.0/30                               Local   Local     00h19m40s  0
       int-PE-1-PE-2                                              0
192.168.14.0/30                               Local   Local     00h19m40s  0
       int-PE-1-PE-4                                              0
-------------------------------------------------------------------------------
No. of Routes: 8
```

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-route-entry 192.168.11.0/24
            next-hop 192.168.12.2
                no shutdown
            exit
        exit
```

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: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      00h00m14s  130
       192.168.12.2                                               0
10.16.12.0/24                                 Remote  Static    00h04m27s  5
       192.168.14.2                                               1
10.16.13.0/24                                 Remote  Static    00h04m27s  5
       192.168.14.2                                               1
10.16.14.0/24                                 Remote  Static    00h04m27s  5
       192.168.14.2                                               1
10.16.15.0/24                                 Remote  Static    00h04m27s  5
       192.168.14.2                                               1
192.0.2.1/32                                  Local   Local     00h24m08s  0
       system                                                     0
192.168.11.0/24                               Remote  Static    00h00m14s  5
       192.168.12.2                                               1
192.168.12.0/30                               Local   Local     00h24m08s  0
       int-PE-1-PE-2                                              0
192.168.14.0/30                               Local   Local     00h24m08s  0
       int-PE-1-PE-4                                              0
-------------------------------------------------------------------------------
No. of Routes: 9
```

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-route-entry 10.16.12.0/22
            next-hop 192.168.23.2
                no shutdown
            exit
        exit
```

The route table on PE-2 looks as follows:

```
*A: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     00h25m17s  0
       system                                                      0
192.168.12.0/30                               Local   Local     00h25m17s  0
       int-PE-2-PE-1                                               0
192.168.23.0/30                               Local   Local     00h25m17s  0
       int-PE-2-PE-3                                               0
-------------------------------------------------------------------------------
No. of Routes: 4
```

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 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
- 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

Most 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.

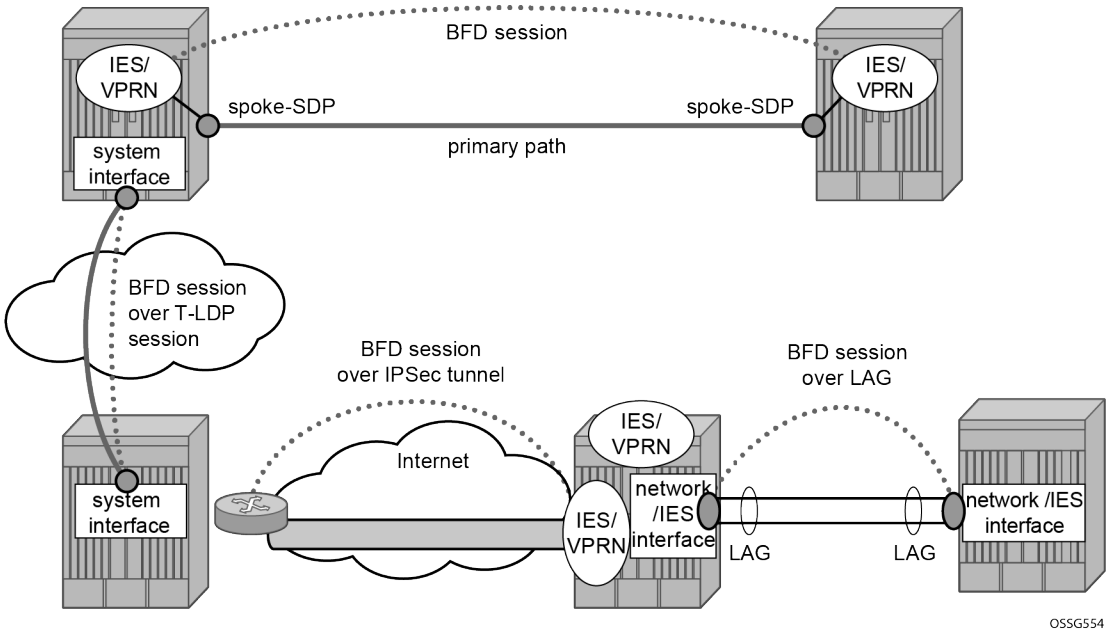
The CPM is able to centrally generate the BFD packets at a subsecond interval as low as 10 ms. The BFD state machine is implemented in software. BFD packet generation can be selectively delegated to CPM hardware as needed. This is applicable when subsecond operations or exceeding the IOM scaling limits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

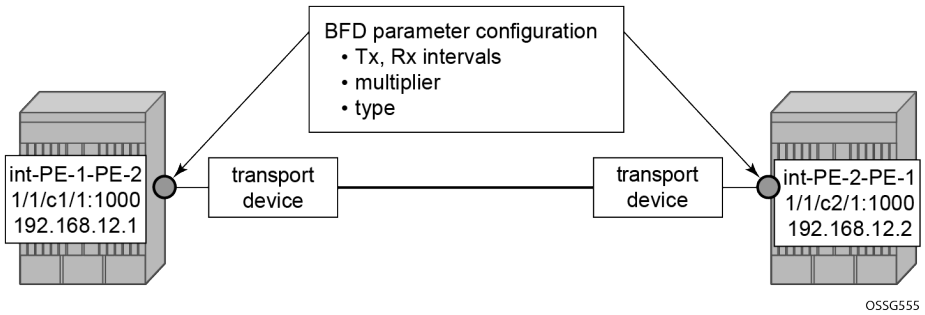
1. configure BFD parameters on the peering interfaces
2. check that the Layer 3 protocol, that is to be bound to BFD, is up and running
3. enable BFD under the Layer 3 protocol interface.

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. In this example, the transmit interval is 100 ms, the receive interval is 100 ms, and the multiplier is 3:

```
# on PE-1:
configure
    router Base
        interface "int-PE-1-PE-2"
            address 192.168.12.1/30
            port 1/1/c1/1:1000
            bfd 100 receive 100 multiplier 3
            no shutdown
        exit
```

```
# on PE-2:
configure
    router Base
        interface "int-PE-2-PE-1"
            address 192.168.12.2/30
            port 1/1/c2/1:1000
            bfd 100 receive 100 multiplier 3
            no shutdown
        exit
```

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: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: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:

```
*A:PE-1>config>router>if# bfd ?
  - bfd <transmit-interval> [receive <receive-interval>] [multiplier <multiplier>] [echo-
receive
    <echo-interval>] [type <type>]
  - no bfd

 <transmit-interval>  : [10..100000] in milliseconds
 <receive-interval>   : [10..100000] in milliseconds
 <multiplier>         : [1..20]
 <echo-interval>      : [100..100000] in milliseconds
 <type>               : cpm-np - use CPM network processor
```

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 default value is 100 ms.

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:

```
*A:PE-1>config>router>if# bfd 10 receive 10 multiplier 3 type cpm-np
INFO: BFD #1001 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:

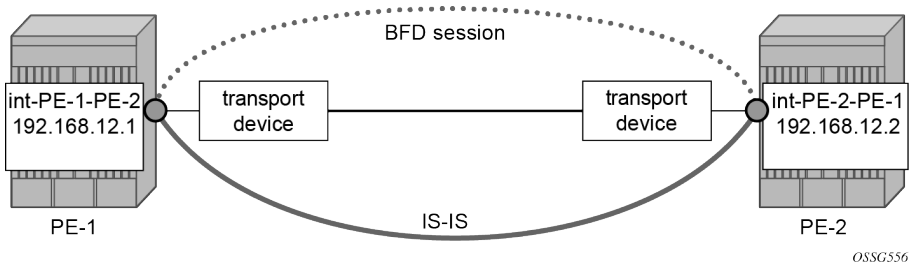
- lower Tx and Rx intervals are desired (down to 10 ms instead of 100 ms supported by local sessions)
- no more local (IOM) sessions are available
- the maximum limit of 500 packets per second per IOM has been reached

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. The topology used is shown in Figure 13: BFD for ISIS.

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-enable ipv4
            exit
```

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

```
*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
-------------------------------------------------------------------------------
int-PE-1-PE-2                                      Down           14          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-enable ipv4
            exit
```

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

On PE-1:

```
*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
-------------------------------------------------------------------------------
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: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: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/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: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
============================
```

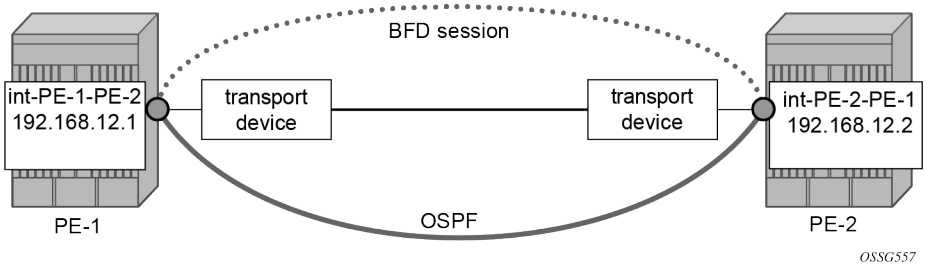
The **show router bfd session src <ip-address> detail** command can help debugging the BFD session. The sent and received counters are not supported for cpm-np type sessions.

```
*A: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      : 1253                     Sent Msgs        : 1305
Up Time        : 0d 00:01:38              Up Transitions   : 1
Last Down Time : 0d 00:00:46              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      : 04/06/2023 15:03:04      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      : 04/06/2023 15:03:04      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
            traffic-engineering
            area 0.0.0.0
                interface "system"
                    no shutdown
                exit
                interface "int-PE-1-PE-2"
                    interface-type point-to-point
                    bfd-enable
                    no shutdown
                exit
            exit
            no shutdown
```

```
# on PE-2:
configure
    router Base
        ospf 0
            traffic-engineering
            area 0.0.0.0
                interface "system"
                    no shutdown
                exit
                interface "int-PE-2-PE-1"
                    interface-type point-to-point
                    bfd-enable
                    no shutdown
                exit
            exit
            no shutdown
```

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

On PE-1:

```
*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
-------------------------------------------------------------------------------
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:PE-2# show router bfd session
```

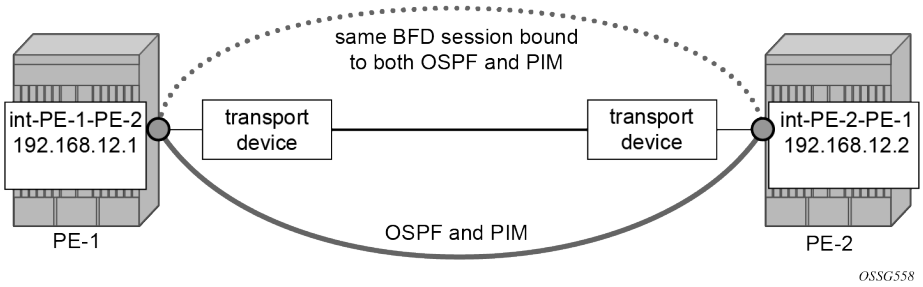
```
===============================================================================
Legend:
  Session Id = Interface Name | LSP Name | Prefix | RSVP Sess Name | Service Id
  wp = Working path   pp = Protecting path
===============================================================================
BFD Session
===============================================================================
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 were previously 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
            interface "int-PE-1-PE-2"
                bfd-enable
            exit
```

```
# on PE-2:
configure
    router Base
        pim
            interface "int-PE-2-PE-1"
```

```
                bfd-enable
            exit
```

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

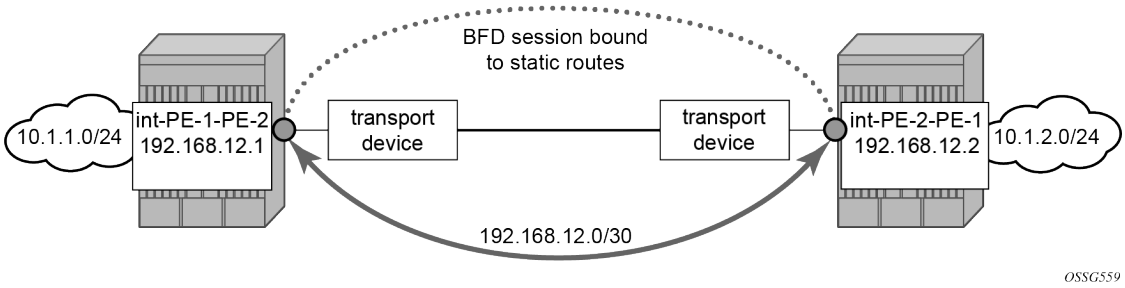
```
*A:PE-1# show router bfd session

===============================================================================
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
-------------------------------------------------------------------------------
int-PE-1-PE-2                                    Up          764        765
  192.168.12.2                                    3          100        100
  ospf2 pim                                     iom          N/A        N/A
  192.168.12.1
-------------------------------------------------------------------------------
No. of BFD sessions: 1
===============================================================================
```

```
*A: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          734        732
  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.

```
# on PE-1:
configure
    router Base
        static-route-entry 10.1.2.0/24
            next-hop 192.168.12.2
                bfd-enable
                no shutdown
            exit
```

```
# on PE-2:
configure
    router Base
        static-route-entry 10.1.1.0/24
            next-hop 192.168.12.1
                bfd-enable
                no shutdown
            exit
```

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

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

The following commands show the BFD session status on PE-1 and PE-2.

```
*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
-------------------------------------------------------------------------------
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: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
-------------------------------------------------------------------------------
```

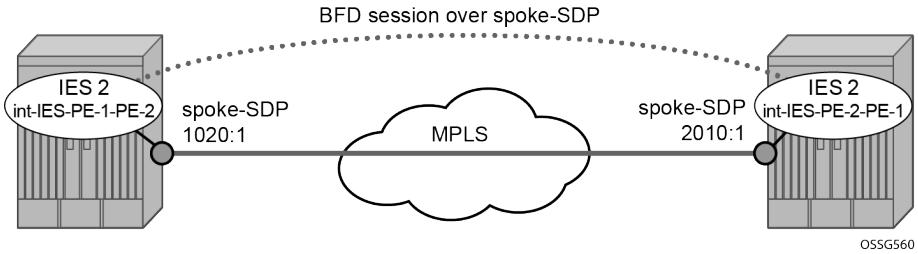
```
No. of BFD sessions: 1
==============================================================================
```

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 mpls create
            far-end 192.0.2.2
            sr-isis
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        ies 2 name "IES-2" customer 1 create
            interface "int-IES-PE-1-PE-2" create
                address 192.168.12.5/30
                spoke-sdp 1020:1 create
                exit
            exit
            no shutdown
        exit
    exit
    router Base
        ospf 0
            area 0.0.0.0
                interface "int-IES-PE-1-PE-2"
                exit
            exit
        exit
```

```
# on PE-2:
configure
    service
        sdp 2010 mpls create
```

```
            far-end 192.0.2.1
            sr-isis
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        ies 2 name "IES-2" customer 1 create
            interface "int-IES-PE-2-PE-1" create
                address 192.168.12.6/30
                spoke-sdp 2010:1 create
                exit
            exit
            no shutdown
        exit
    exit
    router Base
        ospf 0
            area 0.0.0.0
                interface "int-IES-PE-2-PE-1"
                exit
            exit
        exit
```

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

On PE-1:

```
*A: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: 04/06/2023 15:08:19
Last Mgmt Change  : 04/06/2023 15:08:05
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: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       39
```

```
   0.0.0.0
-------------------------------------------------------------------------------
No. of Neighbors: 2
===============================================================================
```

On PE-2:

```
*A: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: 04/06/2023 15:08:19
Last Mgmt Change  : 04/06/2023 15:08:12
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: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       31
   0.0.0.0
int-IES-PE-2-PE-1                192.0.2.1       Full       1    0       32
   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"
                bfd 1000 receive 1000 multiplier 3
            exit
        exit
    exit
    router Base
        ospf 0
```

```
            area 0.0.0.0
                interface "int-IES-PE-1-PE-2"
                    bfd-enable
                exit
            exit
```

```
# on PE-2:
configure
    service
        ies "IES-2"
            interface "int-IES-PE-2-PE-1"
                bfd 1000 receive 1000 multiplier 3
            exit
        exit
    exit
    router Base
        ospf 0
            area 0.0.0.0
                interface "int-IES-PE-2-PE-1"
                    bfd-enable
                exit
            exit
```

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: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: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
```

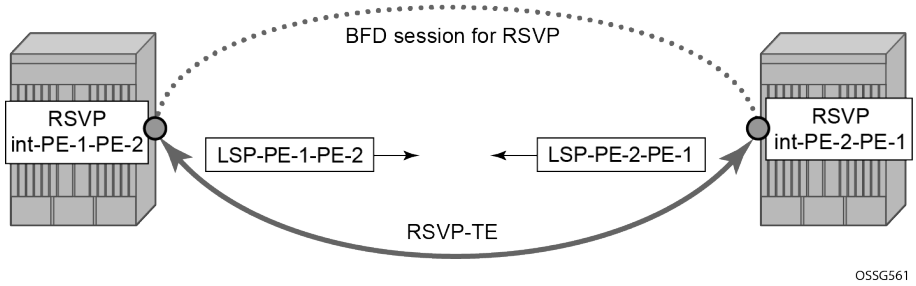
```
===============================================================================
Session Id                                        State      Tx Pkts    Rx Pkts
  Rem Addr/Info/SdpId:VcId                      Multipl     Tx Intvl   Rx Intvl
  Protocols                                        Type     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
            interface "system"
                no shutdown
            exit
            interface "int-PE-1-PE-2"
                no shutdown
            exit
        exit
        rsvp
            interface "system"
                no shutdown
```

```
            exit
            interface "int-PE-1-PE-2"
                no shutdown
            exit
            no shutdown
        exit
        mpls
            path "empty"
                no shutdown
            exit
            lsp "LSP-PE-1-PE-2"
                to 192.0.2.2
                path-computation-method local-cspf
                primary "empty"
                exit
                no shutdown
            exit
            no shutdown
        exit
```

```
# on PE-2:
configure
    router Base
        mpls
            interface "system"
                no shutdown
            exit
            interface "int-PE-2-PE-1"
                no shutdown
            exit
        exit
        rsvp
            interface "system"
                no shutdown
            exit
            interface "int-PE-2-PE-1"
                no shutdown
            exit
            no shutdown
        exit
        mpls
            path "empty"
                no shutdown
            exit
            lsp "LSP-PE-2-PE-1"
                to 192.0.2.1
                path-computation-method local-cspf
                primary "empty"
                exit
                no shutdown
            exit
            no shutdown
        exit
```

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

```
*A:PE-1# show router rsvp session

===============================================================================
RSVP Sessions
===============================================================================
RSVP Session Name
    From               To             Tunnel ID   LSP ID       State
```

```
-------------------------------------------------------------------------------
LSP-PE-1-PE-2::empty
192.0.2.1           192.0.2.2           1           20480        Up

LSP-PE-2-PE-1::empty
192.0.2.2           192.0.2.1           1           42496        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-enable
            exit
```

```
# on PE-2:
configure
    router Base
        rsvp
            interface "int-PE-2-PE-1"
                bfd-enable
            exit
```

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

On PE-1:

```
*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
-------------------------------------------------------------------------------
int-PE-1-PE-2                                   Up           97         91
  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:

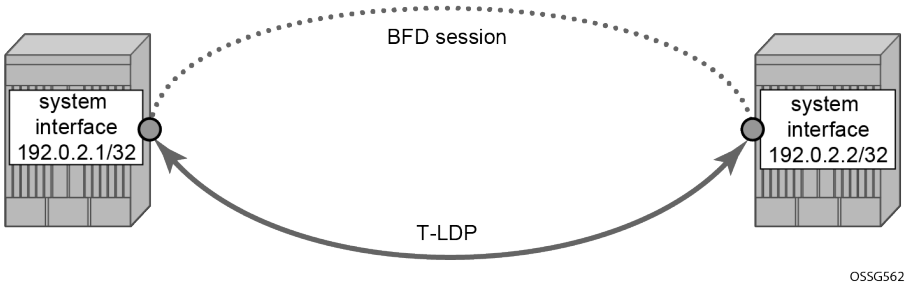
```
*A:PE-2# show router bfd session

===============================================================================
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           70         70
  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"
            bfd 3000 receive 3000 multiplier 3
```

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 re-converge 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
                    bfd-enable
                    no shutdown
                exit
```

```
# on PE-2:
configure
    router Base
        ldp
            targeted-session
                peer 192.0.2.1
                    bfd-enable
                    no shutdown
                exit
```

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: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   71        73        0d 00:05:51
-------------------------------------------------------------------------------
No. of IPv4 Sessions: 1
===============================================================================
```

On PE-2:

```
*A:PE-1# show router ldp session ipv4

===============================================================================
LDP IPv4 Sessions
===============================================================================
Peer LDP Id         Adj Type  State         Msg Sent  Msg Recv  Up Time
-------------------------------------------------------------------------------
192.0.2.2:0         Targeted  Established   71        73        0d 00:05:51
-------------------------------------------------------------------------------
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: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:

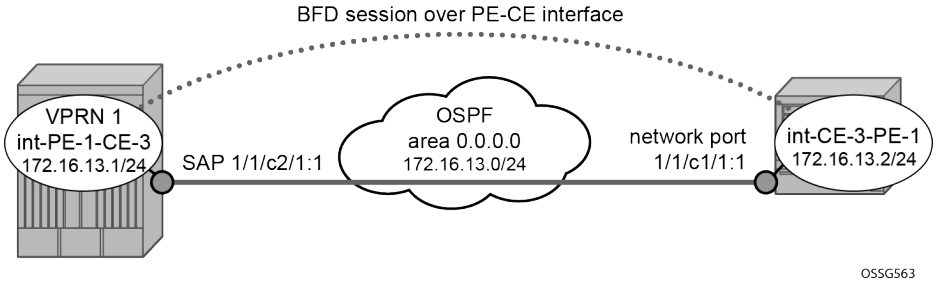
```
*A:PE-2# show router bfd session

===============================================================================
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 BFD session is always created (**cpm-np**) 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 1 configuration includes service interface int-PE-1-CE-1 with BFD parameters.

```
# on PE-1:
configure
    service
        vprn 1 name "VPRN-1" customer 1 create
            interface "int-PE-1-CE-3" create
                address 172.16.13.1/24
                bfd 100 receive 100 multiplier 3
                sap 1/1/c2/1:1 create
                exit
            exit
            ospf
                area 0.0.0.0
                    interface "int-PE-1-CE-3"
                        bfd-enable
                        no shutdown
                    exit
                exit
                no shutdown
            exit
            no shutdown
        exit
```

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"
            address 172.16.13.2/24
            port 1/1/c1/1:1
            bfd 100 receive 100 multiplier 3
            no shutdown
        exit
        interface "system"
            address 192.0.2.3/32
            no shutdown
        exit
        ospf 0
            area 0.0.0.0
                interface "int-CE-3-PE-1"
                    bfd-enable
                    no shutdown
                exit
            exit
```

```
            no shutdown
        exit
```

The following command shows that the OSPF adjacency is up.

On PE-1:

```
*A: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       33
   0.0.0.0
-------------------------------------------------------------------------------
No. of Neighbors: 1
===============================================================================
```

On CE-3:

```
*A: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       34
   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: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          507        500
  172.16.13.2                                     3          100        100
  ospf2                                         iom          N/A        N/A
  172.16.13.1
-------------------------------------------------------------------------------
No. of BFD sessions: 1
```

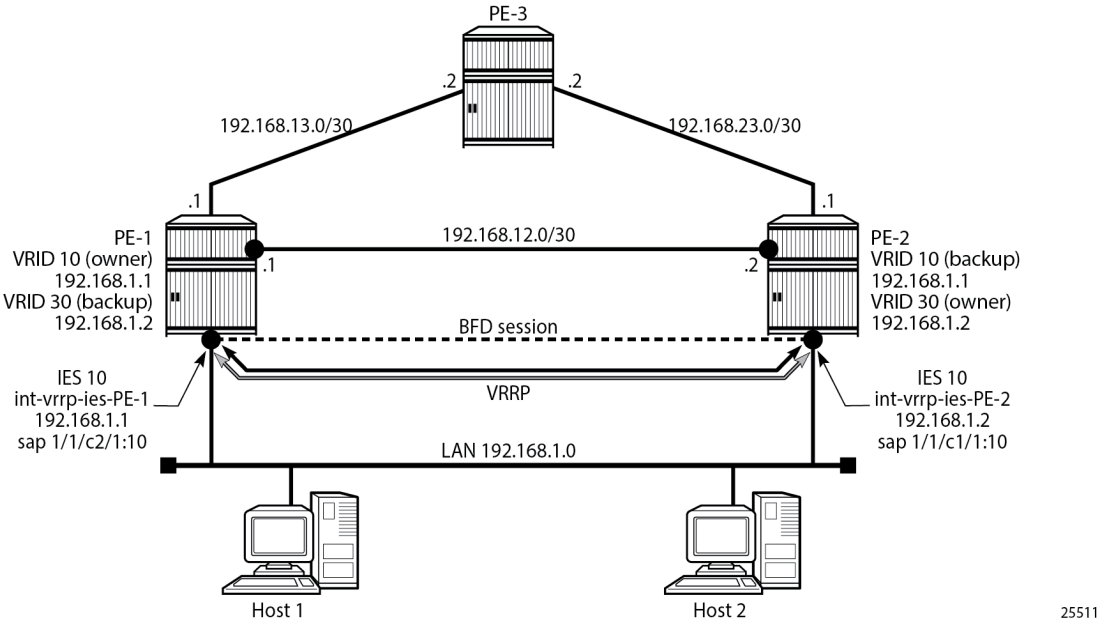
```
===============================================================================
```

```
*A: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          210        209
  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 10 name "IES-10" customer 1 create
            interface "int-vrrp-ies-PE-1" create
                address 192.168.1.1/24
                bfd 100 receive 100 multiplier 10
                sap 1/1/c2/1:10 create
                exit
            exit
            no shutdown
        exit
```

```
# on PE-2:
configure
    service
        ies 10 name "IES-10" customer 1 create
            interface "int-vrrp-ies-PE-2" create
                address 192.168.1.2/24
                bfd 100 receive 100 multiplier 10
                sap 1/1/c1/1:10 create
                exit
            exit
            no shutdown
        exit
```

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

```
*A: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:PE-1# ping 192.168.1.2 rapid
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 = 3.55ms, avg = 3.81ms, max = 4.01ms, stddev = 0.155ms
```

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.

The VRRP configuration on PE-1 is as follows:

```
configure
    service
        ies 10 name "IES-10" customer 1 create
            interface "int-vrrp-ies-PE-1" create
                vrrp 10 owner
                    backup 192.168.1.1
                exit
                vrrp 30
                    backup 192.168.1.2
                    ping-reply
                    telnet-reply
                    ssh-reply
                exit
```

The VRRP configuration on PE-2 is as follows:

```
configure
    service
        ies 10 name "IES-10" customer 1 create
            interface "int-vrrp-ies-PE-2"
                vrrp 10
                    backup 192.168.1.1
                    ping-reply
                    telnet-reply
                    ssh-reply
                exit
                vrrp 30 owner
                    backup 192.168.1.2
                exit
            exit
```

To bind the VRRP instances with a BFD session, add the following command under any VRRP instance: **bfd-enable name <service-name> interface <interface-name> dst-ip <ip-address>**. The IES service ID must be declared where the interface is configured. Instead of configuring the service name, it is possible to configure the service ID: **bfd-enable <service-id> interface <interface-name> dst-ip <ip-address>**.

On PE-1, the following commands enable BFD in IES "IES-10" for VRRP 10 and VRRP 30:

```
configure
    service
        ies 10 name "IES-10" customer 1 create
            interface "int-vrrp-ies-PE-1"
                vrrp 10 owner
                    bfd-enable name "IES-10" interface "int-vrrp-ies-PE-1" dst-ip 192.168.1.2
                exit
```

```
                vrrp 30
                    bfd-enable name "IES-10" interface "int-vrrp-ies-PE-1" dst-ip 192.168.1.2
                exit
            exit
```

On PE-2, the following commands enable BFD in IES "IES-10" for VRRP 10 and VRRP 30:

```
configure
    service
        ies 10 name "IES-10" customer 1 create
            interface "int-vrrp-ies-PE-2"
                vrrp 10
                    bfd-enable name "IES-10" interface "int-vrrp-ies-PE-2" dst-ip 192.168.1.1
                exit
                vrrp 30 owner
                    bfd-enable name "IES-10" interface "int-vrrp-ies-PE-2" dst-ip 192.168.1.1
                exit
            exit
```

The parameters used for the BFD are set by the BFD command under the IP interface. Unlike the previous scenarios, the user can enter the preceding commands, enabling the BFD session, even if the specified interface (int-vrrp-ies-PE-1) has not been configured with BFD parameters.

If the BFD parameters have not been configured yet, the BFD session will be initiated only after the following configuration:

```
# on PE-1:
configure
    service
        ies 10 name "IES-10" customer 1 create
            interface "int-vrrp-ies-PE-1" create
                bfd 100 receive 100 multiplier 10
```

```
# on PE-2:
configure
    service
        ies 10 name "IES-10" customer 1 create
            interface "int-vrrp-ies-PE-2" create
                bfd 100 receive 100 multiplier 10
```

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

```
*A: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      : 36033                    Sent Msgs        : 36032
Up Time        : 0d 03:00:45              Up Transitions   : 1
Last Down Time : 0d 00:00:10              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       : 10
Last Sent      : 04/18/2023 09:50:22      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      : 10
Remote C-flag  : 1
Last Recv      : 04/18/2023 09:50:22      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: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             : 04/18/2023 06:49:27 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          : 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        : 04/18/2023 06:49:27

-------------------------------------------------------------------------------
Masters Seen (Last 32)
-------------------------------------------------------------------------------
Primary IP of Master   Last Seen             Addr List Mismatch     Msg Count
-------------------------------------------------------------------------------
192.168.1.1            04/18/2023 06:49:27   No                           0

-------------------------------------------------------------------------------
Statistics
```

```
-------------------------------------------------------------------------------
Become Master       : 1                   Master Changes    : 1
Adv Sent            : 10948               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             : 04/18/2023 06:49:27 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          : 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        : 04/18/2023 06:49:34
Master Down Interval: 3.609 sec (Expires in 2.700 sec)

-------------------------------------------------------------------------------
Masters Seen (Last 32)
-------------------------------------------------------------------------------
Primary IP of Master   Last Seen             Addr List Mismatch     Msg Count
-------------------------------------------------------------------------------
192.168.1.1            04/18/2023 06:49:31   No                            0
192.168.1.2            04/18/2023 09:51:54   No                        10942

-------------------------------------------------------------------------------
Statistics
-------------------------------------------------------------------------------
Become Master       : 1                   Master Changes    : 2
Adv Sent            : 4                   Adv Received      : 10942
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"
                vrrp 10 owner
                    no bfd-enable name "IES-10" interface "int-vrrp-ies-PE-1" dst-ip
 192.168.1.2
                    bfd-enable interface "int-vrrp-ies-PE-1" dst-ip 192.168.1.2
                exit
```

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: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             : 04/18/2023 06:49:27 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 ID 10 or service name "IES-10" for VRRP instance 10:

```
# on PE-1:
configure
    service
        ies "IES-10"
            interface "int-vrrp-ies-PE-1"
                vrrp 10 owner
                    no bfd-enable interface "int-vrrp-ies-PE-1" dst-ip 192.168.1.2
                    bfd-enable name "IES-10" interface "int-vrrp-ies-PE-1" dst-ip 192.168.1.2
                exit
```

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

Hybrid OpenFlow Switch

This chapter provides information about Hybrid OpenFlow Switch.

Topics in this chapter include:

Applicability

The information and configuration in this chapter are based on SR OS Release 14.0.R5.

Overview

OpenFlow is defined by the Open Networking Foundation and provides a standard interface between the control layer and forwarding layer of a Software Defined Networking (SDN) architecture. The control layer has northbound interfaces to the application layer and translates the requirements from this layer into low-level control protocols on its southbound interfaces toward the forwarding layer. Made up of SDN controllers, the control layer provides an abstraction between the application layer and the forwarding layer. The forwarding layer may consist of physical and/or virtual network elements.

An OpenFlow controller operates at the control layer while an OpenFlow switch operates at the forwarding layer, and the OpenFlow protocol is used for communication between them. The term Hybrid OpenFlow switch refers to switches or routers that fully integrate both OpenFlow operation and conventional Ethernet switching or IP routing. Conversely, OpenFlow-only switches support only OpenFlow operations. SR OS platforms operate as Hybrid OpenFlow switches.

An OpenFlow switch may have one or more flow tables, each of which contains one or more flow entries. A flow is a sequence of packets that matches a specific entry in a flow table. When a packet is processed by a flow table, it is matched against flow entries that contain match fields and a priority to uniquely identify each entry. Match fields consist of criteria to match against a packet, such as ingress port/VLAN, source/destination IP address, protocol, or source/destination port.

The sequence with which a packet is parsed through a flow table that consists of multiple flow entries depends on the priority of each flow entry. The highest priority flow entry is processed first, and if no match is found, the packet continues to the next highest flow entry until a packet is either matched by a flow entry or all flow entries are parsed and no match is found. Priority 0 is reserved for the table-miss flow entry, which is used when a packet does not match any other flow entries in the flow table. In this case, the packet could be forwarded, dropped, or sent to the OpenFlow controller using a Packet-In message.

Each flow entry consists of one or more OpenFlow Protocol Instruction Types (OFPITs) that collectively form an instruction set. The instruction type defines the type of action to be taken, such as Write-Action, Write-Metadata, or Clear-Action. Each instruction type contains an OpenFlow Protocol Action Type

(OFPAT), and the group of actions associated with a flow entry is referred to as an action set. These actions may be to manipulate a packet, or rate-limit packets matching this flow entry, or to output to a specific port, where port may be physical, logical (such as an MPLS or VXLAN tunnel), or a reserved port (such as the control channel with the OpenFlow controller).

Each flow entry is also associated with a 64-bit opaque cookie value assigned by the OpenFlow controller. However, this cookie value is not used for packet lookup or processing. Its purpose is to enable the controller to filter flow statistics and for flow modification/deletion. In SR OS, the cookie value is also used to distinguish between flow entries associated with the base routing instance and those associated with services, as described in more detail later in this chapter.

OpenFlow Protocol

The OpenFlow channel is the interface that connects the OpenFlow switch to a controller and runs over TCP port 6653. By default, the OpenFlow channel is a single TCP connection, but auxiliary connections are also supported. These auxiliary connections may be used for general OpenFlow messages, but are intended to allow for parallel processing of statistics requests and Packet-In messages. Auxiliary connections use the same destination IP address and destination port as the main channel, but are uniquely identified by having a different combination of the switch Datapath ID and an Auxiliary ID on the OpenFlow switch.

The Datapath ID is an 8-byte value used to uniquely identify the switch. To construct it, SR OS uses a concatenation of the OpenFlow switch instance ID (2 bytes) and the chassis MAC (6 bytes). Because the Datapath ID is a switch-wide parameter, it is common to all connections from a switch, but the Auxiliary ID is unique for each channel. In SR OS, the primary channel uses an Auxiliary ID of zero, while auxiliary channels use a unique non-zero value. The Datapath ID and Auxiliary ID are exchanged during the connection setup. After the OpenFlow session is established and Hello messages exchanged, the controller requests a list of supported features from the switch (using an OFPT_FEATURES_REQUEST message). The response from the switch (OFPT_FEATURES_REPLY) contains (among other things) the Datapath ID and Auxiliary ID.

The OpenFlow protocol supports three message types: controller-to-switch, asynchronous, and symmetric.

- Controller-to-switch messages are initiated by the controller and are used to manage or inspect the state of the OpenFlow switch.
- Asynchronous messages are initiated by the switch and are used to notify the controller of network events and changes to switch state.
- Symmetric messages are initiated by either switch or controller and are sent in an unsolicited manner.

OpenFlow specifies the use of a number of different messages within its operation and the use of these messages is constrained to the message type to which they are associated. Table 1 lists the various OpenFlow messages associated with each message type, with a brief description of its usage. Some of the messages will be referred to throughout this chapter, with examples of how and when they are used.

Table 1: OpenFlow Messages

Message Type	Message	Description
Controller-to-switch	Feature	[OFPT_FEATURES_REQUEST/REPLY] Used by controller to query capabilities of the switch. Typically used on session establishment.

Message Type	Message	Description
	Configuration	[OFPT_GET_CONFIG_REQUEST/REPLY, OFPT_SET_CONFIG] Used to set and query configuration parameters in the switch.
	Modify-State	[OFPT_FLOW/PORT/TABLE_MOD] Used to add, delete, and modify flow entries in the OpenFlow tables.
	Read-State	Used to collect information such as configuration, statistics, and capabilities from the switch.
	Packet-Out	[OFPT_PACKET_OUT] Used by the controller to send packets out of a specific port on the switch, and to forward packets received in Packet-In messages.
	Barrier	[OFPT_BARRIER_REQUEST/REPLY] Used to ensure that messages prior to the barrier are processed before any messages after the barrier. Allows for ordering of message processing.
	Role-Request	[OFPT_ROLE_REQUEST/REPLY] Used to set the role of the OpenFlow channel. Can be Master, Slave, or Equal. When multiple controllers are used, only one can be set to Master.
	Asynchronous-Configuration	[OFPT_GET_ASYNC_REQUEST/REPLY, OFPT_SET_ASYNC] Used by the controller to set a filter on the asynchronous messages that it needs to receive.
Asynchronous	Packet-In	[OFPT_PACKET_IN] Used to transfer a packet to the controller (for example, a table-miss flow entry).
	Flow-Removed	[OFPT_FLOW_REMOVED] Used to notify the controller that a flow entry has been removed from the flow table.
	Port-Status	[OFPT_PORT_STATUS] Used to notify the controller of a change in the configuration or status of a port.
	Error	[OFPT_ERROR] Used to notify the controller of an error.

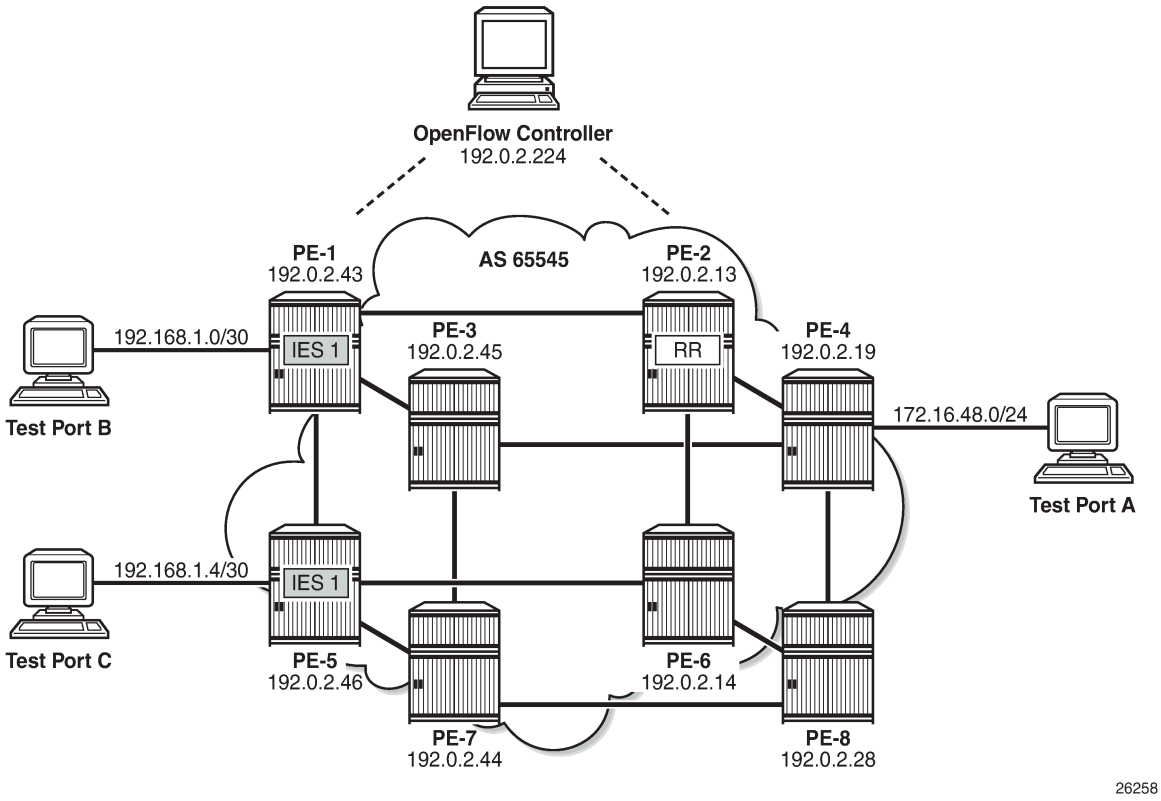
Message Type	Message	Description
Symmetric	Hello	[OFPT_HELLO] Exchanged between controller and switch during session startup.
	Echo	[OFPT_ECHO_REQUEST/REPLY] Used to maintain the liveliness of the OpenFlow channel.
	Experimenter	[OFPT_EXPERIMENTER] Provides a standard way to offer additional proprietary functionality within the standard message space.

OpenFlow messages have a standard header that includes the version of the protocol. SR OS supports OpenFlow specification 1.3.1, which requires the use of OpenFlow protocol version 4. Although the OpenFlow protocol defines the standard through which controllers and switches should communicate, it also allows for additional functionality to be implemented, using Experimenter messages and fields. SR OS uses Experimenter fields as additional match criteria and action types.

Configuration

Figure 22: Example Topology shows an example topology to demonstrate the use of OpenFlow. PE routers PE-1 through PE-8 form part of AS 65545 and run IS-IS and RSVP. All PE routers are IBGP clients of a Route Reflector situated at PE-2 for the IPv4 and VPN-IPv4 address families. An OpenFlow Controller is at address 192.0.2.224, which is reachable from AS 65545. Test port A is connected to PE-4, and test ports B and C are connected to PE-1 and PE-5, respectively. These test ports will be configured to advertise routes and source/sink traffic within the base and service routing contexts to verify OpenFlow operation. More information about specific configurations will be provided within the relevant parts of this chapter.

Figure 22: Example Topology



OpenFlow Switch Configuration

OpenFlow specification 1.3.1 allows for multiple flow tables within an OpenFlow switch that are sequentially numbered starting at zero. A function referred to as pipeline processing subsequently matches packets, first against flow entries of flow table 0, but allows for instructions to optionally direct a packet to another flow table, where the process is repeated. Up to eight Hybrid OpenFlow switch instances can be supported per system. Each switch instance supports a single flow table: table 0.

Flow entries pushed from an OpenFlow controller are dynamically embedded within ingress IP filters provisioned on the system. Within the OpenFlow specification, there is no provision for enabling context-specific flow entries. That is, it is not possible to enable a flow entry explicitly within the base routing context, or enable a flow entry explicitly within a service or VPN context. To overcome this, and provide maximum flexibility without the requirement for proprietary extensions, SR OS makes intelligent use of the 64-bit cookie value that is associated with every flow entry in a Modify Flow Entry (OFPT_FLOW_MOD) message. The high-order 32 bits of the value are subdivided into two parts. Bits 63 to 60 are used to determine whether the flow entry is applicable to a filter on an IES or router interface in the base routing context (also referred to as the Global Routing Table [GRT]), or a System filter, or a filter applied within a VPRN or VPLS service context. For the latter, bits 59 to 32 are then used to define the service ID value. This use of the cookie value in this manner is referred to as a multi-service OpenFlow switch instance.

Table 2: FLOW_MOD Cookie Value

Bits 63 to 60	Bits 59 to 32	Bits 31 to 0	SR OS context
0000	0	Arbitrary	Used by filters in base router
1000	0	Arbitrary	Used by System filter policy
1100	Service ID value	Arbitrary	Used by filters policies within specified service

Note:
The use of a System filter allows for a common rule set defined in an IP filter with a scope of System to be embedded in multiple interface filters, reducing configuration requirements and increasing system scale. The use of System filters is not described within this chapter.

The following output shows the configuration required to define the Hybrid OpenFlow switch and to establish connectivity with the OpenFlow controller:

```
configure
    open-flow
        of-switch "ofs-1"
            aux-channel-enable
            controller 192.0.2.224:6653
            flowtable 0
                switch-defined-cookie
                max-size 4096
            exit
            logical-port-status rsvp-te
            no shutdown
        exit
    exit
```

The **of-switch** command allows for the creation of a switch instance and requires a name of 1 to 32 characters. This creates a new **of-switch** context under which the characteristics of the switch are defined. The **controller** command requires the destination IP address and port of the OpenFlow controller to be entered, separated by a colon. Port 6653 is the standard IANA assigned port for OpenFlow. When this connection is successfully established, it creates the primary channel (with Auxiliary-ID 0) only.

The output shows the configuration of a single controller, but it is possible to configure multiple controllers for redundancy; each controller may create/modify/delete flows entries in the flow table of this switch instance. Also, the OpenFlow switch can use both in-band (base routing context) and out-of-band (management routing context) to establish connectivity with the controller, with preference given to out-of-band if a valid route exists.

The **aux-channel-enable** command establishes the auxiliary channels. When enabled, this command creates an auxiliary channel for statistics with Auxiliary-ID 1, and an auxiliary channel for Packet-In messages with Auxiliary-ID 2. Although these auxiliary channels are assigned an explicit purpose, the switch will still accept any generic OpenFlow messages over these auxiliary channels and will respond in return on the same channel. The **flowtable** command modifies the characteristics of flow table 0. The **max-size** command configures a limit on the number of flow entries that can be populated within each flow-table. Flow-table entries are created in hardware on the line-card datapath and consume Content-Addressable Memory (CAM) entries; therefore, placing a limit on how much of that resource is used by OpenFlow may be needed. The **switch-defined-cookie** command enables the use of a multi-instance OpenFlow switch. This is the recommended approach for deploying an OpenFlow switch in SR OS; it allows for creation of service-specific flow entries, and offers an increased number of traffic actions.

Finally, the **logical-port-status rsvp-te** command instructs the switch to report configuration and/or state changes to RSVP-TE logical ports to the controller, which is achieved using asynchronous Port-Status (OFPT_PORT_STATUS) messages.

When the OpenFlow switch is put into a **no shutdown** state, its operational state can be verified with the command shown in the following output:

```
*B:PE-4# show open-flow of-switch "ofs-1" status

===============================================================================
Open Flow Switch Information
===============================================================================
Switch Name        : ofs-1
Data Path ID       : 00030ca40202d401    Admin Status       : Up
Echo Interval      : 10 seconds          Echo Multiple      : 3
Logical Port Type  : rsvp-te
Buffer Size        : 0                   Num. of Tables     : 1
Description        : (Not Specified)
Capabilities Supp. : flow-stats table-stats port-stats
Aux Channel Enabled: True
===============================================================================
```

The output shows the switch Datapath ID, which together with the Auxiliary ID uniquely identifies a (primary/auxiliary) channel between switch and controller. The output also shows the logical port types in use as being RSVP-TE, support for a single flow table, and a buffer size of 0. The buffer size is used when Packet-In messages are used for a table-miss.

The OpenFlow specification provides an option for a switch to truncate the packet and send only a portion of the packet to the controller in a Packet-In message, together with a buffer-ID, while the remainder of the packet is buffered. When the controller subsequently responds with a Packet-Out message containing a corresponding buffer-ID, the packet is retracted from buffer, re-assembled, and forwarded through the port specified in the Packet-Out message. Rather than buffering, SR OS sends the complete packet to the controller in a Packet-In message, so requires no buffer. Also, SR OS sends only the first packet of a flow in a Packet-In message; any subsequent packets of that flow are dropped at ingress. This avoids overwhelming the controller with table-miss packets, and equally offers a level of protection to the CPM. The expectation is that the controller should create a new flow entry for that flow.

The following output shows the status of the OpenFlow channel to the controller:

```
*B:PE-4# show open-flow of-switch "ofs-1" controller 192.0.2.224:6653 detail

===============================================================================
Open Flow Controller Information
===============================================================================
IP Address            : 192.0.2.224       Port              : 6653
Role                  : equal
Generation ID         : 0

-------------------------------------------------------------------------------
Open Flow Channel Information - Channel ID(1)
-------------------------------------------------------------------------------
Channel ID        : 1                 Version           : 4
Connection Type   : primary           Operational Status: Up
Auxiliary ID      : 0
Source Address    : 192.0.2.19        Source Port       : 56261
Operational Flags : socket-state-established hello-received hello-transmitted
                    handshake
Async Fltr Packet In
 (Master or Equal): table-miss apply-action
 (Slave)          : (Not Specified)
```

```
Async Fltr Port Status
 (Master or Equal): port-add port-delete port-modify
 (Slave)          : port-add port-delete port-modify
Async Fltr Flow Rem
 (Master or Equal): idle-time-out hard-time-out flow-mod-delete group-delete
 (Slave)          : (Not Specified)
Echo Time Expiry  : 0d 00:00:01       Hold Time Expiry  : 0d 00:00:21
Conn. Uptime      : 0d 06:09:59       Conn. Retry       : 0d 00:00:00

-------------------------------------------------------------------------------
Open Flow Channel Stats - Channel ID(1)
-------------------------------------------------------------------------------
Packet Type      Transmitted Packets  Received Packets     Error Packets
-------------------------------------------------------------------------------
Hello            1                    1                    0
Error            1                    0                    0
Echo Request     1722                 508                  0
Echo Reply       508                  1722                 0
Experimenter     0                    0                    0
Feat. Request    0                    1                    0
Feat. Reply      1                    0                    0
---snip---
```

The complete output would show the details of all the OpenFlow channels between the switch and the controller. As the **aux-channel-enable** command is configured, there are three channels in total, but only the primary channel (Auxiliary ID 0) is shown for brevity.

Each controller is assigned a role, which can be master, slave, or equal, with the default being equal. The role determines what access the controller has to the switch and also what asynchronous messages the switch should forward to the controller:

- Equal role: The controller has full access to the switch and is considered equal to other controllers in the same role. All controllers should receive asynchronous messages from the switch.
- Slave role: The controller has read-only access to the switch. Controllers do not receive asynchronous messages from the switch apart from Port-Status messages.
- Master role: The controller has full access to the switch, but only one controller can have the role of master.

If a controller changes its role to master using a Role-Request (OFPT_ROLE_REQUEST) message, the switch modifies all other connections to the role of slave. To ensure that the switch has the latest information on a controller mastership election, controllers coordinate the assignment of a Generation ID, also shown in the output. The Generation ID is a monotonically increasing 64-bit counter; therefore, any OFPT_ROLE_REQUEST message received with a role of master or slave with Generation ID of a lower value than one previously received is ignored.

The version, connection type, and Auxiliary ID have been previously described.

The output shows asynchronous filters (Async Fltr), dependent on the role that the controller is playing. A controller may use Asynchronous Configuration (OFPT_SET_ASYNC) messages to set a filter on the asynchronous messages that it receives from the switch. In the absence of an OFPT_SET_ASYNC message from the controller, the switch sets an initial configuration of asynchronous messages for Packet-In, Port-Status, and Flow-Removal messages and this configuration is shown. The remainder of the output (again truncated) shows detailed statistics for all message types sent and received over this channel.

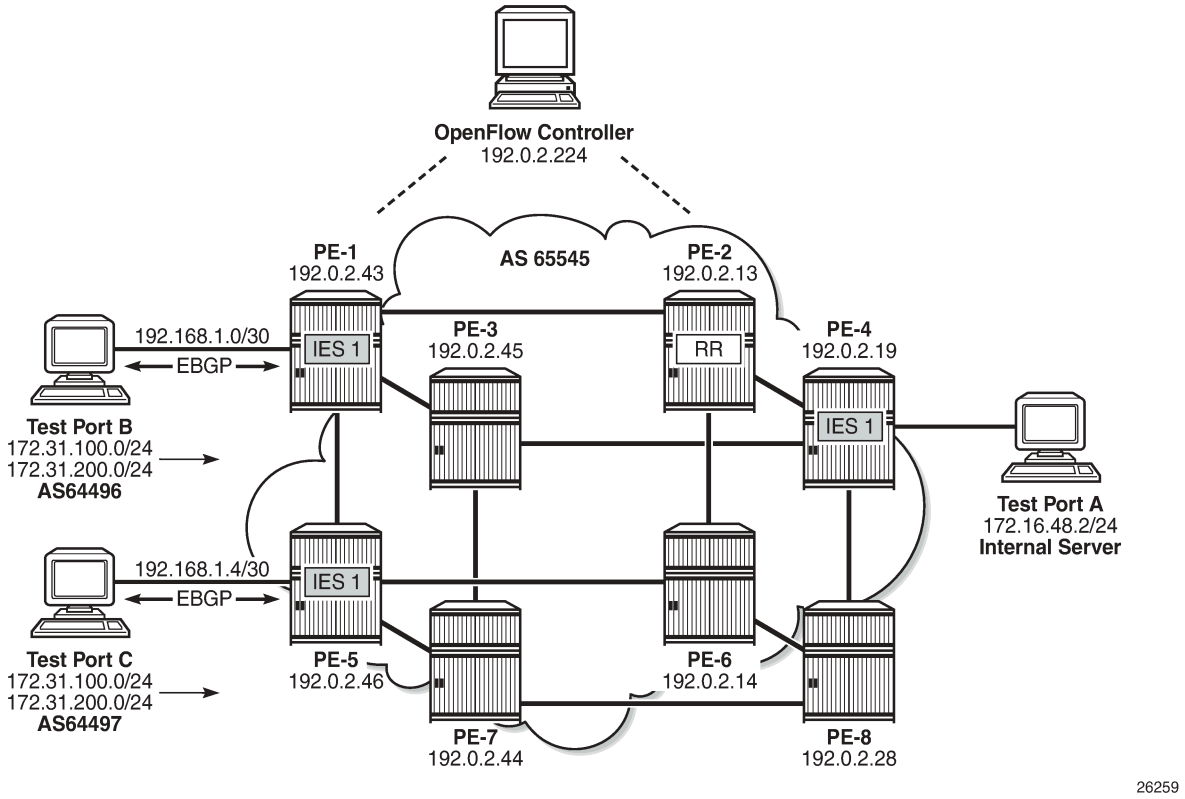
Dynamic Flow Entry Creation

With the basic switch configured and a channel established to the controller, the next step is to configure one or more IP filters that allow for dynamic embedding of OpenFlow flow entries. The following section will describe flow entries created in the base routing instance (also referred to as GRT), followed by entries created within a service instance.

Base Routing Instance

To generate rules within the base routing instance, the example topology is configured as shown in Figure 23: OpenFlow Operation in Base Routing Context. Test ports B and C simulate external peers located in AS 64496 and 64497, respectively. Both external peers advertise prefixes 172.31.100.0/24 and 172.31.200.0/24 to AS 64496, which are propagated internally within AS 65545. PE-4 hosts an internal server on subnet 172.16.48.0/24, which is advertised to the external peers. All three test ports are indexed to IES 1 at the corresponding PE router. BGP is configured within AS 65545 such that next-hops are resolved to shortcut tunnels using RSVP.

Figure 23: OpenFlow Operation in Base Routing Context



An IP filter is configured using the **embed-filter open-flow** command to allow for dynamic embedding of flow entries by an OpenFlow switch instance. In this example, the OpenFlow switch is the previously configured ofs-1. IP filters allow dynamically embedded OpenFlow filter entries to co-exist with static filter entries and other dynamic filter entries created by Flowspec or RADIUS. Therefore, an **offset** is defined

that specifies the start point for dynamically created OpenFlow entries. This ensures that the OpenFlow flow entries can be isolated from other dynamic and static filter entries; FlowSpec filter entries must be created after static entries. In this example, the offset is 100.

```
configure
    filter
        ip-filter 10 create
            description "OpenFlow Basic GRT Filter"
            embed-filter open-flow "ofs-1" offset 100
        exit
    exit
```

The filter is applied as an ingress filter at PE-4 on the SAP connecting test port A, as follows:

```
configure
    service
        ies 1 customer 1 create
            interface "Test-Port-A" create
                address 172.16.48.1/24
                sap 3/1/4:10 create
                    ingress
                        filter ip 10
                    exit
                exit
```

Before any flow entries are initiated from the controller, a single entry with ID 65535 (maximum) is automatically populated in the embedding filter. This entry is inserted by OpenFlow when the **embed-filter open-flow** command is configured in the **filter** context and represents the table-miss entry. When OpenFlow uses filters, it ignores any **default-action** that may be configured in the filter so that filters can be chained. However, a table-miss action must exist and this is represented by entry 65535.

The source/destination addresses are 0.0.0.0/0 and the default primary action is forward (also referred to as fall-through). This primary action is configurable using the **no-match-action** command within the **flowtable 0** context. Other actions include **packet-in** or **drop**. When **packet-in** is configured and a packet of a flow matches entry 65535 (table-miss), SR OS sends only the first packet of that flow to the controller in a Packet-In message, while subsequent packets of that same flow are dropped.

```
*B:PE-4# show filter ip 10
===============================================================================
IP Filter
===============================================================================
Filter Id           : 10                            Applied        : Yes
Scope               : Template                      Def. Action    : Drop
System filter       : Unchained
Radius Ins Pt       : n/a
CrCtl. Ins Pt       : n/a
RadSh. Ins Pt       : n/a
PccRl. Ins Pt       : n/a
Entries             : 0/0/0/1 (Fixed/Radius/Cc/Embedded)
Description         : OpenFlow Basic GRT Filter
-------------------------------------------------------------------------------
Filter Match Criteria : IP
-------------------------------------------------------------------------------
Entry               : 65535
Origin              : Inserted by open-flow (no-match-action)
Description         : (Not Specified)
Log Id              : n/a
Src. IP             : 0.0.0.0/0
Src. Port           : n/a
Dest. IP            : 0.0.0.0/0
```

```
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-syn             : Off                         TCP-ack        : Off
Option-pres         : Off
Egress PBR          : Disabled
Primary Action      : Forward
Ing. Matches        : 0 pkts
Egr. Matches        : 0 pkts

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

An OpenFlow IP filter is also automatically created by the system with a filter ID of _tmnx_ofs_<name>:<number>, where <name> is the name of the OpenFlow switch instance and <number> is a numerical integer. This is shown in the following output as _tmnx_ofs_ofs-1:8. This system-created filter ID contains all of the active flow entries dynamically created by the OpenFlow switch ofs-1 for the base (GRT) context, effectively acting as a repository for that routing context.

Any filter that is subsequently configured to dynamically embed GRT OpenFlow filter entries from the same OpenFlow switch instance will inherit all of the current entries contained in this filter. That is, if a new filter is configured to embed GRT OpenFlow entries from ofs-1, all of the flow-entries contained in _tmnx_ofs_ofs-1:8 will be automatically cloned into that new filter. There is no requirement for any active flow entries to be re-sent by the controller in order to populate this new filter. This approach allows filters to be enabled for OpenFlow embedding before or after flow entries have been received by the OpenFlow switch, thereby removing any order dependency.

```
*B:PE-4# show filter ip filter-type openflow
===============================================================================
Openflow IP Filters                                          Total:     1
===============================================================================
Filter-Id                     Description
-------------------------------------------------------------------------------
_tmnx_ofs_ofs-1:8             Filter for OFS 'ofs-1' for grt context
```

OpenFlow Filtering in Action

Before initiating any flow entries from the controller, the following traffic flows are sourced from test port A connected to PE-4.

- A UDP-based flow with a destination IP address of 172.31.100.1/24 at a rate of 1000 packets/s.
- A UDP-based flow with a destination address of 172.31.200.1/24, again at a rate of 1000 packets/s.

Both test port B and C are advertising the preceding prefixes, which are advertised internally by PE-1 and PE-5, respectively. At PE-4, the preferred next-hop for these prefixes is PE-1 (192.0.2.43).

```
*B:PE-4# show router bgp routes 172.31.0.0/16 longer
===============================================================================
 BGP Router ID:192.0.2.19       AS:65545       Local AS:65545
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
```

```
BGP IPv4 Routes
===============================================================================
Flag  Network                                          LocalPref   MED
      Nexthop (Router)                                 Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
u*>i  172.31.100.0/24                                  100         0
      192.0.2.43                                       None        -
      64496 64500
u*>i  172.31.200.0/24                                  100         0
      192.0.2.43                                       None        -
      64496 64500
-------------------------------------------------------------------------------
Routes : 2
===============================================================================
```

BGP next-hops are resolved to RSVP shortcut tunnels. For this test, there are two RSVP LSPs, one to PE-1 and one to PE-5, and they are viewed as logical ports by the OpenFlow switch. Because PE-1 is the preferred next-hop for the advertised prefixes, this resolves to the LSP named PE-4-PE-1-RSVP.

```
*B:PE-4# show open-flow of-switch "ofs-1" port
===============================================================================
Open Flow Port Stats
===============================================================================
Port ID        Type           Transmitted Packets      Transmitted Bytes
Port Name
-------------------------------------------------------------------------------
1073741833     logical        0                        0
PE-4-PE-1-RSVP
1073741834     logical        0                        0
PE-4-PE-5-RSVP
===============================================================================
```

The following output shows, as expected, that PE-1 is egressing traffic at a rate of 2000 packets/s toward test port B, representing the sum of the two 1000 packets/s test streams.

```
B:PE-1# monitor service id 1 sap 5/1/3:10 rate
===============================================================================
Monitor statistics for Service 1 SAP 5/1/3:10
===============================================================================
---snip---
-------------------------------------------------------------------------------
At time t = 11 sec (Mode: Rate)
-------------------------------------------------------------------------------
                        Packets                 Octets                  % Port
                                                                        Util.
Egress Queue 1
For. In/InplusProf    : 0                       7                       ~0.00
For. Out/ExcProf      : 2000                    1023907                 0.08
Dro. In/InplusProf    : 0                       0                       0.00
Dro. Out/ExcProf      : 0                       0                       0.00
```

The controller initiates an OFPT_FLOW_MOD message containing an OFPFC_ADD command to the switch to create a new flow entry. The flow entry is viewed using the command shown in the following output:

```
*B:PE-4# tools dump open-flow of-switch "ofs-1"
===============================================================================
Switch: ofs-1
===============================================================================
Table     : 0                           Flow Pri  : 0
```

```
Cookie    : 0x0000000000000000          CookieType: grt
Controller: :::0
Filter Hnd: 0xC30000080000FFFF
Filter    : _tmnx_ofs_ofs-1:8 entry 65535

In Port   : *
VID       : *                           Outer VID : *
EthType   : *
Src IP    : *
Dst IP    : *
IP Proto  : *                           DSCP      : *
Src Port  : *                           Dst Port  : *
ICMP Type : *                           ICMP Code : *
Label     : *
IPv6ExtHdr: (Not Specified)

Action    : Fall-through

Flow Flags: IPv4/6 [!E] [RO] [DEF]
Up Time   : 0d 00:18:47                 Add TS    : 405757580
Mod TS    : 0                           Stats TS  : 405870240
#Packets  : 1638207                     #Bytes    : 838761984
-------------------------------------------------------------------------------
Table     : 0                           Flow Pri  : 1635
Cookie    : 0x0000000000000100          CookieType: grt
Controller: 192.0.2.224:6653
Filter Hnd: 0x8300000080000F99C
Filter    : _tmnx_ofs_ofs-1:8 entry 63900

In Port   : *
VID       : *                           Outer VID : *
EthType   : 0x0800
Src IP    : *
Dst IP    : 172.31.100.0/24
IP Proto  : *                           DSCP      : *
Src Port  : *                           Dst Port  : *
ICMP Type : *                           ICMP Code : *
Label     : *

Action    : Forward LspId 10
            Lsp PE-4-PE-5-RSVP

Flow Flags: IPv4 [FR]
Up Time   : 0d 00:01:57                 Add TS    : 405858646
Mod TS    : 0                           Stats TS  : 405870241
#Packets  : 115951                      #Bytes    : 59366912
-------------------------------------------------------------------------------
Number of flows: 2
===============================================================================
```

The first flow entry shown is the table-miss entry with an action of fall-through (or forward). The second entry contains the new flow entry.

The cookie associated with the message has a value of 0x0000000000000100 and, as shown in Table 2, because the high-order bits are set to zero, the cookie represents a flow entry that is used by filters within the base routing instance (shown as Global Routing Table or GRT). The filter used by this flow entry is _tmnx_ofs_ofs-1:8. This is the system-created OpenFlow IP filter for OpenFlow switch ofs-1 and contains all active GRT flow entries for that switch. Any filters that embed GRT OpenFlow entries from switch instance ofs-1 will automatically inherit all the active flow entries contained within this filter. In this example, the flow entries in _tmnx_ofs_ofs-1:8 are inherited only by IP filter 10. In addition, IP filter 10 will include ingress packets/bytes matched for each entry.

The priority field indicates a value of 1635 and, as previously described, determines the order with which flow entries are processed. Because OpenFlow states that the highest priority should be processed first and SR OS processes packets starting with the lowest numeric entry ID within a filter, the formula [65535 - flow_priority + embedding offset] is used to convert the cookie priority into a filter entry ID. This yields a filter entry ID of 63900. When a packet matches an entry in the filter, the packet is subject to the action defined in that entry, and is not subject to further filter entry processing.

The OpenFlow match fields specify an Ethertype of IPv4 (0x0800) and a destination prefix of 172.31.100.0/24, which are converted directly into filter entry match criteria. The OpenFlow instruction type is Write_Actions (OFPIT_WRITE_ACTIONS) in order to create the new flow, and has an action type of Output (OFPAT_OUTPUT). The output is directed to a (logical) port, which is the LSP PE-4-PE-5-RSVP.

The Modify Flow Entry (OFPT_FLOW_MOD) message contains a field for flags that are associated with each flow entry. These flags, together with some internal flags, are indicated in the Flow Flags field. Their meanings are described in Table 3.

Table 3: FLOW_MOD Flags

Flag	Meaning	Description
!E	Not evictable	Entry cannot be removed.
RO	Read only	Entry cannot be modified.
DEF	Default	
FR	SEND_FLOW_REM	If set, the switch must send a Flow-Removed message when the flow entry is deleted.
CO	CHECK_OVERLAP	If set, the switch must check that there are no conflicting entries with the same priority before inserting it into the flow entry table. An error is returned if a conflict exists.
RC	RESET_COUNTS	Reset flow packet and byte counts.
!PC	NO_PKT_COUNTS	When set, the switch does not need to keep track of the flow packet count.
!BC	NO_BYT_COUNTS	When set, the switch does not need to keep track of the byte count.

The dynamic OpenFlow flow entry redirecting traffic destined for prefix 172.31.100.0/24 is now in place as entry 63900 within IP filter _tmnx_ofs_ofs-1:8, and subsequently IP filter 10. The following two outputs show a monitor command run against PE-1's SAP toward test port B and PE-5's SAP toward test port C. Both SAPs are equally spreading the load of the two test streams of 1000 packets/s. Traffic for prefix 172.31.200.0/24 is routed toward PE-1 based on a route-table lookup. Traffic for prefix 172.31.100.0/24 is forwarded to PE-5 as a result of the OpenFlow redirect.

```
*A:PE-5# monitor service id 1 sap lag-1:10 rate
===============================================================================
Monitor statistics for Service 1 SAP lag-1:10
===============================================================================
---snip---
                        Packets                 Octets                  % Port
```

```
                                                                        Util.
Egress Queue 1
For. In/InplusProf    : 0                     0                       0.00
For. Out/ExcProf      : 1000                  512186                  0.04
Dro. In/InplusProf    : 0                     0                       0.00
Dro. Out/ExcProf      : 0                     0                       0.00
```

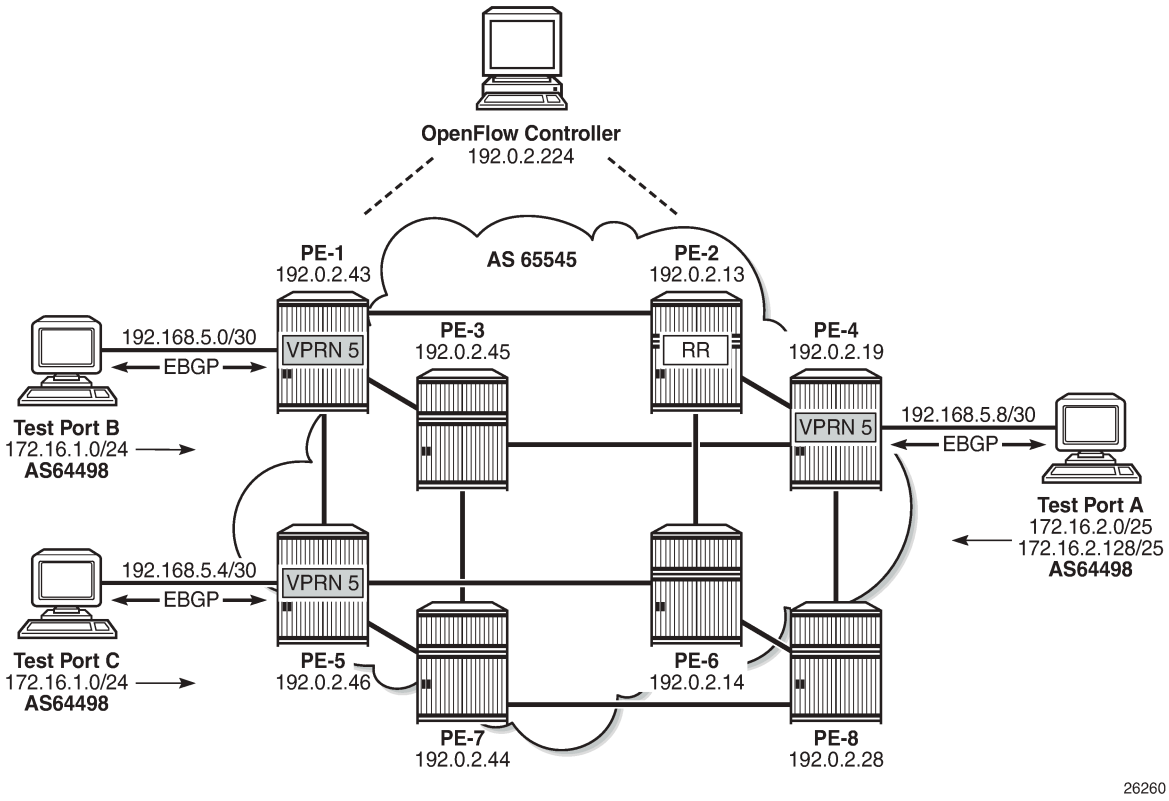
```
B:PE-1# monitor service id 1 sap 5/1/3:10 rate
===============================================================================
Monitor statistics for Service 1 SAP 5/1/3:10
===============================================================================
---snip---
                        Packets                 Octets                  % Port
                                                                       Util.
Egress Queue 1
For. In/InplusProf    : 0                     7                        ~0.00
For. Out/ExcProf      : 1000                  512186                   0.04
Dro. In/InplusProf    : 0                     0                        0.00
Dro. Out/ExcProf      : 0                     0                        0.00
```

FLOW_MOD messages allow for flow entries to be associated with hard and idle timeouts, which are not currently used by SR OS. Although timeout values can be passed by a controller in a FLOW_MOD message, they are effectively ignored. As a result, dynamic flow entries remain in place as filter entries until removed by the controller, or the OpenFlow switch instance is placed in a **shutdown** state. If the LSP transitions to an operationally down state while the redirect flow entry is still active, the switch will notify the controller of the change of state using a Port-Status message, and traffic will be subject to a forward action. If the LSP becomes operational again, the flow entry becomes active again.

Service Routing Instance

To generate rules within a VPRN routing instance, the example topology is configured as shown in Figure 24: Example Topology for OpenFlow within a Service Routing Context. Test ports A, B, and C belong to VPRN 5. Test ports B and C simulate CE routers in a dual-homed site, advertising prefix 172.16.1.0/24 in EBGP to PE-1 and PE-5, respectively. Test port A simulates a CE router at a different site, advertising prefixes 172.16.2.0/25 and 172.16.2.128/25 in EBGP to PE-4. The PE to CE (WAN) links are also advertised into VPN-IPv4 by the respective PE routers, to provide complete visibility of the VPN.

Figure 24: Example Topology for OpenFlow within a Service Routing Context



An IP filter is configured using the **embed-filter open-flow** command to allow for dynamic embedding of flow entries by an OpenFlow switch instance. In this example, the OpenFlow switch remains as ofs-1. The command also specifies **service 5** to make this filter applicable to interfaces within that service instance. Thereafter, this filter can only be deployed in the configured service. An **offset** is also defined to specify the start point for dynamically created OpenFlow entries and allow them to remain isolated from other dynamic and static filter entries.

```
configure
    filter
        ip-filter 20 create
            description "OpenFlow Service Filter"
            embed-filter open-flow "ofs-1" service 5 offset 100
        exit
```

The **embed-filter** command has the option to configure a service ID or a SAP ID. The former is applicable to embedding filters applied in VPRN or VPLS services. The latter is applicable only to VPLS services. It requires that the embedding filter has the scope of exclusive (as opposed to the default scope of template) and can only be deployed on the SAP specified in the argument.

The filter is applied at PE-4 on the SAP connecting test port A, as follows:

```
configure
    service
        vprn 5 customer 1 create
            interface "Test-Port-A" create
```

```
                address 192.168.5.9/30
                sap 3/1/4:5 create
                    ingress
                        filter ip 20
                    exit
                exit
            exit
```

As with the example of the base routing context, before any flow entries are initiated from the controller, a single entry with ID 65535 (maximum) is automatically populated in the filter, representing the table-miss entry. As before, when OpenFlow uses filters, it ignores any **default-action** that may be configured in the filters, so that filters can be chained. However, a table-miss action must exist and this is represented by entry 65535, as follows:

```
B:PE-4# show filter ip 20
===============================================================================
IP Filter
===============================================================================
Filter Id           : 20                            Applied        : Yes
Scope               : Template                      Def. Action    : Drop
System filter       : Unchained
Radius Ins Pt       : n/a
CrCtl. Ins Pt       : n/a
RadSh. Ins Pt       : n/a
PccRl. Ins Pt       : n/a
Entries             : 0/0/0/1 (Fixed/Radius/Cc/Embedded)
Description         : OpenFlow Service Filter
-------------------------------------------------------------------------------
Filter Match Criteria : IP
-------------------------------------------------------------------------------
Entry               : 65535
Origin              : Inserted by open-flow (no-match-action)
Description         : (Not Specified)
Log Id              : n/a
Src. IP             : 0.0.0.0/0
Src. Port           : n/a
Dest. IP            : 0.0.0.0/0
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-syn             : Off                           TCP-ack        : Off
Option-pres         : Off
Egress PBR          : Disabled
Primary Action      : Forward
Ing. Matches        : 8193635 pkts (4194988287 bytes)
Egr. Matches        : 0 pkts

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

An OpenFlow IP filter, _tmnx_ofs_ofs-1:16, is also automatically created by the system and contains all of the flow entries dynamically created by the OpenFlow switch ofs-1 for service ID 5. This filter acts as a repository for active flow entries specific to that service context and its purpose has been previously described. If a new filter is configured to embed OpenFlow entries for service ID 5, the entries from _tmnx_ofs_ofs-1:16 will be cloned into that new filter.

```
B:PE-4# show filter ip filter-type openflow
===============================================================================
Openflow IP Filters                                          Total:     2
```

```
===============================================================================
Filter-Id                    Description
-------------------------------------------------------------------------------
_tmnx_ofs_ofs-1:15           Filter for OFS 'ofs-1' for grt context
_tmnx_ofs_ofs-1:16           Filter for OFS 'ofs-1' for service [5] context
===============================================================================
```

OpenFlow Filtering in Action

To validate flow entries initiated by the controller, the following traffic flows are sourced from test port A connected to PE-4:

- A UDP-based flow with a source address of 172.16.2.1/25 and a destination address of 172.16.1.1/24 at a rate of 1000 packets/s.
- A UDP-based flow with a source address of 172.16.2.129/25 and a destination address of 172.16.1.1/24, again at a rate of 1000 packets/s.

Both test port B and C are advertising the preceding prefixes, which are advertised internally in VPN-IPv4 by PE-1 and PE-5, respectively. At PE-4, the preferred next-hop for 172.16.1.0/24 within VPRN 5 is PE-1 (192.0.2.43), as follows:

```
B:PE-4# show router 5 route-table 172.16.1.0/24
===============================================================================
Route Table (Service: 5)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
172.16.1.0/24                                 Remote  BGP VPN   01h47m36s  170
       192.0.2.43 (tunneled:RSVP:9)                                 0
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

PE-1 is egressing traffic at a rate of 2000 packets/s toward test port B, representing the sum of the two 1000 packets/s test streams, as follows:

```
B:PE-1# monitor service id 1 sap 5/1/3:10 rate
===============================================================================
Monitor statistics for Service 1 SAP 5/1/3:10
===============================================================================
---snip---
-------------------------------------------------------------------------------
At time t = 22 sec (Mode: Rate)
-------------------------------------------------------------------------------
                        Packets                 Octets                  % Port
                                                                        Util.
Egress Queue 1
For. In/InplusProf    : 0                       7                       ~0.00
For. Out/ExcProf      : 2000                    1023907                 0.08
Dro. In/InplusProf    : 0                       0                       0.00
Dro. Out/ExcProf      : 0                       0                       0.00
```

An OFPT_FLOW_MOD message containing an OFPFC_ADD command is initiated by the controller and can be viewed using the command in the following output:

```
B:PE-4# tools dump open-flow of-switch "ofs-1"

===============================================================================
Switch: ofs-1
===============================================================================
Table     : 0                           Flow Pri  : 0
Cookie    : 0x0000000000000000          CookieType: grt
Controller: :::0
Filter Hnd: 0xC300000F0000FFFF
Filter    : _tmnx_ofs_ofs-1:15 entry 65535

In Port   : *
VID       : *                           Outer VID : *
EthType   : *
Src IP    : *
Dst IP    : *
IP Proto  : *                           DSCP      : *
Src Port  : *                           Dst Port  : *
ICMP Type : *                           ICMP Code : *
Label     : *
IPv6ExtHdr: (Not Specified)

Action    : Fall-through

Flow Flags: IPv4/6 [!E] [RO] [DEF]
Up Time   : 0d 05:01:44                 Add TS    : 422384764
Mod TS    : 0                           Stats TS  : 424195136
#Packets  : 27501425                    #Bytes    : 14080300394
-------------------------------------------------------------------------------
Table     : 0                           Flow Pri  : 1535
Cookie    : 0xC000000500000038          CookieType: service 5
Controller: 192.0.2.224:6653
Filter Hnd: 0x8300001000000FA00
Filter    : _tmnx_ofs_ofs-1:16 entry 64000

In Port   : *
VID       : *                           Outer VID : *
EthType   : 0x0800
Src IP    : 172.16.2.128/25
Dst IP    : *
IP Proto  : *                           DSCP      : *
Src Port  : *                           Dst Port  : *
ICMP Type : *                           ICMP Code : *
Label     : *

Action    : Forward On Nhop(Indirect)
            Nhop: 192.168.5.6

Flow Flags: IPv4 [FR]
Up Time   : 0d 00:00:44                 Add TS    : 424190707
Mod TS    : 0                           Stats TS  : 424195137
#Packets  : 44301                       #Bytes    : 22682112
-------------------------------------------------------------------------------
Number of flows: 2
===============================================================================
```

The first flow entry with cookie value 0x0000000000000000 is the table-miss entry with a fall-through or forward action. The second entry with cookie value 0xC000000500000038 contains the new flow entry. The high-order bits of the cookie are set to 0xC (or 1100), which (as shown in Table 2) means that this

represents a flow entry that is used by filters used within a service instance. Bits 59 to 32 encode the service instance, which in this case is 5.

The filter used by this second flow entry is _tmnx_ofs_ofs-1:16, which is the system-created OpenFlow filter for OpenFlow switch ofs-1, and contains all active flows entries initiated by that switch for service ID 5. Any filters embedding OpenFlow flow entries from ofs-1 in service ID 5 will clone all of the entries contained in _tmnx_ofs_ofs-1:16. In this example, the entries in _tmnx_ofs_ofs-1:16 are cloned into IP filter 20. IP filter 20 will also include ingress packets/bytes matched for each entry.

The priority field indicates a value of 1535 and, as previously described, determines the order in which flow entries are processed, using the formula [65535 - flow_priority + embedding offset].

The OpenFlow Match fields specify an Ethertype of IPv4 (0x0800) for source prefix 172.16.2.128/25, and are mapped directly into filter entry match criteria. The OpenFlow instruction type is Write_Actions (OFPIT_WRITE_ACTIONS) in order to create the new flow entry, and has an action type of Forward to Next-Hop IP Address. Because OpenFlow has no standard action type of Forward to Next-Hop IP Address, an Experimenter (OFPAT_EXPERIMENTER) is used for this purpose, which encompasses the use of both direct and indirect next-hops. In this example, an indirect next-hop of 192.168.5.6 is used.

The preferred next-hop for traffic destined to prefix 172.16.1.0/24 is PE-1. The indirect next-hop address of 192.168.5.6 represents the (simulated) CE WAN address of test port C, and is known in the routing table of VPRN 5 with a next-hop of PE-5 (192.0.2.46), as follows:

```
B:PE-4# show router 5 route-table 192.168.5.6
===============================================================================
Route Table (Service: 5)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
192.168.5.4/30                                Remote  BGP VPN   19h19m22s  170
       192.0.2.46 (tunneled:RSVP:10)                                0
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The dynamic OpenFlow flow entry redirecting traffic with a source address of 172.16.2.128/25 is now in place as entry 64000 within IP filter _tmnx_ofs_ofs-1:16, and cloned into IP filter 20. The effect of this OpenFlow flow entry on the two test streams is as follows:

- Traffic sourced from prefix 172.16.2.0/25 to prefix 172.16.1.0/24 is routed in accordance with the VPRN 5 routing table, with a next-hop of PE-1.
- Traffic sourced from prefix 172.16.2.128/25 to prefix 172.16.1.0/24 is subject to policy-based routing and, rather than being routed directly toward the destination prefix known via PE-1, is forwarded to an indirect next-hop of 192.168.5.6, known via PE-5.

This is validated in the following two outputs, which show a monitor command run against PE-1's SAP toward test port B and PE-5's SAP toward test port C. The outputs show that each SAP is egressing 1000 packets/s:

```
B:PE-1# monitor service id 5 sap 5/1/3:5 rate
===============================================================================
Monitor statistics for Service 5 SAP 5/1/3:5
===============================================================================
---snip---
```

```
                        Packets                 Octets                  % Port
                                                                        Util.
Egress Queue 1
For. In/InplusProf    : 0                       8                       ~0.00
For. Out/ExcProf      : 1000                    512000                  0.04
Dro. In/InplusProf    : 0                       0                       0.00
Dro. Out/ExcProf      : 0                       0                       0.00
```

```
A:PE-5# monitor service id 5 sap lag-1:5 rate
===============================================================================
Monitor statistics for Service 5 SAP lag-1:5
===============================================================================
---snip---
                        Packets                 Octets                  % Port
                                                                        Util.
Egress Queue 1
For. In/InplusProf    : 0                       0                       0.00
For. Out/ExcProf      : 1000                    512000                  0.04
Dro. In/InplusProf    : 0                       0                       0.00
Dro. Out/ExcProf      : 0                       0                       0.00
```

As previously described, dynamic flow entries will remain in place as filter entries until removed by the controller or the OpenFlow switch instance is put in a **shutdown** state.

Supported Redirect Actions

Table 4 lists the redirect actions supported in SR OS together with the applicability and associated action types. Experimenter encodings are described in user guides. Unless otherwise stated, all instruction types are WRITE_ACTION/APPLY_ACTION.

Table 4: Supported Redirect Actions

Action	Applicability	Action Type	Remarks
Redirect to IP Next-Hop	IPv4/IPv6 traffic ingressing an IP interface	OFPAT_EXPERIMENTER (ALU_AXN_REDIRECT_TO_NEXTHOP)	Next-hop can be direct or indirect
Redirect to Routing Context (GRT or VRF)		OFPAT_OUTPUT <logical_port> <logical_port> encoding: Bits 31-28=0100, bits 27-24=0001, bits 23-0=VPRN service ID or 0 for GRT	
Redirect to Next-Hop and VRF/GRT Routing Context		Action 1: OFPAT_EXPERIMENTER (ALU_AXN_REDIRECT_TO_NEXTHOP)	Next-hop must be indirect
		Action 2: OFPAT_OUTPUT <logical_port> <logical_port> encoding: Bits 31-28=0100, bits 27-24=0001,	

Action	Applicability	Action Type	Remarks
		bits 23-0=VPRN service ID or 0 for GRT	
Redirect to LSP		OFPAT_OUTPUT <logical_ port> <logical_port> encoding: Bits 31-28=0100, bits 27-24=0000, bits 23-0=RSVP-TE Tunnel ID	
Redirect to SAP	Traffic ingressing a VPLS interface	Action 1: OFPAT_OUTPUT <port> <port> encoding: OXM_OF_IN_ PORT: TmnxPortID for Ethernet port or LAG	TmnxPortId encoding in TIMETRA-CHASSIS-MIB (port) or LAG TIMETRA-TC-MIB (LAG)
		Action 2: OFPAT_SET_FIELD <vlan_encoding> VLAN encoding: OXM_OF_ VLAN_ID (null, dot1Q, or inner QinQ tag) Optional EXPERIMENTER OFL_OUT_VLAN_ID (outer QinQ tag)	
Redirect to SDP		OFPAT_EXPERIMENTER (ALU_AXN_REDIRECT_TO_ SDP)	Possible to match against entire SAP using OXM_OF_ IN_PORT encoding Tmnx PortID, OXM_OF_VLAN_ID (null tag, dot1Q tag, inner Q-in-Q tag) and optional EXPERIMENTER OFL_ OUT_VLAN_ID (outer Q-in-Q tag)

Resource Consumption

Dynamic OpenFlow flow entries are embedded in filters as filter entries, and as such, consume CAM entries in the same way as statically configured filter entries and/or other dynamic filter entries, such as those created by BGP FlowSpec or RADIUS. When a flow entry is created and dynamically embedded as a filter entry, it will consume one or more ingress ACL/QoS entries from the line card to which the filter is attached. If a flow entry is embedded in multiple filters, an ingress ACL/QoS entry will be consumed for each filter. If a flow entry is embedded in a single filter with a default scope of template, and this filter is attached to multiple SAPs on the same line card, only a single entry is consumed.

As with conventional ACL resource consumption, a standard four- or five-tuple match will consume a single entry. Defining a range of ports, for example, will consume multiple entries, as follows:

```
B:PE-4# tools dump system-resources 3
Resource Manager info at 049 d 12/01/16 09:10:18.148:
Hardware Resource Usage for Slot #3, CardType imm12-10gb-sf+, Cmplx #0:
                                |   Total   | Allocated |    Free
  ------------------------------|-----------|-----------|------------
---snip---
         Ingress ACL/QoS Entries |      65536|          5|      65531
---snip---
```

Debugging

A number of OpenFlow debug commands are available. For troubleshooting and interoperability purposes, detailed packet-level debug commands are available for all OpenFlow message types. Also, the ability to debug OpenFlow switch errors is useful. An example is provided in the following output:

```
debug
    open-flow
        of-switch "ofs-1"
            error
            packet flow-mod detail
        exit
    exit
exit
```

Conclusion

OpenFlow has a number of use-cases in the WAN. The dynamic insertion of flow entries from a controller can be used for flow placement in an SDN environment implementing some business logic. Equally, it could be used to implement security measures, or off-ramping of traffic to a DDoS scrubbing center.

This chapter described how to configure and deploy Hybrid OpenFlow in SR OS. It described how to configure the OpenFlow switch, and how filter entries are dynamically embedded in GRT filters and service filters. These examples are intended to provide an overview of functionality.

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 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 *7450 ESS, 7750 SR, and 7950 XRS MPLS Advanced Configuration Guide for Classic CLI*; for SRLGs, see chapter "Shared Risk Link Groups for RSVP-Based LSPs" in *7450 ESS, 7750 SR, and 7950 XRS MPLS Advanced Configuration Guide for Classic CLI*.

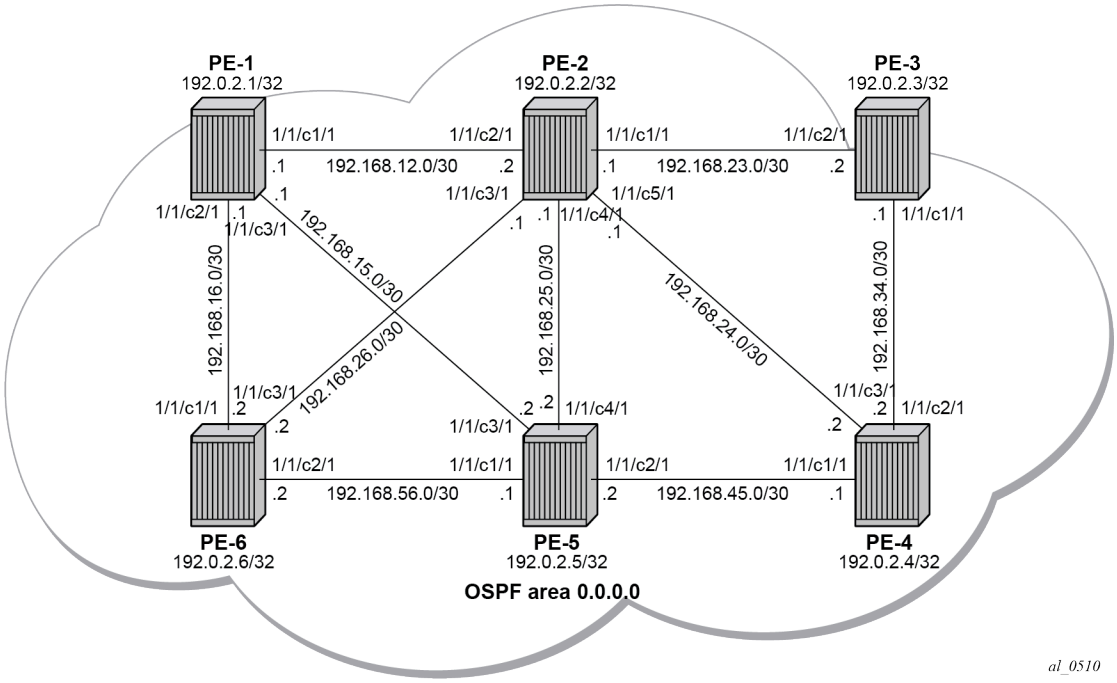
In the SR OS implementation, IP admin groups and SRLGs are locally significant, meaning they are not advertised by the IGP. Only the admin groups and SRLGs bound to an MPLS interface are advertised in TE link TLVs and sub-TLVs when the traffic engineering option is enabled in the IGP protocol. IES and 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 25: 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 25: 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 other PEs, as follows. The LDP LSP metric follows the IGP cost.

```
*A: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:
configure
    router Base
        ospf 0
            loopfree-alternates

*A: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:
```

```
configure
    router Base
       ldp
            fast-reroute

*A: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: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: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 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
       router Base
           route-next-hop-policy
               template <template name>
   ```

   Commands within a route next hop policy template follow the **begin-abort-commit** model. After a **commit**, 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
       router Base
           if-attribute
               admin-group <group-name> value <group-value>
   ```

   Second, configure the admin group membership of the IP interfaces (network, IES, or VPRN), as follows. Maximum five admin groups can be assigned to an IP interface in one command but the command can be applied multiple times. 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> [ <group-name> ... (up to 5 max)]

configure
    service
        vprn <svc-id>
            interface <itf-name>
                if-attribute
                    admin-group <group-name> [ <group-name> ... (up to 5 max)]

configure
    service
        ies <svc-id>
            interface <itf-name>
                if-attribute
                    admin-group <group-name> [ <group-name> ... (up to 5 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 **pref** 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
    router Base
        route-next-hop-policy
            template <template-name>
                begin
                exclude-group <ip-admin-group-name>
                include-group <ip-admin-group-name> [pref <preference>]
                commit
```

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
    router Base
        if-attribute
```

```
            srlg-group <group-name> value <group-value> [penalty-weight <penalty-weight>]
```

Second, configure the SRLG group membership of the IP interfaces (network, IES, or VPRN), as follows. Up to five SRLG groups can be applied to an IP interface in one 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> [ <group-name> ... (up to 5 max)]

configure
    service
        vprn <svc-id>
            interface <itf-name>
                if-attribute
                    srlg-group <group-name> [ <group-name> ... (up to 5 max)]

configure
    service
        ies <svc-id>
            interface <itf-name>
                if-attribute
                    srlg-group <group-name> [ <group-name> ... (up to 5 max)]
```

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
    router Base
        route-next-hop-policy
            begin
            template <template-name>
                srlg-enable
            exit
            commit
```

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
       router Base
           route-next-hop-policy
               begin
               template <template-name>
                   protection-type {link|node}
               exit
               commit
   ```

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
    router Base
        route-next-hop-policy
            begin
            template <template-name>
                nh-type {ip|tunnel}
            exit
            commit
```

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>] [<router-id>]
            area <area-id>
                interface <itf-name>
                    lfa-policy-map route-nh-template <template-name>

configure
    router Base
        ospf3 [<ospf-instance>] [<router-id>
            area <area-id>
                interface <itf-name>
                    lfa-policy-map route-nh-template <template-name>

configure
    service
        vprn <svc-id>
            ospf [<router-id>]
                area <area-id>
                    interface <itf-name>
                        lfa-policy-map route-nh-template <template-name>

configure
    service
        vprn <svc-id>
            ospf3 [<router-id>] [<ospf-instance>]
                area <area-id>
                    interface <itf-name>
                        lfa-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 25: 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: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: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
    router Base
        if-attribute
            admin-group "red" value 1
        exit
        interface "int-PE-2-PE-1"
            if-attribute
                admin-group "red"
            exit
        exit
        interface "int-PE-2-PE-6"
            if-attribute
                admin-group "red"
            exit
        exit
```

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

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

```
configure
    router Base
        route-next-hop-policy
            begin
            template "LFA_NH_exclRed"
                exclude-group "red"
            exit
            commit
```

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"
                    lfa-policy-map route-nh-template "LFA_NH_exclRed"
                exit
                interface "int-PE-2-PE-6"
                    lfa-policy-map route-nh-template "LFA_NH_exclRed"
                exit
```

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: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: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
    router Base
        if-attribute
            srlg-group "blue" value 2
        exit
        interface "int-PE-2-PE-1"
            if-attribute
                srlg-group "blue"
            exit
        exit
        interface "int-PE-2-PE-6"
            if-attribute
                srlg-group "blue"
            exit
        exit
```

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

```
# on PE-2:
configure
    router Base
        route-next-hop-policy
            begin
            template "LFA_NH_SRLG"
                srlg-enable
            exit
            commit
```

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"
                    lfa-policy-map route-nh-template "LFA_NH_SRLG"
                exit
                interface "int-PE-2-PE-6"
                    lfa-policy-map route-nh-template "LFA_NH_SRLG"
                exit
```

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: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: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"
                    no lfa-policy-map
                exit
                interface "int-PE-2-PE-6"
                    no lfa-policy-map
                exit
```

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
    router Base
        ip-fast-reroute
```

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"
            exit
            path "path-PE-2-PE-4-PE-5-PE-6"
                hop 10 192.168.24.2 strict
                hop 20 192.168.45.2 strict
                hop 30 192.168.56.2 strict
```

```
            no shutdown
        exit
        lsp "LSP-PE-2-PE-6-strict"
            to 192.0.2.6
            primary "path-PE-2-PE-4-PE-5-PE-6"
            exit
            no shutdown
        exit
        no shutdown
```

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
                tunnel-next-hop
                    family ipv4
                        resolution filter
                        resolution-filter
                            rsvp
                        exit
                    exit
                exit
                no shutdown
            exit
        exit
        mpls
            lsp "LSP-PE-2-PE-6-strict"
                igp-shortcut lfa-only
            exit
        exit
```

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: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: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: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
    router Base
        route-next-hop-policy
            begin
            template "LFA_NH_Tunnel"
                nh-type tunnel
            exit
            commit
```

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"
                    lfa-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: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: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
    router Base
        policy-options
            begin
            prefix-list "lo0-PE-1"
                prefix 192.0.2.1/32 exact
            exit
            policy-statement "LFA_Exclude_PE-1"
                entry 10
                    from
                        prefix-list "lo0-PE-1"
                    exit
                    action accept
                    exit
                exit
            exit
```

```
            commit
```

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

```
# on PE-2:
configure
    router Base
        ospf 0
            loopfree-alternates
                exclude
                    prefix-policy "LFA_Exclude_PE-1"
                exit
```

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: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 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

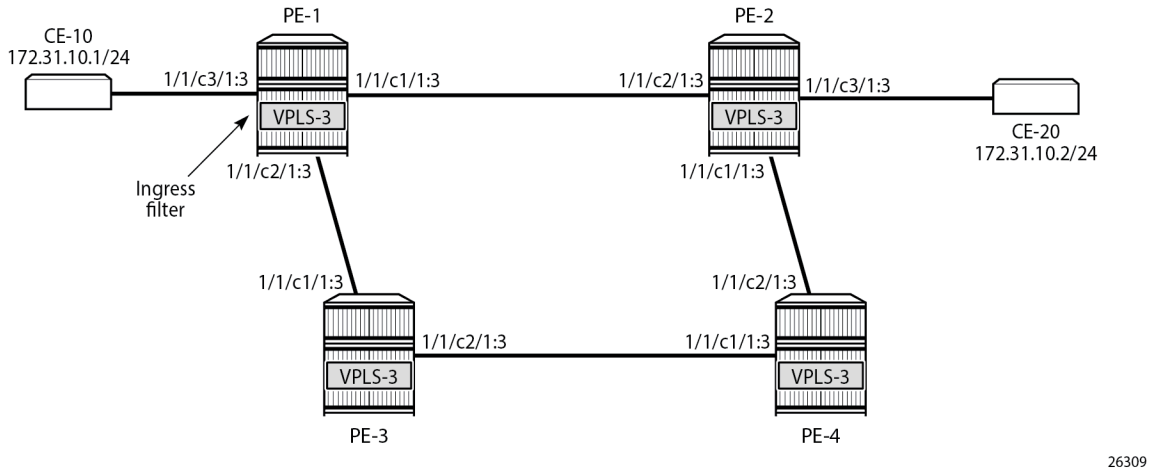
In most situations, PBR/PBF works on inbound unicast packets; therefore, a filter is applied at the ingress of access or network interfaces. In this chapter, examples will be shown for IPv4 filters and MAC filters applied on SAP ingress. IPv6 filters are also supported, but the examples in this chapter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

```
*A:PE-1>config>filter>ip-filter# entry 10 ?
  - entry <entry-id> [create]
  - no entry <entry-id>

 <entry-id>           : [1..2097151]
 <create>             : keyword - mandatory while creating an entry.


 [no] action          + Configure action for the filter entry
 [no] description     - Description for this filter entry
 [no] egress-pbr      - Enable egress PBR
 [no] filter-sample   - Enable/Disable Cflowd sampling
 [no] interface-disa* - Disable/Enable Cflowd sampling on the interfaces
 [no] log             - Configure log for the filter entry
 [no] match           + Configure match criteria for this ip filter entry
 [no] pbr-down-actio* - Configure action that overrides default PBR/PBF down action.
                        'no pbr-down-action-override' preserves default PBR/PBF down action,
                        which varies for different actions.
 [no] sample-profile  - Cflowd sample profile which will be used for packets matching this
                        filter entry
 [no] sticky-dest     - Set stickiness of PBR/PBF destinations and hold-time-up for stickiness
                        to take effect
```

Figure 26: 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 26: 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 1 name "IP-1" create
            entry 10 create
                match
```

```
                    dst-ip 172.31.10.2/24
                    src-ip 172.31.10.1/24
                exit
                action
                    forward sap 1/1/c1/1:3
                exit
            exit
```

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

```
*A:PE-1>config>filter>mac-filter>entry$ ?
 [no] action          + Configure action for the filter entry
 [no] description     - Description for this filter entry
 [no] log             - Configure log for the filter entry
 [no] match           + Configure match criteria for this mac filter entry
 [no] pbr-down-actio* - Configure action that overrides default PBR/PBF down action.
                        'no pbr-down-action-override' preserves default PBR/PBF down action,
                        which varies for different actions.
 [no] sticky-dest     - Set stickiness of PBF destinations and hold-time-up for stickiness
                        to take effect
```

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 2 name "MAC-2" create
            entry 10 create
                match frame-type 802dot3
                    src-mac 00:00:5e:00:53:01 ff:ff:ff:ff:ff:ff
                exit
                action
                    forward sap 1/1/c1/1:3
                exit
            exit
        exit
```

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, 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 1 name "IP-1" create
            entry 10 create
                match
                    dst-ip 172.31.10.2/24
                    src-ip 172.31.10.1/24
                exit
                action
                    forward sap 1/1/c1/1:3
                exit
                action secondary
```

```
                forward sap 1/1/c2/1:3
            exit
        exit
```

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: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:

```
*A:PE-1>config>filter>ip-filter>entry# sticky-dest ?
  - no sticky-dest
  - sticky-dest <hold-time-up>
  - sticky-dest no-hold-time-up

 <hold-time-up>       : 0..65535 seconds
```

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.

```
*A:PE-1>config>filter>ip-filter>entry# pbr-down-action-override ?
  - no pbr-down-action-override
  - pbr-down-action-override <filter-action>

 <filter-action>      : drop|forward|filter-default-action
```

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 5: Primary and secondary forwarding actions shows the allowed primary and secondary forwarding action combinations within a filter entry.

Table 5: 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>

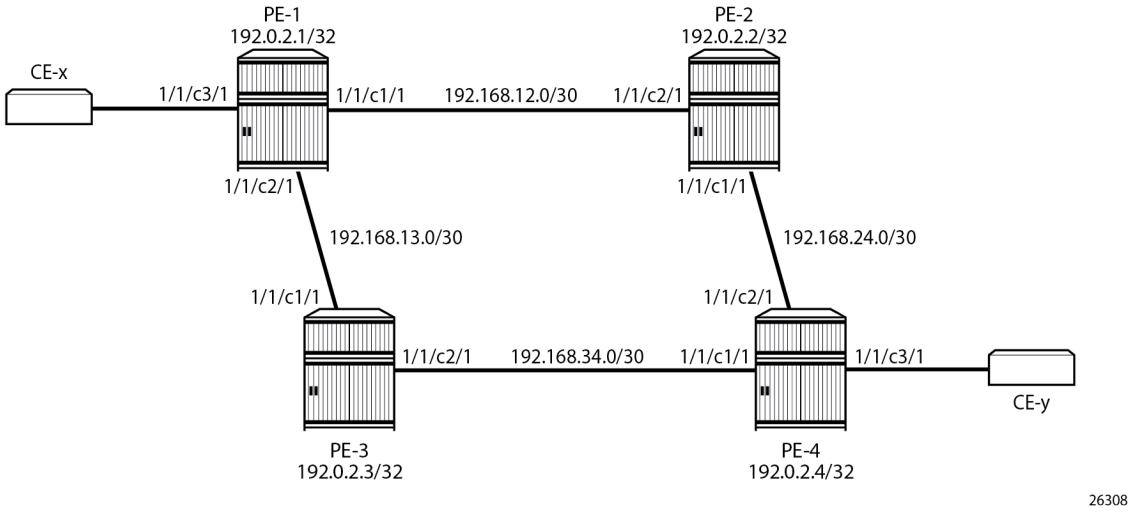
Configuration

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

Figure 27: Example topology



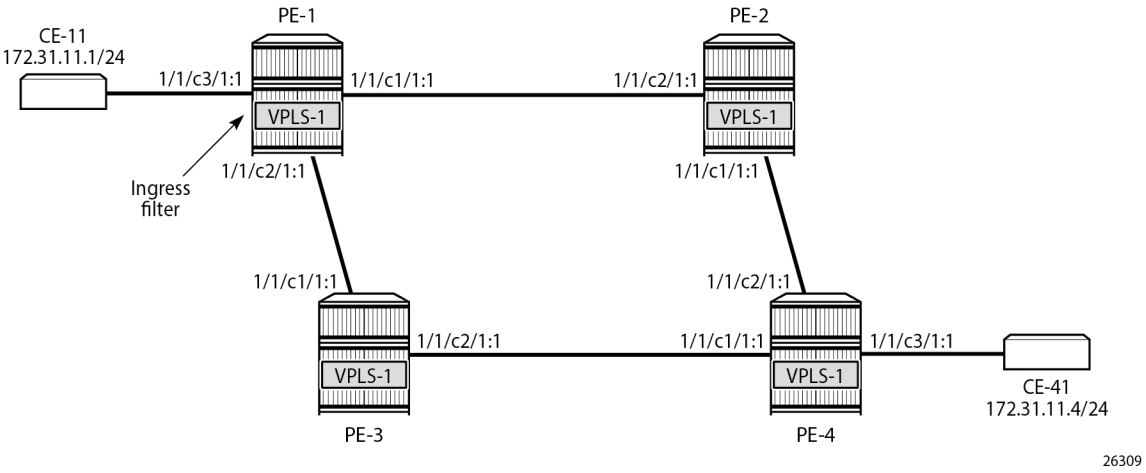
The initial configuration is as follows:

- 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 28: 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 28: 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 put in a shutdown state. 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 put in a shutdown state. The default action is drop.
4. Both primary and secondary action SAPs 1/1/c1/1:1 and 1/1/c2/1:1 are put in a shutdown state. The **pbr-down-action-override** parameter is configured with action *forward*.
5. The secondary action SAP 1/1/c2/1:1 is put in a no shutdown state. The secondary action is executed.
6. The primary action SAP 1/1/c1/1:1 is put in a no shutdown state. 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.

```
configure
    filter
        ip-filter 10 name "IP-10" create
            entry 10 create
                action
                    forward sap 1/1/c1/1:1
                exit
                action secondary
                    forward sap 1/1/c2/1:1
                exit
```

```
            exit
```

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 mpls create
            far-end 192.0.2.2
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        sdp 13 mpls create
            far-end 192.0.2.3
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        vpls 1 name "VPLS-1" customer 1 create
            stp
                shutdown
            exit
            sap 1/1/c1/1:1 create
                no shutdown
            exit
            sap 1/1/c2/1:1 create
                no shutdown
            exit
            sap 1/1/c3/1:1 create
                ingress
                    filter ip 10
                exit
                no shutdown
            exit
            spoke-sdp 12:1 create
                no shutdown
            exit
            spoke-sdp 13:1 create
                no shutdown
            exit
            no shutdown
        exit
```

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 200 ICMP echo requests and, afterward, the port statistics on PE-1 are verified.

```
*A:PE-1# clear port 1/1/c[1..3]/1 statistics
```

```
*A:PE-1# ping router 11 172.31.11.4 source 172.31.11.1 rapid count 200
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 = 2.46ms, avg = 2.82ms, max = 6.40ms, stddev = 0.350ms
```

```
*A:PE-1# show port 1/1/c[1..3]/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                 Ingress Octets
Id                           Egress Packets                  Egress Octets
-------------------------------------------------------------------------------
1/1/c1/1                                 203                          21490
                                         202                          21366
===============================================================================

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                 Ingress Octets
Id                           Egress Packets                  Egress Octets
-------------------------------------------------------------------------------
1/1/c2/1                                   4                            414
                                           5                            528
===============================================================================

===============================================================================
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: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        : 200 pkts (21200 bytes)
Egr. Matches        : 0 pkts

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

Primary action PBR target down

The primary action SAP 1/1/c1/1:1 is put in a shutdown state. 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
                shutdown

*A: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        : 400 pkts (42400 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
```

```
                shutdown
```

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: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        : 400 pkts (42400 bytes)
Egr. Matches        : 0 pkts

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

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

```
*A:PE-1# ping router 11 172.31.11.4 source 172.31.11.1 rapid count 50
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 in a shutdown state. 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:PE-1# ping router 11 172.31.11.4 source 172.31.11.1 rapid count 200
PING 172.31.11.4 56 data bytes
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
---snip---
---- 172.31.11.4 PING Statistics ----
200 packets transmitted, 200 packets received, 0.00% packet loss
round-trip min = 2.14ms, avg = 2.71ms, max = 4.40ms, stddev = 0.261ms
```

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: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: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        : 850 pkts (90100 bytes)
Egr. Matches        : 0 pkts

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

Secondary action up - revertive behavior

The primary action SAP 1/1/c1/1:1 remains in a shutdown state, 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
                no shutdown
```

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

```
*A: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        : 1050 pkts (111300 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
                no shutdown
```

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: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        : 1250 pkts (132500 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 put in a shutdown state, so the secondary action is active, as follows. The hold remain timer keeps counting down.

```
# on PE-1:
configure
    service
        vpls "VPLS-1"
            sap 1/1/c1/1:1
                shutdown
```

```
*A: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    : 100
Ing. Matches        : 1450 pkts (153700 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
                no shutdown


*A: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    : 54
Ing. Matches        : 1650 pkts (174900 bytes)
Egr. Matches        : 0 pkts
```

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

In the preceding output, the secondary action is active and the hold remain time is 54 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: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        : 1650 pkts (174900 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
    filter
        ip-filter "IP-10"
            entry 10
                sticky-dest no-hold-time-up        # sticky-dest 0
```

```
*A:PE-1>config>filter# info
----------------------------------------------
        ip-filter 10 name "IP-10" create
            entry 10 create
                action
```

```
                forward sap 1/1/c1/1:1
            exit
            action secondary
                forward sap 1/1/c2/1:1
            exit
            pbr-down-action-override forward
            sticky-dest 0
        exit
```

Initially, the primary action is executed, but when the primary action SAP 1/1/c1/1:1 is put in a shutdown state, the secondary action is executed, as follows:

```
# on PE-1:
configure
    service
        vpls "VPLS-1"
            sap 1/1/c1/1:1
                shutdown
```

```
*A: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        : 1850 pkts (196100 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
                no shutdown


*A: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        : 2050 pkts (217300 bytes)
Egr. Matches        : 0 pkts

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

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

```
*A: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: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        : 2250 pkts (238500 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
    filter
        mac-filter 20 name "MAC-20" create
            entry 10 create
                match
                    src-mac 00:00:5e:00:53:11 ff:ff:ff:ff:ff:ff
                exit
                action
                    forward sap 1/1/c1/1:1
                exit
                action secondary
                    forward sap 1/1/c2/1:1
                exit
                pbr-down-action-override forward
                sticky-dest 0
            exit
        exit
```

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:

```
*A:PE-1>config>service>vpls>sap>ingress# filter mac 20
MINOR: SVCMGR #1631 There is another filter already defined for the SAP
```

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
                    no filter            # remove filter
                    filter mac 20
```

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

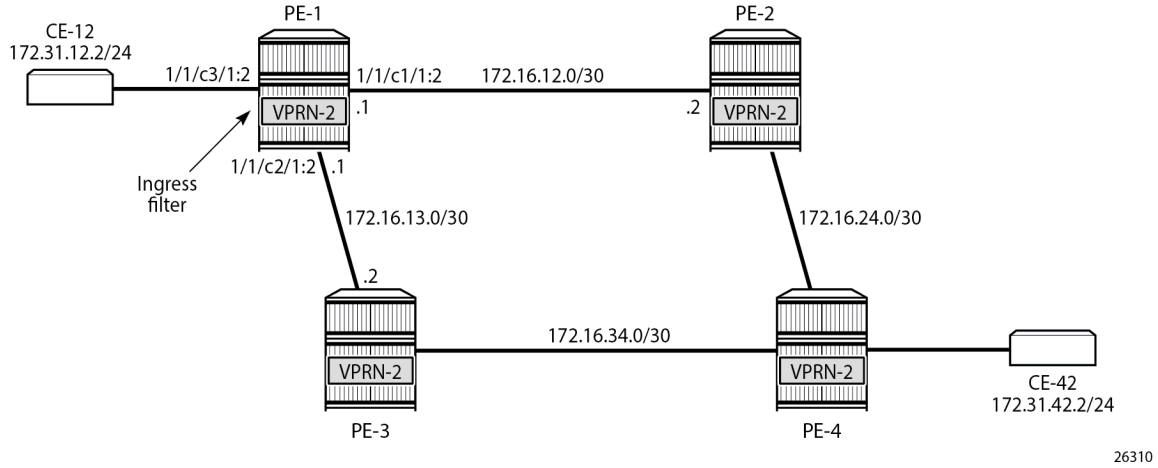
```
*A:PE-1# show filter mac "MAC-20"

===============================================================================
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 29: PBR in a VPRN shows the example topology used with the "VPRN-2" service configured on each PE and the CEs configured as VPRN 12 on PE-1 and PE-4.

Figure 29: PBR in a VPRN



The following IPv4 filter is configured on PE-1:

```
configure
    filter
        ip-filter 30 name "IP-30" create
            entry 10 create
                action
                    forward next-hop 172.16.12.2 router 2
                exit
                action secondary
                    forward next-hop 172.16.13.2 router 2
                exit
            exit
```

```
        exit
```

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 2 name "VPRN-2" customer 1 create
            interface "int-VPRN-2-PE-1-CE-12" create
                address 172.31.12.1/30
                sap 1/1/c3/1:2 create
                    ingress
                        filter ip 30
                    exit
                exit
            exit
            interface "int-VPRN-2-PE-1-PE-2" create
                address 172.16.12.1/30
                sap 1/1/c1/1:2 create
                exit
            exit
            interface "int-VPRN-2-PE-1-PE-3" create
                address 172.16.13.1/30
                sap 1/1/c2/1:2 create
                exit
            exit
            bgp-ipvpn
                mpls
                    route-distinguisher 64496:2
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

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: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
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        : 200 pkts (21200 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:

```
*A:PE-1>config>filter>ip-filter>entry# action extended-action ?
  - extended-action
  - no extended-action

      remark          - Activate dscp remarking for packets matching the entry
```

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
                shutdown
```

```
*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        : 200 pkts (21200 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 40 name "IP-40" create
            entry 10 create
                action
                    forward next-hop 192.0.2.2 router "Base"
                exit
                action secondary
                    forward next-hop 192.0.2.3 router "Base"
                exit
            exit
```

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:

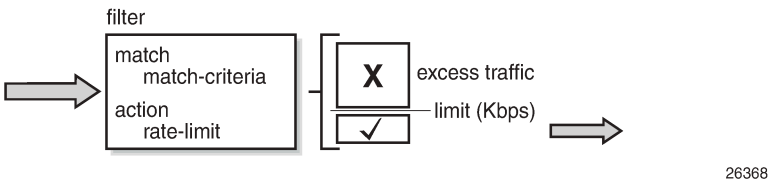
Applicability

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 30: Filter Based Rate Limiting.

Figure 30: 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.

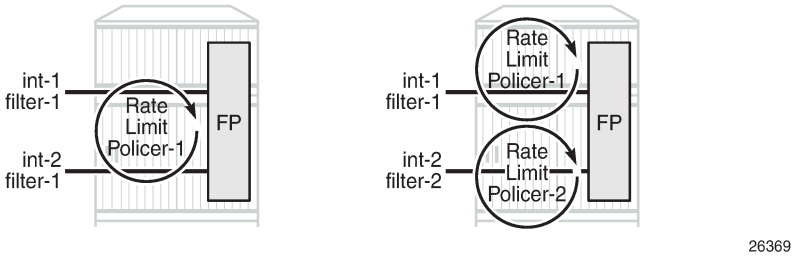
If multiple interfaces use the same rate-limiting filter policy on the same FP, the system will allocate a single rate-limiter resource to the FP; a common aggregate rate limit is applied to those interfaces.

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 31: 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 31: 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:
*A:PE-1>config>filter#         info
------------------------------------------------
        ip-filter|ipv6-filter|mac-filter <filter-id> name <filter-name> create
```

```
            default-action forward|drop
            description "<filter-description>"
            entry <entry-id> create
                match
                    ** match criteria, e.g.: IP/Port/MAC **
                exit
                action
                    rate-limit <value-Kbps>|max
                exit
            exit
        exit
```

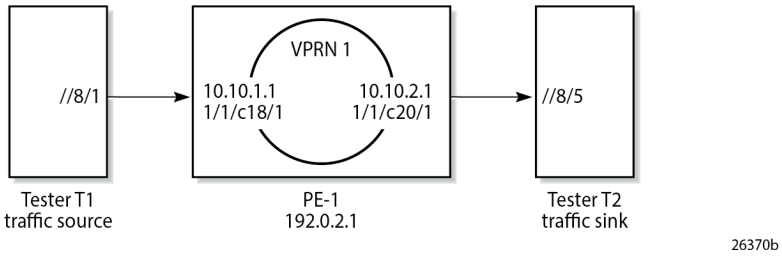
All regular IP and MAC match criteria are supported with the **action rate-limit**.

Configuration

Figure 32: 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 32: Example Configuration



The configuration of VPRN 1 on PE-1 is as follows:

```
# on PE-1:
configure
    service
        vprn 1 name "VPRN 1" customer 1 create
            description "rate limit action for ip filter"
            interface "int-TST-source" create
                address 10.10.1.1/24
                sap 1/1/c18/1 create
                ingress
                    filter ip 3
                exit
                no shutdown
                exit
            exit
            interface "int-TST-receiver" create
                address 10.10.2.1/24
                sap 1/1/c20/1 create
                exit
            exit
            bgp-ipvpn
                mpls
                    route-distinguisher 65536:1
```

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

The filter configuration is as follows:

```
# on PE-1:
configure
    filter
        ip-filter 3 name "ip-filter-1M" create
            default-action forward
            description "IP filter test for rate limit action"
            entry 10 create
                match
                    dst-ip 10.10.2.2/32
                    src-ip 10.10.1.2/32
                exit
                action
                    rate-limit 1024    # in Kbps ; 1024000/8/128 = 1000 packets/s
                exit
            exit
        exit
```

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:PE-1# monitor filter ip 3 entry 10 rate repeat 6 interval 11

===============================================================================
Monitor statistics for IP filter 3 entry 10
===============================================================================
-------------------------------------------------------------------------------
At time t = 0 sec (Base Statistics)
-------------------------------------------------------------------------------
Ing. Matches        : 1 pkts (128 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 (63977 bytes)
Egr. Matches        : 0 pkts
Ing. Rate-limiter
  Offered           : 500 pkts (63988 bytes)
  Forwarded         : 500 pkts (63988 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:PE-1# monitor filter ip 3 entry 10 rate repeat 6 interval 11

===============================================================================
Monitor statistics for IP filter 3 entry 10
===============================================================================
-------------------------------------------------------------------------------
At time t = 0 sec (Base Statistics)
-------------------------------------------------------------------------------
Ing. Matches        : 3 pkts (384 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        : 1499 pkts (191907 bytes)
Egr. Matches        : 0 pkts
Ing. Rate-limiter
  Offered           : 1500 pkts (191942 bytes)
  Forwarded         : 996 pkts (127465 bytes)
  Dropped           : 504 pkts (64477 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:PE-1# monitor service id 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)
-------------------------------------------------------------------------------

-------------------------------------------------------------------------------
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 16:50:30

                        Packets                 Octets                  % Port
                                                                        Util.
CPM Ingress           : 0                       0                       0.00

Forwarding Engine Stats
Dropped               : 0                       0                       0.00
Received Valid        : 455                     58182                   ~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:PE-1# monitor service id 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                     127454
Aggregate Dropped     : 0                       0
-------------------------------------------------------------------------------
Sap Statistics
-------------------------------------------------------------------------------
Last Cleared Time     : 04/12/2024 16:56:05

                        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                     127454                  ~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                     127454                  ~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-id> counters**, or **clear service statistics sap <sap-id> all** or **clear service statistics sap <sap-id> counters**, followed by: **show service id <service-id> sap <sap-id> base**, **show service id <service-id> sap <sap-id> stats** and **show service id <service-id> sap <sap-id> sap-stats** after the end of the counting period
- **clear filter ip|ipv6|mac <filter-id>**, followed by: **show filter ip|ipv6|mac <filter-id> 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* is recommended reading.

Overview

Equal Load Balancing

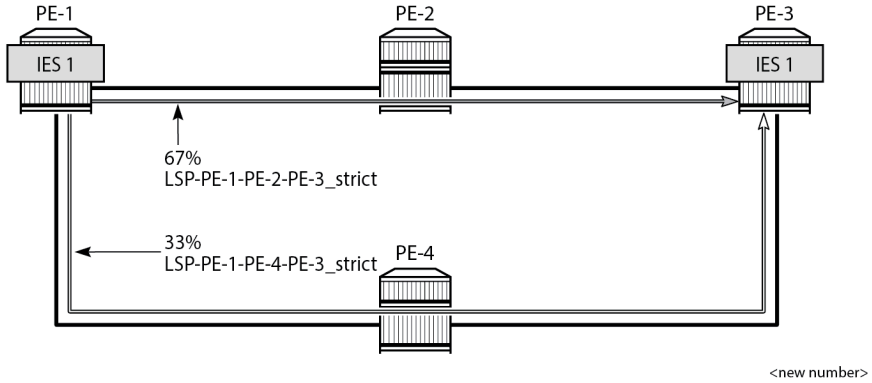
In this chapter, ECMP refers to spraying traffic flows over multiple RSVP-TE LSPs within an ECMP set. ECMP spraying consists of hashing the relevant fields in the packet header and selecting the tunnel next-hop based on the modulo operation of the output of the hash and the number of RSVP-TE LSPs present 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 33: 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 33: 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:

```
*A:PE-1# configure router Base mpls lsp-template "LSPtemplate1" load-balancing-weight ?
  - no load-balancing-weight
  - load-balancing-weight <weight>

 <weight>             : [0..4294967295] Default - 0
```

The following command is used to configure the weight on an LSP (for example on "LSP-PE-1-PE-2-PE-3_strict"):

```
*A:PE-1# configure router Base mpls lsp "LSP-PE-1-PE-2-PE-3_strict" load-balancing-weight ?
  - load-balancing-weight <weight>
  - no load-balancing-weight

 <weight>             : [0..4294967295] Default - 0
```

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"
                hop 10 192.168.12.2 strict
                hop 20 192.168.23.2 strict
                no shutdown
            exit
            lsp "LSP-PE-1-PE-2-PE-3_strict"
                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"
                exit
```

```
                no shutdown
            exit
```

Weighted ECMP for 6PE over RSVP-TE LSPs is enabled in the **bgp next-hop-resolution** context as follows:

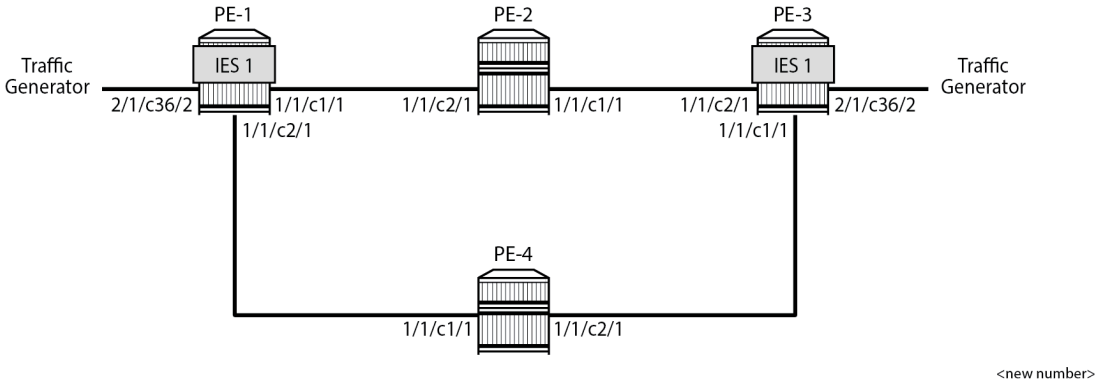
```
configure
    router Base
        bgp
            next-hop-resolution
                weighted-ecmp
```

The **weighted-ecmp** option controls load balancing to the same next-hop only.

Configuration

Figure 34: 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 34: 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"
            address 192.168.12.1/30
            port 1/1/c1/1
        exit
        interface "int-PE-1-PE-4"
            address 192.168.14.1/30
            port 1/1/c2/1
        exit
        interface "system"
            address 192.0.2.1/32
        exit
        isis 0
            area-id 49.0001
            traffic-engineering
            interface "system"
            exit
            interface "int-PE-1-PE-2"
                interface-type point-to-point
            exit
            interface "int-PE-1-PE-4"
                interface-type point-to-point
            exit
            no shutdown
        exit
        mpls
            interface "int-PE-1-PE-2"
            exit
            interface "int-PE-1-PE-4"
            exit
            path "path-PE-1-PE-2-PE-3_strict"
                hop 10 192.168.12.2 strict
                hop 20 192.168.23.2 strict
                no shutdown
            exit
            path "path-PE-1-PE-4-PE-3_strict"
                hop 10 192.168.14.2 strict
                hop 20 192.168.34.1 strict
                no shutdown
            exit
            lsp "LSP-PE-1-PE-2-PE-3_strict"
                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"
                exit
                no shutdown
            exit
            lsp "LSP-PE-1-PE-4-PE-3_strict"
                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"
                exit
                no shutdown
            exit
            no shutdown
        exit
        rsvp
```

```
            no shutdown
        exit
```

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 1 name "IES-1" customer 1 create
            description "6PE-1"
            interface "int-PE-1-STC" create
                ipv6
                    address 2001:db8::11:1/120
                exit
                sap 2/1/c36/2 create
                exit
            exit
            no shutdown
        exit
```

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
            split-horizon
            next-hop-resolution
                weighted-ecmp
                labeled-routes
                    transport-tunnel
                        family label-ipv6
                            resolution-filter
                                no ldp
                                rsvp
                            exit
                            resolution filter
                        exit
                    exit
                exit
            exit
            group "iBGP"
                export "export-6PE-1"
                peer-as 64496
                path-mtu-discovery
```

```
                neighbor 192.0.2.3
                    family label-ipv6
                exit
            exit
        exit
```

The configuration on PE-3 is similar.

On PE-1 and PE-3, the following export policy is configured:

```
configure
    router Base
        policy-options
            begin
            policy-statement "export-6PE-1"
                entry 10
                    from
                        protocol direct
                    exit
                    action accept
                    exit
                exit
                default-action drop
                exit
            exit
            commit
        exit
```

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: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 00h01m12s  170
       192.0.2.3 (tunneled:RSVP:3)                                  100
2001:db8::33:0/120 [2]                        Remote  BGP_LABEL 00h01m12s  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 3 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 3 is twice as much as the weight of RSVP tunnel 4, so the load on RSVP-TE LSP 3 is twice as high as the load on RSVP LSP 4.

```
*A: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                   : 00h01m12s
  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:3)
      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 3 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: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  3         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:PE-1# monitor port 1/1/c1/1 1/1/c2/1 2/1/c36/2 rate interval 3 repeat 3
```

```
===============================================================================
Monitor statistics for Ports
===============================================================================
                                                  Input                 Output
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
At time t = 6 sec (Mode: Rate)
-------------------------------------------------------------------------------
Port 1/1/c1/1
-------------------------------------------------------------------------------
Octets                                              21                 441717
Packets                                             0                    428
Errors                                              0                      0
Bits                                              168                3533736
Utilization (% of port capacity)                ~0.00                   0.03

Port 1/1/c2/1
-------------------------------------------------------------------------------
Octets                                              99                 187619
Packets                                             1                    182
Errors                                              0                      0
Bits                                              792                1500952
Utilization (% of port capacity)                ~0.00                   0.01

Port 2/1/c36/2
-------------------------------------------------------------------------------
Octets                                         623957                      0
Packets                                           609                      0
Errors                                              0                      0
Bits                                          4991656                      0
Utilization (% of port capacity)                 0.05                   0.00
---snip---
===============================================================================
```

This can also be verified as follows:

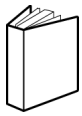
```
*A:PE-1# show port 1/1/c1/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                 Ingress Octets
Id                           Egress Packets                  Egress Octets
-------------------------------------------------------------------------------
1/1/c1/1                                 66                           7524
                                      11038                       11331501
===============================================================================
*A:PE-1# show port 1/1/c2/1 statistics

===============================================================================
Port Statistics on Slot 1
===============================================================================
Port                        Ingress Packets                 Ingress Octets
Id                           Egress Packets                  Egress Octets
-------------------------------------------------------------------------------
1/1/c2/1                                 64                           7556
                                      4710                        4805198
===============================================================================
*A: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                               15624                       15998976
                                            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