



7450 Ethernet Service Switch  
7750 Service Router  
7950 Extensible Routing System  
Virtualized Service Router  
Releases up to 24.10.R2

## MPLS Advanced Configuration Guide for MD CLI

---

3HE 20797 AAAC TQZZA  
Edition: 01  
March 2025

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.

---

This document includes Nokia proprietary and confidential information, which may not be distributed or disclosed to any third parties without the prior written consent of Nokia.

This document is intended for use by Nokia's customers ("You"/"Your") in connection with a product purchased or licensed from any company within Nokia Group of Companies. Use this document as agreed. You agree to notify Nokia of any errors you may find in this document; however, should you elect to use this document for any purpose(s) for which it is not intended, You understand and warrant that any determinations You may make or actions You may take will be based upon Your independent judgment and analysis of the content of this document.

Nokia reserves the right to make changes to this document without notice. At all times, the controlling version is the one available on Nokia's site.

No part of this document may be modified.

NO WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY OF AVAILABILITY, ACCURACY, RELIABILITY, TITLE, NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE, IS MADE IN RELATION TO THE CONTENT OF THIS DOCUMENT. IN NO EVENT WILL NOKIA BE LIABLE FOR ANY DAMAGES, INCLUDING BUT NOT LIMITED TO SPECIAL, DIRECT, INDIRECT, INCIDENTAL OR CONSEQUENTIAL OR ANY LOSSES, SUCH AS BUT NOT LIMITED TO LOSS OF PROFIT, REVENUE, BUSINESS INTERRUPTION, BUSINESS OPPORTUNITY OR DATA THAT MAY ARISE FROM THE USE OF THIS DOCUMENT OR THE INFORMATION IN IT, EVEN IN THE CASE OF ERRORS IN OR OMISSIONS FROM THIS DOCUMENT OR ITS CONTENT.

Copyright and trademark: Nokia is a registered trademark of Nokia Corporation. Other product names mentioned in this document may be trademarks of their respective owners.

© 2025 Nokia.

# Table of contents

List of tables.....	5
List of figures.....	6
Preface.....	13
Automatic Bandwidth Adjustment in P2P LSPs.....	14
Automatic Creation of RSVP-TE LSPs.....	52
BFD for RSVP-TE and LDP LSPs.....	69
BFD for RSVP-TE LSPs with Failure Action.....	88
DiffServ Traffic Engineering.....	102
Entropy Label.....	142
IGP Shortcuts.....	157
Inter-Area TE Point-to-Point LSPs.....	213
LDP FEC to BGP Label Route Stitching.....	237
LDP over RSVP Using OSPF as IGP.....	266
LDP Point-to-Point LSPs.....	307
LDP-IGP Synchronization.....	322
LDP-SR Stitching for IPv4 Prefixes (IS-IS).....	334
MPLS LDP FRR using ISIS as IGP.....	350
Multicast Label Distribution Protocol.....	375

---

<b>Path MTU Discovery.....</b>	<b>408</b>
<b>Remote Loop-Free Alternate Node Protection.....</b>	<b>430</b>
<b>RSVP Point-to-Point LSPs.....</b>	<b>447</b>
<b>RSVP Signaled Point-to-Multipoint LSPs.....</b>	<b>489</b>
<b>Seamless MPLS: Isolated IGP/LDP Domains and Labeled BGP.....</b>	<b>534</b>
<b>Shared Risk Link Groups for RSVP-Based LSPs.....</b>	<b>555</b>
<b>Static Point-to-Point LSPs.....</b>	<b>572</b>
<b>Topology-Independent Loop-Free Alternate for Link Protection.....</b>	<b>583</b>
<b>Tunneling of ICMP Reply Packets over MPLS LSPs.....</b>	<b>610</b>
<b>Unnumbered Interfaces in RSVP-TE and LDP.....</b>	<b>634</b>

---

# List of tables

Table 1: Comparison bandwidth constraint models.....	109
Table 2: RSVP LSP Role As Outcome of LSP Level and IGP Level Configuration Options.....	180
Table 3: Terminology.....	375
Table 4: MTU types.....	408
Table 5: MTU values for Ethernet frames.....	409
Table 6: Values of the max-sr-frr-labels parameter in TI-LFA.....	597

# List of figures

Figure 1: Auto-Bandwidth Adjustment Implementation.....	15
Figure 2: Underflow-Triggered Auto-Bandwidth Implementation.....	22
Figure 3: Example Setup for Auto-Bandwidth Point-to-Point LSPs.....	26
Figure 4: Example Topology.....	53
Figure 5: IGP Shortcuts with RSVP-TE Auto-Mesh.....	57
Figure 6: Example Topology for Single-Hop LDP over RSVP with ECMP.....	64
Figure 7: MPLS LSP BFD session establishment: BFD handshake.....	70
Figure 8: BFD for RSVP-TE LSPs - topology.....	71
Figure 9: BFD for LDP LSPs - topology.....	81
Figure 10: Topology for failure action failover.....	89
Figure 11: Topology for failure action down.....	96
Figure 12: Example topology.....	103
Figure 13: Mapping of TE classes.....	104
Figure 14: Bandwidth reservation for the CTs.....	105
Figure 15: Bandwidth reservation in Maximum Allocation Model for three CTs.....	110
Figure 16: Bandwidth reservation in Russian Doll Model for three CTs.....	110
Figure 17: Paths from PE-1 to PE-3.....	116
Figure 18: MAM bandwidth allocation.....	118
Figure 19: Reserved and unreserved bandwidth.....	121
Figure 20: Reserved and unreserved bandwidth on PE-1.....	123
Figure 21: Bandwidth reservation.....	125

---

Figure 22: Russian Doll Model for three class types.....	126
Figure 23: Reserved bandwidth for LSP with CT2 (one session).....	130
Figure 24: Bandwidth reservation for LSP with CT2 and LSP with CT1 (two sessions).....	131
Figure 25: Reserved bandwidth on both interfaces of PE-1 (three sessions).....	133
Figure 26: Reserved bandwidth on both interfaces on PE-1 (four sessions).....	134
Figure 27: Reserved bandwidth on both interfaces of PE-1 (five sessions).....	136
Figure 28: Reserved bandwidth on both interfaces on PE-1 (six sessions).....	139
Figure 29: Load-balancing of flows based on hash label or entropy label.....	143
Figure 30: Label stack with hash label versus label stack with EL and ELI.....	143
Figure 31: Downstream LERs signal EL capability to ILER.....	144
Figure 32: Example topology.....	145
Figure 33: RSVP LSP "LSP-PE-1-PE-2" from PE-1 to PE-2 via P-3 and P-4.....	146
Figure 34: Normal SPF Tree Sourced by PE-1.....	158
Figure 35: SPF Tree Sourced by PE-1 Using LSP Shortcuts.....	158
Figure 36: Example Topology.....	159
Figure 37: LSPs Between PE-1 and PE-6.....	173
Figure 38: RSVP Shortcuts LFA Use Case Example.....	186
Figure 39: Network Topology to Verify Installation of Shortcuts into the RTM.....	189
Figure 40: Shortcuts Within a VRF Topology Network.....	206
Figure 41: Inter-area TE LSP setup.....	214
Figure 42: Inter-area TE LSP path.....	215
Figure 43: ABR protection.....	224
Figure 44: Protection of all nodes/links along the LSP path.....	225

---

Figure 45: Admin group example.....	228
Figure 46: Share Risk Link Groups.....	232
Figure 47: LDP FEC to BGP label route stitching.....	238
Figure 48: Example topology.....	240
Figure 49: BGP enabled with P-4 as RR.....	243
Figure 50: End-to-end Epipe service.....	249
Figure 51: Label stacks for traffic from AN-1 to AN-8.....	256
Figure 52: Block BGP label bindings to LDP DU peer PE-9.....	261
Figure 53: Initial example topology.....	267
Figure 54: VPRN 1 with LDP over RSVP and no intra-area PE connectivity.....	283
Figure 55: VPRN 1 with LDP over RSVP and intra-area PE connectivity.....	303
Figure 56: Generic MPLS network, MPLS label operations.....	308
Figure 57: MPLS example topology.....	309
Figure 58: Example topology.....	326
Figure 59: Shortest path between PE-1 and PE-5.....	326
Figure 60: Rerouting via P-3 and P-4 until LDP synchronization timer terminates.....	329
Figure 61: Restored link with one LDP synchronization timer terminated.....	331
Figure 62: Example topology.....	335
Figure 63: Initial example topology.....	351
Figure 64: Data verification in the direction from PE-1 to PE-5 using Epipe service.....	362
Figure 65: LFA computation: Inequality 1 for prefix PE-5 (D) on PE-1 (S).....	368
Figure 66: LFA computation: Inequality 3 for prefix PE-5 (D) on PE-1 (S).....	369
Figure 67: IS-IS overload on PE-2, Inequality 1 for 192.168.24.0/30 (D) on PE-1 (S).....	374



---

Figure 68: Setup of mLDP P2MP LSP.....	377
Figure 69: Example topology.....	378
Figure 70: LDP P2MP LSP.....	384
Figure 71: New LDP P2MP LSP after metric change.....	402
Figure 72: L2 services MTUs for Ethernet frames.....	410
Figure 73: Minimum network port MTU for Ethernet frames in MPLS encapsulation.....	411
Figure 74: Path MTU.....	411
Figure 75: ICMP "Destination Unreachable" Message - Fragmentation Needed.....	413
Figure 76: ICMPv6 "Packet Too Big" message.....	413
Figure 77: Example topology.....	414
Figure 78: Multiple Epipes Using LDP SDPs.....	418
Figure 79: Multiple Epipes between PE-1 and PE-4 - IPv6.....	421
Figure 80: BGP-IPv4.....	423
Figure 81: BGP-IPv6.....	426
Figure 82: LFA node protection - topology & denominations.....	431
Figure 83: Node protecting extended P-space.....	432
Figure 84: Link protecting Q-space.....	432
Figure 85: One candidate PQ-router – repair tunnel.....	433
Figure 86: Two candidate PQ routers – repair tunnel.....	434
Figure 87: Example topology.....	435
Figure 88: Link protection extended P-space calculation.....	438
Figure 89: Link protecting Q-space calculation.....	438
Figure 90: Repair tunnel.....	439

---

Figure 91: Node protecting extended P-space calculation.....	442
Figure 92: Link protecting Q-space calculation.....	443
Figure 93: Validating candidate PQ routers - repair tunnel calculation.....	444
Figure 94: Generic MPLS network, MPLS label operations.....	448
Figure 95: MPLS example topology.....	450
Figure 96: LSP with dynamic path takes IGP best route.....	458
Figure 97: RSVP-TE LSP with dynamic path using TE metric.....	461
Figure 98: Fast reroute one-to-one detour tunnels.....	464
Figure 99: FRR facility bypass tunnels.....	467
Figure 100: FRR facility without node protection.....	471
Figure 101: Admin groups 'blue' and 'red'.....	473
Figure 102: LSP and bypass within admin group 'blue'.....	475
Figure 103: LSP and FRR bypass tunnel excluding admin group 'red'.....	477
Figure 104: P2MP example topology.....	490
Figure 105: P2MP LSP LSP-p2mp-1 with bypass tunnels.....	495
Figure 106: P2MP LSP p-to-mp-1 with metric change.....	514
Figure 107: P2MP LSP LSP-p2mp-1 with strict S2L path toward PE-7.....	517
Figure 108: Intelligent remerge, case 1.....	519
Figure 109: Intelligent remerge, case 2.....	524
Figure 110: Intelligent remerge, case 3.....	529
Figure 111: Seamless MPLS - network topology, control and data plane.....	535
Figure 112: Seamless MPLS - IGP/LDP domains.....	536
Figure 113: Seamless MPLS - BGP.....	538

---

Figure 114: End-to-End Epipe service.....	543
Figure 115: L3 VPN service.....	545
Figure 116: Label stacks for traffic from AN-1 to AN-5.....	548
Figure 117: Example topology.....	556
Figure 118: SRLG topology.....	557
Figure 119: Path primary RSVP-TE LSP.....	562
Figure 120: FRR bypass tunnels originating in PE-1 with and without SRLG.....	564
Figure 121: SRLG for secondary path.....	567
Figure 122: SRLG database example.....	570
Figure 123: Generic MPLS network, MPLS label operations.....	573
Figure 124: MPLS example topology.....	574
Figure 125: Static LSP running over PE-1, PE-2, PE-5, PE-6.....	576
Figure 126: Post-failure LFA path does not match post-convergence path.....	584
Figure 127: Post-failure TI-LFA path matches post-convergence path.....	586
Figure 128: Example topology.....	587
Figure 129: Example topology with regular LFA configured on PE-4.....	588
Figure 130: No post-failure LFA path when PE-4 loops back traffic.....	591
Figure 131: Example topology for remote LFA.....	593
Figure 132: PQ node in remote LFA.....	594
Figure 133: Extended P space of PE-1 and Q space of PE-4 are one hop apart.....	597
Figure 134: Directed LFA with P router and Q router one hop apart.....	599
Figure 135: Post-failure TI-LFA path coincides with post-convergence path.....	600
Figure 136: Use of TTL: uniform versus pipe.....	611

---

Figure 137: Use of TTL in an L2 VPN service in pipe mode.....	612
Figure 138: Use of TTL in an L3 VPN service in pipe mode.....	612
Figure 139: Tunneling of ICMP reply packets over an MPLS LSP.....	613
Figure 140: MPLS label stack object.....	618
Figure 141: ICMP extension header.....	619
Figure 142: ICMP extension object: object header and payload.....	619
Figure 143: Example configuration.....	620
Figure 144: Tunnel from iLER PE-3 to eLER PE-6 via LSR PE-2.....	624
Figure 145: UDP traceroute in VPRN with iLER in uniform mode.....	626
Figure 146: UDP traceroute in VPRN without TTL propagation to LDP.....	628
Figure 147: Tunnel from iLER PE-3 to eLER PE-6 with multiple LSRs.....	629
Figure 148: UDP traceroute with iLER in uniform mode.....	630
Figure 149: Example topology for unnumbered interfaces in RSVP and LDP.....	635
Figure 150: Configuration example for unnumbered Interfaces in RSVP and LDP.....	642
Figure 151: LSP-PE-4-PE-2 on unnumbered interfaces.....	646
Figure 152: LSP and FRR facility bypass tunnels.....	653
Figure 153: FRR one-to-one only supported on numbered interfaces.....	656
Figure 154: FRR on iLER: no bypass on unnumbered interfaces.....	658
Figure 155: FRR facility and admin groups.....	661
Figure 156: SRLG-FRR strict: no bypass on PE-4.....	664
Figure 157: LDP FRR LFA link protection on PE-4.....	678

# 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 24.10.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.

# Automatic Bandwidth Adjustment in P2P LSPs

This chapter provides information about automatic bandwidth adjustment in P2P LSPs.

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

## Applicability

Automatic bandwidth adjustment was first introduced in SR OS Release 8.0.R4. Overflow triggers are supported from SR OS Release 8.0.R4 onward; underflow triggers from SR OS Release 12.0.R1 onward. From SR OS Release 12.0.R4 onward, auto-bandwidth adjustment is also supported on LSPs that have secondary paths.

Initially, this chapter was written for SR OS Release 13.0.R2, but the CLI in the current edition corresponds to SR OS Release 23.10.R2.

## Overview

Automatic bandwidth adjustment refers to the capability of an ingress Label Edge Router (iLER) to dynamically adjust the bandwidth of a Resource Reservation Protocol (RSVP) Label Switched Path (LSP) tunnel based on active measurement of the traffic rate into the tunnel. The bandwidth assigned to an RSVP LSP tunnel is taken into account by the control plane, to verify that sufficient bandwidth is available for a new LSP or for an increase or decrease in bandwidth for an existing LSP. The actual bandwidth in the data plane is not capped by this setting. QoS mechanisms can be set up to filter and police the traffic in the data plane, but that is beyond the scope of this chapter.

Auto-bandwidth adjustment uses the existing LSP egress statistics feature to track the bandwidth on a specific LSP. When egress statistics are enabled, the Control Processing Module (CPM) collects statistics from all IOMs forwarding traffic belonging to the LSP (whether the traffic is currently leaving the ingress LER via the primary path, a secondary path, or an FRR detour/bypass path). The egress statistics have counts for the number of packets and bytes forwarded per LSP on a per-forwarding class, per priority (in-profile versus out-of-profile) basis.

For the actual bandwidth adjustment, Make-Before-Break (MBB) is used. No traffic interruption is noticed. If an auto-bandwidth attempt fails, there are 5 retries and, if they all fail, the bandwidth remains unchanged. The next attempt may occur with the next trigger.

Retries follow the retry-limit (5 in this case), retry-timer (by default 30s), and exponential back-off timer, if enabled in MPLS.

Auto-bandwidth adjustment can be triggered in four different ways:

1. Periodic trigger

The iLER determines at the end of each adjust-interval whether to attempt an auto-bandwidth adjustment.

2. Overflow or underflow trigger

The measured bandwidth of an LSP has increased or decreased significantly since the start of the current adjust-interval. It may be preferable to adjust the bandwidth of the LSP after a number of overflow/underflow samples, rather than wait for the adjust-interval to end (default: 24 h).

3. Manual trigger

An operator launches a **tools** command to trigger an auto-bandwidth adjustment.

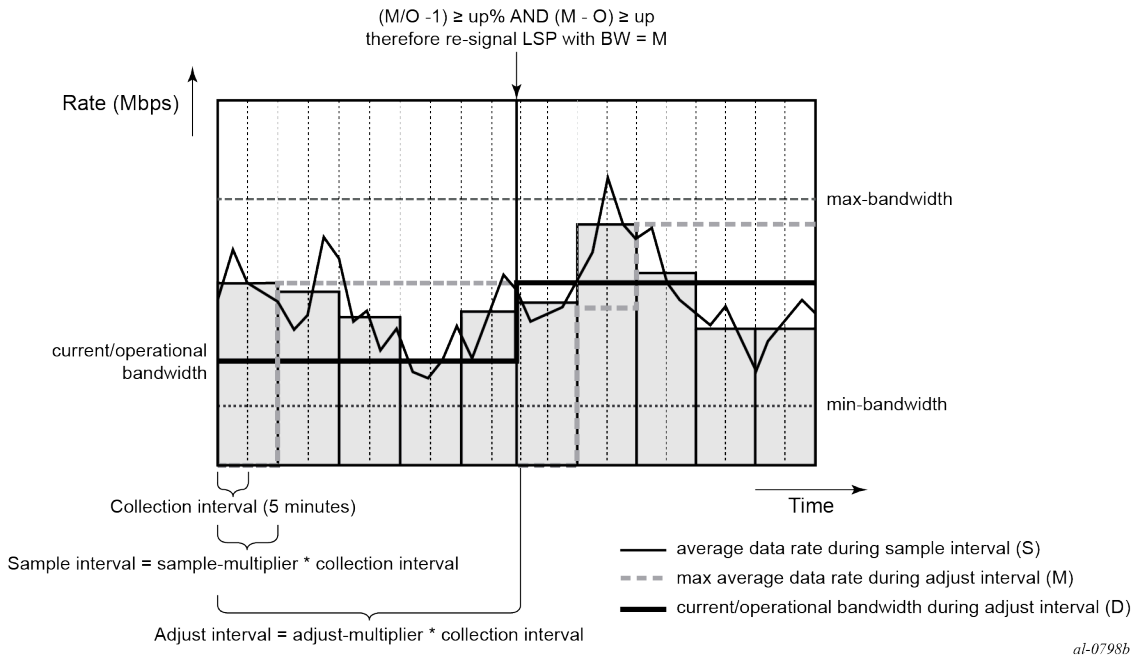
4. Active path change

The LSP has a primary and one or more secondary paths. When there is a change from the primary path to a secondary path without the LSP going down, an auto-bandwidth MBB is triggered. When the primary path becomes active again, another auto-bandwidth MBB is triggered.

### Periodic Trigger

Figure 1: Auto-Bandwidth Adjustment Implementation shows the different time intervals and bandwidths defined in the auto-bandwidth adjustment implementation. In this example, there is an auto-bandwidth attempt when the adjust-interval elapses (periodic trigger). If the auto-bandwidth algorithm is met, the current bandwidth is increased. The parameters are explained after the figure.

Figure 1: Auto-Bandwidth Adjustment Implementation



The time intervals are:

- Collection interval in minutes. This is a global parameter to be set in an accounting policy. Range: 5 to 120 minutes. Default: 5 minutes.

For this kind of record type, the minimum interval is 5 minutes. For policies using a record type of SAA or PM, the minimum is 1 minute.

```
# On PE-1:
configure {
    log accounting-policy 10 collection-interval ?

    collection-interval <number>
    <number>          - <1..120> - minutes
    Dynamic Default - 5

    Accounting collection interval

configure {
    log accounting-policy 10 collection-interval 1
MINOR: LOG #12: configure log accounting-policy 10 collection-interval
    - Inconsistent Value error
    - minimum value is 5 minutes for this record type.
```

- Sample interval: sample-multiplier \* collection interval
- Sample-multiplier is configurable globally in the **mpls** context or per LSP. Default value: 1. In [Figure 1: Auto-Bandwidth Adjustment Implementation](#), the sample multiplier equals 2 for a sample interval of 2 \* 5 minutes = 10 minutes.
- Adjust-interval: adjust-multiplier \* collection interval
  - Nokia recommends that the adjust-multiplier is an integer multiple of the sample-multiplier.
  - Adjust-multiplier is configurable globally in the **mpls** context or per LSP. Default value: 288 (288 \* 5 minutes = 1440 minutes = 24 h). In [Figure 1: Auto-Bandwidth Adjustment Implementation](#), the adjust multiplier equals 10 for an adjust-interval of 10 \* 5 minutes = 50 minutes.

```
# On PE-1:
configure {
    router "Base" mpls auto-bandwidth-multipliers ?

    auto-bandwidth-multipliers

    adjust-multiplier    - Collection intervals in a sample interval
    sample-multiplier    - Collection intervals in a sample interval

configure {
    router "Base" mpls auto-bandwidth-multipliers adjust-multiplier ?

    adjust-multiplier <number>
    <number>          - <1..16383>
    Default          - 288

    Collection intervals in a sample interval

configure {
    router "Base" mpls auto-bandwidth-multipliers sample-multiplier ?

    sample-multiplier <number>
    <number>          - <1..511>
    Default          - 1
```



#### Collection intervals in a sample interval

```
# On PE-1:
configure {
  router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth multipliers ?

  multipliers

  adjust-multiplier    - Number of collection intervals in the adjust interval
  sample-multiplier    - Multiplier for collection intervals in a sample interval

configure {
  router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth multipliers adjust-multiplier ?

  adjust-multiplier <number>
  <number> - <1..16383>

  Number of collection intervals in the adjust interval

configure {
  router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth multipliers sample-multiplier ?

  sample-multiplier <number>
  <number> - <1..511>

  Multiplier for collection intervals in a sample interval
```

The different bandwidths are:

- Minimum bandwidth: configured minimum bandwidth in Mbps that the auto-bandwidth adjustment can signal for the LSP. Granularity: 1 Mbps. Default: 0 Mbps.

```
# On PE-1:
configure {
  router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth min-bandwidth ?

  min-bandwidth <number>
  <number> - <0..6400000> - megabps
  Default - 0

  Minimum bandwidth of auto-bandwidth to request for the LSP
```

- Maximum bandwidth: configured maximum bandwidth in Mbps that the auto-bandwidth adjustment can signal for the LSP. Granularity: 1 Mbps. Default: 100 Mbps.

```
# On PE-1:
configure {
  router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth max-bandwidth ?

  max-bandwidth <number>
  <number> - <0..6400000> - megabps
  Default - 100000

  Maximum bandwidth of auto-bandwidth to request for LSP
```

- Current bandwidth or operational bandwidth (O): currently reserved bandwidth in Mbps for the LSP in the control plane. This is the operational bandwidth that is maintained in the Management Information Base (MIB) and is the bandwidth that is auto-adjusted. Granularity: 1 Mbps.

- Sampled bandwidth (S): average data rate for the last sample interval.
- Measured bandwidth (M): maximum averaged (per sample interval) data rate in the current adjust-interval. The SR OS keeps track of the maximum average data rate of each LSP since the last reset of the adjust-count.
- Signaled bandwidth: bandwidth in Mbps that is provided to the CSPF algorithm and signaled in the RSVP SENDER\_TSPEC and FLOWSPEC objects, when an auto-bandwidth adjustment is attempted. Granularity: 1 Mbps.

The other auto-bandwidth parameters for periodically triggered auto-bandwidth adjustment are:

- Up% (adjust-up in percent): minimum increase in bandwidth from current to measured bandwidth, expressed as a percentage of the current bandwidth. Default: 5%.
- Up (adjust-up bw): minimum increase in bandwidth as absolute bandwidth in Mbps. Up = measuredBW – currentBW. Granularity: 1 Mbps. Default: 0 Mbps.

```
# On PE-1:
configure {
    router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth adjust-up ?

    adjust-up

    bw                - Minimum difference in absolute bandwidth
    percent           - Minimum difference in percent

    configure {
        router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth adjust-up bw ?

        bw <number>
        <number> - <0..6400000> - megabps
        Default - 0

        Minimum difference in absolute bandwidth

    configure {
        router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth adjust-up percent ?

        percent <number>
        <number> - <0..100> - percent
        Default - 5

        Minimum difference in percent
    }
}
```

- Down% (adjust-down in percent): minimum decrease in bandwidth from current to measured bandwidth, expressed as a percentage of the current bandwidth. Default: 5%.
- Down (adjust-down bw): minimum decrease in bandwidth as absolute bandwidth in Mbps. Down = currentBW – measuredBW. Granularity: 1 Mbps. Default: 0 Mbps.

```
# On PE-1:
configure {
    router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth adjust-down ?

    adjust-down

    bw                - Minimum difference in absolute bandwidth
    percent           - Minimum difference in percent
}
```

```

configure {
  router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth adjust-down bw ?

  bw <number>
  <number> - <0..6400000> - megabps
  Default - 0

  Minimum difference in absolute bandwidth

configure {
  router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth adjust-down percent ?

  percent <number>
  <number> - <0..100> - percent
  Default - 5

  Minimum difference in percent

```

In [Figure 1: Auto-Bandwidth Adjustment Implementation](#), the minimum and maximum bandwidths mark the bandwidth range where auto-bandwidth adjustments are allowed. The sample interval is two collection intervals long ( $2 * 5$  minutes = 10 minutes). The adjust-interval is 10 collection intervals long ( $10 * 5$  minutes = 50 minutes). Initially, the current bandwidth (O) equals the configured bandwidth for the primary path. It is good practice to give that same value to the minimum bandwidth for auto-bandwidth. The system doesn't confirm this and these bandwidths are independent from each other.

In this example, the sampled bandwidth exceeds the current bit rate in most of the sample intervals. The maximum sampled bandwidth in the current adjust-interval corresponds to the measured bandwidth (M). When auto-bandwidth adjustment is triggered at the end of the adjust-interval, this measured bandwidth is signaled and, after a successful adjustment, is the new current bandwidth. After the auto-bandwidth adjustment, a new adjust-interval starts and the measured bandwidth is reset to 0. As long as the first sample interval of the new adjust-interval is not finished, the measured bandwidth equals 0 and auto-adjustment is impossible even when triggered manually.

The auto-bandwidth attempt follows these rules:

- When  $\text{measuredBW} \geq \text{currentBW}$ 
  - if  $\{( \text{measuredBW} / \text{currentBW} - 1 ) \geq \text{up}\% \}$  &&  $\{ (\text{measuredBW} - \text{currentBW}) \geq \text{up} \}$  then  $\text{signaledBW} = \text{max}\{(\text{min}(\text{measuredBW}, \text{maxBW})), \text{minBW}\}$
- When  $\text{measuredBW} < \text{currentBW}$ 
  - if  $\{ (1 - \text{measuredBW} / \text{currentBW}) \geq \text{down}\% \}$  &&  $\{ (\text{currentBW} - \text{measuredBW}) \geq \text{down} \}$  then  $\text{signaledBW} = \text{min}\{(\text{max}(\text{measuredBW}, \text{minBW})), \text{maxBW}\}$

CLI configured bandwidths have a granularity of 1 Mbps, while the threshold calculations with measured bandwidth are performed at full precision. This means that the signaled bandwidth in the RSVP message is rounded up to the nearest integer multiple of 1 Mbps.

## Overflow/Underflow Trigger

Auto-bandwidth adjustment can also be triggered by overflow or underflow. When the bandwidth changes drastically, the bandwidth can be auto-adjusted after a number of consecutive overflow/underflow samples. In this case, there is no need to wait for the adjust-interval to end (by default: 24 h).

The parameters used in case of overflow are:

- Overflow sample: a sample interval counts as an overflow sample if the sampled bandwidth is higher than the current bandwidth by at least the configured overflow thresholds.
- Overflow-limit/overflow-count: an auto-bandwidth adjustment occurs after this number of consecutive overflow samples.
- Threshold%: minimum difference between sampled bandwidth and current bandwidth, expressed as a percentage of the current bandwidth.
- Threshold bw: minimum difference between sampled bandwidth and current bandwidth in Mbps. Default value: 0.

```
# On PE-1:
configure {
  router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth overflow-limit ?

  overflow-limit

  bw                - Minimum difference in absolute bandwidth
  number            ^ Number of overflow samples to trigger an overflow auto-
                    bandwidth adjustment attempt
  threshold         - Minimum difference in percent between the LSP bandwidth and
                    the sampled data rate

  configure {
    router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth overflow-limit bw ?

    bw <number>
    <number> - <1..6400000> - megabps

    Minimum difference in absolute bandwidth

  configure {
    router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth overflow-limit number ?

    number <number>
    <number> - <1..10>

    'number' is: mandatory

    Number of overflow samples to trigger an overflow auto-bandwidth adjustment attempt

  configure {
    router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth overflow-limit threshold ?

    threshold <number>
    <number> - <0..100> - percent
    Default - 0

    Minimum difference in percent between the LSP bandwidth and the sampled data rate
```

The rules for overflow-triggered auto-bandwidth adjustment are as follows:

- Overflow sample:  $\{(sampledBW / currentBW - 1) \geq threshold\% \} \&\& \{(sampledBW - currentBW) \geq thresholdBW\}$
- The signaled bandwidth is:
  - if (measuredBW  $\geq$  maxBW) then signaledBW = maxBW
  - if (measuredBW < minBW) then signaledBW = minBW

- else signaledBW = measuredBW

The parameters used in case of underflow are:

- Underflow sample: a sample interval counts as an underflow sample if the sampled bandwidth is lower than the current bandwidth by at least the configured underflow thresholds.
- Underflow-limit/underflow-count: an auto-bandwidth adjustment occurs after this number of consecutive underflow samples.
- Threshold%: minimum difference between current bandwidth and sampled bandwidth, expressed as a percentage of the current bandwidth.
- Threshold bw: minimum difference between current bandwidth and sampled bandwidth in Mbps. Default value: 0.
- Maximum underflow bandwidth (MU): maximum sampled bandwidth in the consecutive underflow samples.

```
# On PE-1:
configure {
  router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth underflow-limit ?

  underflow-limit

  bw                - Minimum difference in absolute bandwidth
  number            ^ Number of overflow samples to trigger an overflow auto-
                    bandwidth adjustment att
  threshold         - Minimum difference in percent between the LSP bandwidth and
                    the sampled data ra

configure {
  router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth underflow-limit bw ?

  bw <number>
  <number> - <1..6400000> - megabps

  Minimum difference in absolute bandwidth

configure {
  router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth underflow-limit number ?

  number <number>
  <number> - <1..10>

  'number' is: mandatory

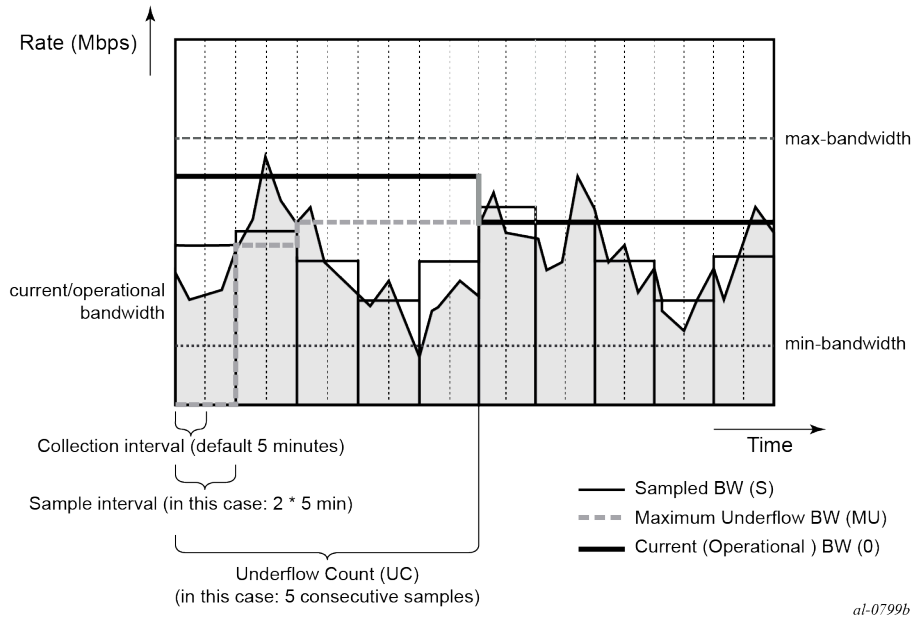
  Number of overflow samples to trigger an overflow auto-bandwidth adjustment attempt

configure {
  router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth underflow-limit threshold ?

  threshold <number>
  <number> - <0..100> - percent
  Default - 0

  Minimum difference in percent between the LSP bandwidth and the sampled data rate
```

Figure 2: Underflow-Triggered Auto-Bandwidth Implementation



In [Figure 2: Underflow-Triggered Auto-Bandwidth Implementation](#), the adjust-interval is not displayed. It is assumed to be the default of 288 collection intervals (24 h). The figure only shows five consecutive underflow samples. The underflow-limit equals 5. In each of the samples, the sample bandwidth is below the underflow threshold. The maximum sampled bandwidth of these five samples corresponds to the maximum underflow bandwidth. This bandwidth is signaled when auto-bandwidth adjustment is triggered because the underflow count is reached.

The rules for underflow-triggered auto-bandwidth adjustment are as follows:

- Underflow sample:
  - $\{(1 - \text{sampledBW} / \text{currentBW}) \geq \text{threshold}\% \} \&\& \{(\text{currentBW} - \text{sampledBW}) \geq \text{thresholdBW}\}$
- Underflow count/underflow limit: after that many consecutive underflow samples, an auto-bandwidth adjustment is triggered.
- The signaled bandwidth is:
  - if  $(\text{maxUnderflowBW} \geq \text{maxBW})$  then signaledBW = maxBW
  - if  $(\text{maxUnderflowBW} < \text{minBW})$  then signaledBW = minBW
  - else signaledBW = maxUnderflowBW

If the adjustment is successful, the sample counter within the adjust-interval is reset, along with other parameters, such as the maximum underflow bandwidth, the measured bandwidth, and the underflow count. The next adjust-interval will elapse in 24 h.

If the adjustment fails, there are 5 retries. If they all fail, only the underflow count and the maximum underflow bandwidth are reset. The current adjust-interval continues.

## Manual Trigger

Besides the periodic trigger and the overflow/underflow trigger, an operator can launch a **tools** command to trigger an auto-bandwidth adjustment.

```
# On PE-1:
tools perform router "Base" mpls adjust-autobandwidth ?

adjust-autobandwidth [lsp <string>]
adjust-autobandwidth lsp <string> [force]
adjust-autobandwidth lsp <string> force [bandwidth <number>]

lsp                - string '<1..64 characters>'

# On PE-1:
tools perform router "Base" mpls adjust-autobandwidth
```

This **tools** command can be launched with or without explicit LSP name. In the latter case, all active LSPs are attempted for auto-bandwidth.

```
# On PE-1:
tools perform router "Base" mpls adjust-autobandwidth lsp "LSP-PE-1-PE-2" ?

adjust-autobandwidth [lsp <string>]
adjust-autobandwidth lsp <string> [force]
adjust-autobandwidth lsp <string> force [bandwidth <number>]

force                - keyword

# On PE-1:
tools perform router "Base" mpls adjust-autobandwidth lsp "LSP-PE-1-PE-2" force ?

adjust-autobandwidth lsp <string> [force]
adjust-autobandwidth lsp <string> force [bandwidth <number>]

bandwidth            - rate-in-mbps

# On PE-1:
tools perform router "Base" mpls adjust-autobandwidth lsp "LSP-PE-1-PE-2" force bandwidth ?

bandwidth (<number> | <expression>)
<expression> - range ([])
[0..6400000]

# On PE-1:
tools perform router "Base" mpls adjust-autobandwidth lsp "LSP-PE-1-PE-2"
```

This command (without the keyword **force**) triggers a new auto-bandwidth calculation according to the rules of periodic triggered type. If the LSP already has the correct reserved bandwidth, the following message is returned.

```
# On PE-1:
tools perform router "Base" mpls adjust-autobandwidth lsp "LSP-PE-1-PE-2"
MINOR: CLI #2005: Error while processing command
- lsp LSP-PE-1-PE-2 active path is already at the requested value 10 Mbps.
```

If the keyword **force** is added without a specific value for the bandwidth, there is no threshold checking. The bandwidth can also be adjusted if the difference in bandwidth is below the thresholds. The granularity remains 1 Mbps.

```
# On PE-1:
tools perform router "Base" mpls adjust-autobandwidth lsp "LSP-PE-1-PE-2" force
```

The rules for the signaled bandwidth are unchanged:

- if (measuredBW  $\geq$  maxBW) then signaledBW = maxBW
- if (measuredBdBW < minBW) then signaledBW = minBW
- else signaledBW = measuredBW

If the keyword **force** with **bandwidth** (in Mbps) option is given, the signaled bandwidth is set to this configured bandwidth, even if it is a value below the minimum or higher than the maximum bandwidth.

```
# On PE-1:
tools perform router "Base" mpls adjust-autobandwidth lsp "LSP-PE-1-PE-2" force bandwidth 30
```

After a manually triggered auto-bandwidth MBB, no counters are reset. The ongoing adjust-interval is not aborted.

A **clear** command resets all counters and timers associated with auto-bandwidth adjustment on a specified LSP.

```
# On PE-1:
clear router "Base" mpls lsp-autobandwidth "LSP-PE-1-PE-2"
```

## Passive Monitoring

The system offers the option to measure the bandwidth of an LSP without taking any action to adjust the bandwidth reservation.

```
# On PE-1:
configure {
  router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth monitor-bandwidth
```

## Auto-Bandwidth Based on Forwarding Class

From SR OS Release 11.0.R4 onward, the bandwidth can be calculated as a weighted sum of all the traffic in the eight forwarding classes. By default, all forwarding classes have the same weight: 100%, but that sampling weight is configurable.

```
# On PE-1:
configure {
  router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth fc ?

  [fc-name] (<keyword> | <expression>)
  <keyword>      - (be|l2|af|l1|h2|ef|h1|nc)
  <expression>  - wildcard (*), regular expression (') or range ([])

  Forwarding class name for LSP
```



```
configure {
  router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth fc fc-name ?

  [fc-name] (<keyword> | <expression>)
  <keyword>      - (be|l2|af|l1|h2|ef|h1|nc)
  <expression> - wildcard (*), regular expression (') or range ([])

  Forwarding class name for LSP

configure {
  router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth fc fc-name be ?

  fc

  apply-groups          - Apply a configuration group at this level
  apply-groups-exclude - Exclude a configuration group at this level
  sampling-weight       - Forwarding class weight for LSP

configure {
  router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth fc fc-name be sampling-weight ?

  sampling-weight <number>
  <number> - <0..100> - percent
  Default - 100

  Forwarding class weight for LSP
```

## Active Path Change

Auto-bandwidth adjustment is also supported on LSPs that have secondary paths. If the secondary path is standby, an auto-bandwidth MBB can be triggered when the active path changes from primary to secondary. The secondary/standby path is only initialized at its configured bandwidth when it is established, and the bandwidth is adjusted only when it becomes active. This happens when the primary path goes down or becomes degraded. When another path becomes active, the bandwidth used to signal the auto-bandwidth MBB is the operational bandwidth of the previous path.

The definition for current bandwidth is modified for this feature:

- Current bandwidth: last known reserved bandwidth for the LSP. This may be for a different path than the active one.

Auto-bandwidth adjustment will only take place on the active path. When the active path changes, the current bandwidth is updated to the operational bandwidth of the new active path.

- For a secondary path that is signaled as standby, if the active path for an LSP changes without the LSP going down, an auto-bandwidth MBB is triggered on the new active path. The signaled bandwidth is the operational bandwidth of the previous path. The reserved bandwidth of the new active path is its configured bandwidth until the MBB succeeds.
- For a secondary path where the active path goes down, the LSP will go down temporarily until the secondary path is set up. When the LSP goes down, all statistics and counters are cleared, so the previous path operational bandwidth is lost. There is no immediate bandwidth adjustment on the secondary path.

The following rules apply to determine the signaled bandwidth of the new active path.

- For a path that is operationally down, signaledBW = configuredBW.

- For the first 5 MBB attempts on the path that just became active, signaledBW = currentBW (operational bandwidth of the previous path).

For the remaining MBB attempts, signaledBW = operationalBW.

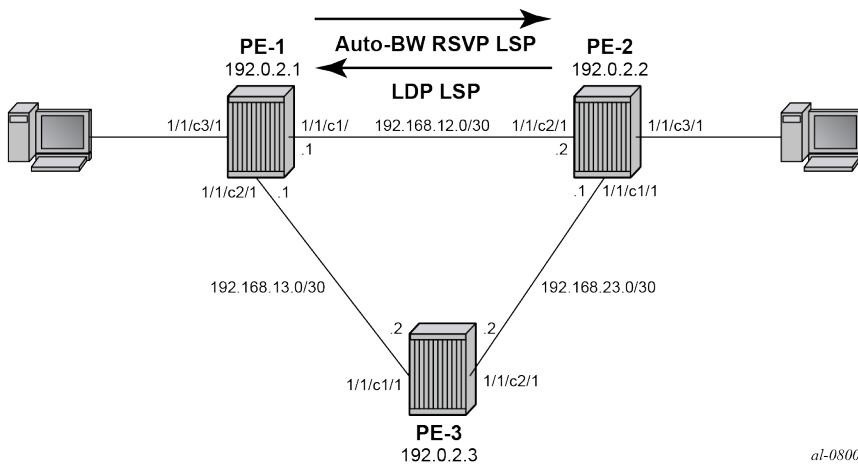
- For all MBBs other than auto-bandwidth MBB on the active path, MBB signaledBW = operationalBW.
- For an MBB on the inactive (standby) path, MBB signaledBW = configuredBW.

When the system reverts from a secondary standby path to the primary path, a Delayed Retry MBB is attempted to bring the bandwidth on the standby path back to the configured bandwidth. MBB is attempted once, and if it fails, the standby is torn down. A Delayed Retry MBB has the highest priority among MBBs, so it will take precedence over any other MBB in progress on the standby path, such as configuration change or pre-emption.

## Configuration

[Figure 3: Example Setup for Auto-Bandwidth Point-to-Point LSPs](#) shows the example setup. The focus is on the RSVP LSP from PE-1 to PE-2.

Figure 3: Example Setup for Auto-Bandwidth Point-to-Point LSPs



## Base Configuration

The cards, MDAs and ports need to be configured.

Configure the interfaces on all nodes.

```
# On PE-1:
configure {
  router "Base" {
    interface "int-PE-1-PE-2" {
      port 1/1/c1/1
      ipv4 {
        primary {
          address 192.168.12.1
        }
      }
    }
  }
}
```



```

    }
  }
}

```

Enable MPLS and RSVP on all nodes and add all interfaces to the **mpls** context. They will automatically be added to the **rsvp** context.

```

# On PE-1:
configure {
  router "Base" {
    mpls {
      admin-state enable
      interface "int-PE-1-PE-2" { }
      interface "int-PE-1-PE-3" { }
    }
    rsvp {
      admin-state enable
      interface "int-PE-1-PE-2" { }
      interface "int-PE-1-PE-3" { }
    }
  }
}

```

Configure a path with no explicitly defined hops and LSP LSP-PE-1-PE-2 on PE-1:

```

# On PE-1:
configure {
  router "Base" {
    mpls {
      path "loose" {
        admin-state enable
      }
      lsp "LSP-PE-1-PE-2" {
        admin-state enable
        type p2p-rsvp
        to 192.0.2.2
        primary "loose" { }
      }
    }
  }
}

```

In the example, traffic needs to be injected into the LSP tunnel. For that, a VPLS service is created. On PE-1, an SDP using the RSVP LSP to PE-2 is created.

```

# On PE-1:
configure {
  service {
    sdp 212 {
      admin-state enable
      description "SDP-PE-1-PE-2-overRSVP-TE"
      delivery-type mpls
      far-end {
        ip-address 192.0.2.2
      }
      lsp "LSP-PE-1-PE-2" { }
    }
  }
}

```

On PE-2, an SDP using LDP is created toward PE-1.

```
# On PE-2:
configure {
  service {
    sdp 121 {
      admin-state enable
      description "SDP-PE-2-PE-1-overLDP"
      delivery-type mpls
      ldp true
      far-end {
        ip-address 192.0.2.1
      }
    }
  }
}
```

On PE-1 and PE-2, a VPLS is created.

```
# On PE-1:
configure {
  service {
    vpls "VPLS 100" {
      admin-state enable
      service-id 100
      customer "1"
      spoke-sdp 212:100 { }
      sap 1/1/c3/1 { }
    }
  }
}
```

The configuration on PE-2 is similar.

## Pre-requisites for Auto-Bandwidth LSP Configuration

Enable Constrained Shortest Path First (CSPF) on the LSP by adding **path-computation-method local-cspf**.

```
# On PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-2" {
        path-computation-method local-cspf
      }
    }
  }
}
```

The bandwidth of the LSP is adjusted in a Make-Before-Break (MBB) manner. Enable MBB on the LSP by adding **adaptive true** to the primary path.

```
# On PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-2" {
```

```

        primary "loose" {
            adaptive true
        }
    }
}

```

Enter a value for the bandwidth in Mbps for the primary path. It is good practice to configure the same value as for the minimum bandwidth in the auto-bandwidth settings.

```

# On PE-1:
configure {
    router "Base" {
        mpls {
            lsp "LSP-PE-1-PE-2" {
                primary "loose" {
                    bandwidth 2
                }
            }
        }
    }
}

```

## Auto-Bandwidth LSP Configuration

MPLS auto-bandwidth adjustment allows the ingress LER to dynamically adjust the bandwidth of an RSVP tunnel based on active measurements of the traffic rate into the tunnel. Therefore, LSP egress statistics need to be enabled on the iLER.

Auto-bandwidth adjustment requires an accounting policy to be defined and operational. The accounting policy specifies the collection interval for LSP statistics collection, which is fundamental to the auto-bandwidth algorithm. The minimum interval for this type of collection is 5 minutes, which is the default value.

```

# On PE-1:
configure {
    log {
        accounting-policy 10 {
            admin-state enable
            record combined-mpls-lsp-egress
            destination {
                null
            }
            collection-interval 5 ## default (in minutes)
        }
    }
}

```

An accounting policy of record type **combined-mpls-lsp-egress** doesn't need a reference to a specific file ID.

From the moment auto-bandwidth is enabled within an **lsp** context, the record **combined-mpls-lsp-egress** inside the accounting policy will also take bandwidth measurements.

```

# On PE-1:
configure {
    log {

```

```

        accounting-policy 10 {
            destination {
                null
            }
        }
    }
}

```

When the **destination { null }** is configured, no LSP statistics are stored anymore. The MPLS auto-bandwidth feature retrieves its LSP statistics information directly from the statistics module.

However, the accounting policy can reference a file and, therefore, a CF card. An additional CF card may be required in each node as a storage location.

```

# On PE-1:
configure {
    log {
        file "77" {
            rollover 15
            retention 1
            compact-flash-location {
                primary cfl
            }
        }
        accounting-policy 77 {
            admin-state enable
            record combined-mpls-lsp-egress
            destination {
                file "77"
            }
        }
    }
}

```

In the remainder of the example, the accounting policy will reference to destination { null }.

After the accounting policy has been created, egress statistics can be enabled on the LSP.

```

# On PE-1:
configure {
    router "Base" {
        mpls {
            lsp "LSP-PE-1-PE-2" {
                egress-statistics {
                    admin-state enable
                    collect-stats true
                    accounting-policy 10
                }
            }
        }
    }
}

```

The system does not verify whether egress statistics have been enabled on the LSP. When a user configures auto-bandwidth adjustment, but without enabling egress statistics, no auto-bandwidth measurements and adjustments are performed. The operational state of auto-bandwidth (**AB OpState**) is down.

Enable auto-bandwidth with default settings by adding the keyword **auto-bandwidth** to the LSP.

```

# On PE-1:

```

```
configure {
  router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth
```

The actual values are shown in the following output. They are explained after the output.

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth

=====
MPLS LSP (Auto Bandwidth)
=====
Legend :
  + - Inherited
=====
-----
Type : Originating
-----
LSP Name   : LSP-PE-1-PE-2
Auto BW    : Enabled                AB OpState      : Up
Auto BW Min : 0 Mbps                Auto BW Max     : 100000 Mbps
AB Up Thresh : 5 percent            AB Down Thresh  : 5 percent
AB Up BW    : 0 Mbps                AB Down BW     : 0 Mbps
AB Curr BW  : 2 Mbps                AB Samp Intv   : 5 Mins
AB Adj Mul  : 288+                  AB Samp Mul    : 1+
AB Adj Time : 1440 Mins             AB Samp Time   : 5 Mins
AB Adj Cnt  : 0                    AB Samp Cnt    : 0
AB Last Adj : n/a                  AB Next Adj    : 1440 Mins
ABMaxAvgRt  : 0 Mbps                AB Lst AvgRt   : 0 Mbps
AB Ovfl Lmt : 0                    AB Ovfl Cnt    : 0
AB0vflThres : 0 percent            AB Ovfl BW     : 0 Mbps
AB UndflLmt : 0                    AB Undrfl Cnt  : 0
ABUndflThrs : 0 percent            AB Undrfl BW   : 0 Mbps
ABMaxUndflBW : 0 Mbps
AB Adj Cause : none                AB Monitor BW  : False
Be Weight   : 100 percent          Af Weight      : 100 percent
L1 Weight   : 100 percent          L2 Weight      : 100 percent
Nc Weight   : 100 percent          Ef Weight      : 100 percent
H1 Weight   : 100 percent          H2 Weight      : 100 percent
Use Last Adj BW : Disabled
=====
```

The plus sign (+) indicates that the value is inherited from the global MPLS settings (**AB Adj Mul**: 288+ and **AB Samp Mul**: 1+). The sample-multiplier and the adjust-multiplier can both be configured globally in the **mpls** context or overruled by the settings per LSP. In this example, nothing has been configured in the **mpls** context or in the LSP. Therefore, the default values as defined in the **mpls** context are applicable.

## Auto-Bandwidth – Periodic Trigger (Normal)

The default collection interval is 5 minutes. The sample-multiplier is 1, by default. The sample interval equals 1 \* 5 minutes = 5 minutes, by default. The adjust-multiplier is 288, by default. The adjustment interval equals 288 \* 5 minutes = 1440 minutes (24 hours), by default.

The auto-bandwidth settings for the LSP are modified as follows:

```
# On PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-2" {
```





```

LSP Name      : LSP-PE-1-PE-2
Auto BW       : Enabled
Auto BW Min   : 2 Mbps
AB Up Thresh  : 10 percent
AB Up BW      : 1 Mbps
AB Curr BW    : 2 Mbps
AB Adj Mul    : 3
AB Adj Time   : 15 Mins
AB Adj Cnt    : 1
AB Last Adj   : n/a
ABMaxAvgRt    : 10 Mbps
AB Ovfl Lmt   : 0
ABOvflThres  : 0 percent
AB UndflLmt   : 0
ABUndflThrs  : 0 percent
ABMaxUndflBW : 0 Mbps
AB Adj Cause  : none
Be Weight     : 100 percent
L1 Weight     : 100 percent
Nc Weight     : 100 percent
H1 Weight     : 100 percent
Use Last Adj BW : Disabled

AB OpState    : Up
Auto BW Max   : 20 Mbps
AB Down Thresh : 5 percent
AB Down BW    : 0 Mbps
AB Samp Intv  : 5 Mins
AB Samp Mul   : 1
AB Samp Time  : 5 Mins
AB Samp Cnt   : 0
AB Next Adj   : 10 Mins
AB Lst AvgRt  : 10 Mbps
AB Ovfl Cnt   : 0
AB Ovfl BW    : 0 Mbps
AB Undrfl Cnt : 0
AB Undrfl BW  : 0 Mbps

AB Monitor BW : False
Af Weight     : 100 percent
L2 Weight     : 100 percent
Ef Weight     : 100 percent
H2 Weight     : 100 percent
=====

```

After 10 minutes, another collection interval has elapsed in the adjust-interval (**AB Adj Cnt = 2**) and the next adjustment time is in 5 minutes (**AB Next Adj = 5 Min**).

```

[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth

=====
MPLS LSP (Auto Bandwidth)
=====
Legend :
+ - Inherited
=====
-----
Type : Originating
-----
LSP Name      : LSP-PE-1-PE-2
Auto BW       : Enabled
Auto BW Min   : 2 Mbps
AB Up Thresh  : 10 percent
AB Up BW      : 1 Mbps
AB Curr BW    : 2 Mbps
AB Adj Mul    : 3
AB Adj Time   : 15 Mins
AB Adj Cnt    : 2
AB Last Adj   : n/a
ABMaxAvgRt    : 10 Mbps
AB Ovfl Lmt   : 0
ABOvflThres  : 0 percent
AB UndflLmt   : 0
ABUndflThrs  : 0 percent
ABMaxUndflBW : 0 Mbps
AB Adj Cause  : none
Be Weight     : 100 percent
L1 Weight     : 100 percent
Nc Weight     : 100 percent
H1 Weight     : 100 percent
Use Last Adj BW : Disabled

AB OpState    : Up
Auto BW Max   : 20 Mbps
AB Down Thresh : 5 percent
AB Down BW    : 0 Mbps
AB Samp Intv  : 5 Mins
AB Samp Mul   : 1
AB Samp Time  : 5 Mins
AB Samp Cnt   : 0
AB Next Adj   : 5 Mins
AB Lst AvgRt  : 10 Mbps
AB Ovfl Cnt   : 0
AB Ovfl BW    : 0 Mbps
AB Undrfl Cnt : 0
AB Undrfl BW  : 0 Mbps

AB Monitor BW : False
Af Weight     : 100 percent
L2 Weight     : 100 percent
Ef Weight     : 100 percent
H2 Weight     : 100 percent
=====

```

After 15 minutes, auto-bandwidth adjustment occurs. **AB Adj Cause** is normal for periodically triggered adjustments. The next adjustment interval will elapse in 15 minutes. The measured bandwidth **ABMaxAvgRt** is reset to 0 after a successful adjustment.

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth

=====
MPLS LSP (Auto Bandwidth)
=====
Legend :
+ - Inherited
=====
-----
Type : Originating
-----
LSP Name      : LSP-PE-1-PE-2
Auto BW       : Enabled                AB OpState      : Up
Auto BW Min   : 2 Mbps                 Auto BW Max     : 20 Mbps
AB Up Thresh  : 10 percent              AB Down Thresh  : 5 percent
AB Up BW      : 1 Mbps                  AB Down BW      : 0 Mbps
AB Curr BW    : 10 Mbps                 AB Samp Intv    : 5 Mins
AB Adj Mul    : 3                      AB Samp Mul     : 1
AB Adj Time   : 15 Mins                 AB Samp Time    : 5 Mins
AB Adj Cnt    : 0                      AB Samp Cnt     : 0
AB Last Adj   : 03/19/2024 21:08:58    AB Next Adj     : 15 Mins
ABMaxAvgRt   : 0 Mbps                AB Lst AvgRt    : 10 Mbps
AB Ovfl Lmt   : 0                      AB Ovfl Cnt     : 0
ABOvflThres  : 0 percent                AB Ovfl BW      : 0 Mbps
AB UndflLmt   : 0                      AB Undrfl Cnt   : 0
ABUndflThrs  : 0 percent                AB Undrfl BW    : 0 Mbps
ABMaxUndflBW : 0 Mbps
AB Adj Cause  : normal                  AB Monitor BW   : False
Be Weight     : 100 percent              Af Weight       : 100 percent
L1 Weight     : 100 percent              L2 Weight       : 100 percent
Nc Weight     : 100 percent              Ef Weight       : 100 percent
H1 Weight     : 100 percent              H2 Weight       : 100 percent
Use Last Adj BW : Disabled
=====
```

The periodic trigger type rules for auto-bandwidth are:

- When measuredBW ≥ currentBW
  - if {(measuredBW / currentBW – 1) ≥ up%} &&{(measuredBW – currentBW) ≥ up}
    - then signaledBW = max{(min(measuredBW, maxBW)), minBW}

In this case, the measuredBW (10 Mbps) is greater than the currentBW (2 Mbps). The increase is at least 10% (up%) and at least 1 Mbps (up). The bandwidth is adjusted. The new bandwidth that is signaled is calculated as follows:

```
signaledBW = max{(min(measuredBW, maxBW)), minBW}
signaledBW = max {(min (10 Mbps, 20 Mbps)), 2 Mbps}
signaledBW = max {10 Mbps, 2 Mbps}
signaledBW = 10 Mbps
```

Whenever an auto-bandwidth adjustment is performed, a message is stored in log 99.

```
[/]
A:admin@PE-1# show log log-id 99 application "mpls"
```

```

=====
Event Log 99 log-name 99
=====
Description : Default System Log
Memory Log contents [size=500  next event=197  (not wrapped)]

196 2024/03/19 21:08:57.987 UTC WARNING: MPLS #2014 Base VR 1:
"LSP path LSP-PE-1-PE-2::loose resigned as a result of autoBandwidth MBB"

```

When the maximum bandwidth is modified to a value that is lower than the current bandwidth, an adjustment occurs at the end of the adjustment interval.

```

# On PE-1:
configure {
  router "Base" mpls lsp "LSP-PE-1-PE-2" auto-bandwidth max-bandwidth 9
}

```

The current bandwidth is reduced to 9 Mbps (for a measured bandwidth of 10 Mbps).

```

[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth

=====
MPLS LSP (Auto Bandwidth)
=====
Legend :
+ - Inherited
=====
-----
Type : Originating
-----
LSP Name      : LSP-PE-1-PE-2
Auto BW       : Enabled                AB OpState    : Up
Auto BW Min   : 2 Mbps                 Auto BW Max   : 9 Mbps
AB Up Thresh  : 10 percent              AB Down Thresh : 5 percent
AB Up BW      : 1 Mbps                 AB Down BW    : 0 Mbps
AB Curr BW   : 9 Mbps                 AB Samp Intv  : 5 Mins
AB Adj Mul    : 3                      AB Samp Mul   : 1
AB Adj Time   : 15 Mins                 AB Samp Time  : 5 Mins
AB Adj Cnt    : 1                      AB Samp Cnt   : 0
AB Last Adj   : 03/19/2024 21:23:58    AB Next Adj   : 10 Mins
ABMaxAvgRt    : 10 Mbps                 AB Lst AvgRt  : 10 Mbps
AB Ovfl Lmt   : 0                      AB Ovfl Cnt   : 0
ABOvflThres  : 0 percent                AB Ovfl BW    : 0 Mbps
AB UndflLmt   : 0                      AB Undrfl Cnt : 0
ABUndflThrs  : 0 percent                AB Undrfl BW  : 0 Mbps
ABMaxUndflBW : 0 Mbps
AB Adj Cause  : normal                  AB Monitor BW : False
Be Weight     : 100 percent              Af Weight     : 100 percent
L1 Weight     : 100 percent              L2 Weight     : 100 percent
Nc Weight     : 100 percent              Ef Weight     : 100 percent
H1 Weight     : 100 percent              H2 Weight     : 100 percent
Use Last Adj BW : Disabled
=====

```

## Auto-Bandwidth - Passive Monitoring

When passive monitoring is enabled, no automatic bandwidth adjustments occurs. When the maximum bandwidth is again raised to 20 Mbps, the bandwidth will not be auto-adjusted even if the measured bandwidth is high enough.

```
# On PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-2" {
        auto-bandwidth {
          max-bandwidth 20
          monitor-bandwidth true
        }
      }
    }
  }
}
```

The system monitors the bandwidth, but without taking action at the end of the adjust-interval.

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth

=====
MPLS LSP (Auto Bandwidth)
=====
Legend :
+ - Inherited
=====
Type : Originating
-----
LSP Name   : LSP-PE-1-PE-2
Auto BW    : Enabled                AB OpState      : Up
Auto BW Min : 2 Mbps                Auto BW Max     : 20 Mbps
AB Up Thresh : 10 percent           AB Down Thresh  : 5 percent
AB Up BW    : 1 Mbps                AB Down BW      : 0 Mbps
AB Curr BW  : 9 Mbps                AB Samp Intv    : 5 Mins
AB Adj Mul  : 3                    AB Samp Mul     : 1
AB Adj Time : 15 Mins               AB Samp Time    : 5 Mins
AB Adj Cnt  : 2                    AB Samp Cnt     : 0
AB Last Adj : 03/19/2024 21:23:58  AB Next Adj     : 5 Mins
ABMaxAvgRt  : 10 Mbps               AB Lst AvgRt    : 10 Mbps
AB Ovfl Lmt : 0                    AB Ovfl Cnt     : 0
AB0vflThres : 0 percent             AB Ovfl BW      : 0 Mbps
AB UndflLmt : 0                    AB Undrfl Cnt   : 0
ABUndflThrs : 0 percent             AB Undrfl BW    : 0 Mbps
ABMaxUndflBW : 0 Mbps
AB Adj Cause : normal              AB Monitor BW : True
Be Weight   : 100 percent           Af Weight       : 100 percent
L1 Weight   : 100 percent           L2 Weight       : 100 percent
Nc Weight   : 100 percent           Ef Weight       : 100 percent
H1 Weight   : 100 percent           H2 Weight       : 100 percent
Use Last Adj BW : Disabled
=====
```

The value for the parameter **AB Monitor BW** is True

For the remainder of the chapter, there is no passive monitoring. The settings are restored to normal:

```
# On PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-2" {
        auto-bandwidth {
          delete monitor-bandwidth
        }
      }
    }
  }
}
```

## Auto-Bandwidth – Overflow and Underflow Trigger Type

With default settings, the adjustment interval is 24 hours. If the bandwidth changes significantly since the start of the current adjust-interval, overflow and underflow triggers can be used. This will speed up the auto-bandwidth adjustment.

Stop auto-bandwidth in order to force an MBB attempt toward the configured primary path bandwidth (2 Mbps in this example).

```
# On PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-2" {
        delete auto-bandwidth
      }
    }
  }
}
```

Check the operational bandwidth of the LSP.

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-2" detail

=====
MPLS LSPs (Originating) (Detail)
=====
Legend :
+ - Inherited
-----
Type : Originating
-----
LSP Name   : LSP-PE-1-PE-2
LSP Type   : RegularLsp           LSP Tunnel ID       : 1
LSP Index  : 1                   TTM Tunnel Id       : 1
From       : 192.0.2.1
To         : 192.0.2.2
Adm State  : Up                   Oper State           : Up
---snip---
Primary(a) : loose
```

```

Bandwidth           : 2 Mbps
Up Time             : 0d 01:15:19
=====
* indicates that the corresponding row element may have been truncated.

```

Enable auto-bandwidth with similar settings as before and add overflow and underflow triggers. The multipliers are default. Therefore, a periodically triggered auto-adjustment will only take place once every 24 hours.

```

# On PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-2" {
        auto-bandwidth {
          max-bandwidth 20
          min-bandwidth 2
          adjust-up {
            percent 10
            bw 1
          }
          multipliers {
            sample-multiplier 1
            adjust-multiplier 288
          }
          overflow-limit {
            threshold 10
            bw 2
            number 1
          }
          underflow-limit {
            threshold 10
            bw 2
            number 3
          }
        }
      }
    }
  }
}

```

The overflow count indicates the number of consecutive times that the overflow condition is detected at the end of a sample interval. Auto-bandwidth adjustment occurs after that number of overflow samples is reached, in this case, after the first overflow sample (overflow-limit = 1). The conditions for an overflow sample are:

$$\{(sampledBW / currentBW - 1) \geq threshold\% \} \&\& \{(sampledBW - currentBW) \geq thresholdBW\}$$

$$\{(7 \text{ Mbps} / 2 \text{ Mbps} - 1) \geq 0,1 \} \&\& \{(7 \text{ Mbps} - 2 \text{ Mbps}) \geq 2 \text{ Mbps}\}$$

The signaled bandwidth is:

- if (measuredBW ≥ maxBW) then signaledBW = maxBW
- if (measuredBW < minBW) then signaledBW = minBW  
else signaledBW = measuredBW
- if (7 Mbps ≥ 20 Mbps) then signaledBW = 20 Mbps;
- if (7 Mbps < 2 Mbps) then signaledBW = 2 Mbps;  
else signaledBW = 7 Mbps

Display the auto-bandwidth data. The **AB Adj Cause** is now *overflow*. The overflow limit is the configured value of 1 (**AB Ovfl Lmt**). The overflow count has been reset (**AB Ovfl Cnt** = 0) after the auto-bandwidth was adjusted, along with the **ABMaxAvgRt**. This is the start of a new adjust-interval of 24 hours.

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth

=====
MPLS LSP (Auto Bandwidth)
=====
Legend :
+ - Inherited
=====
-----
Type : Originating
-----
LSP Name      : LSP-PE-1-PE-2
Auto BW       : Enabled                AB OpState      : Up
Auto BW Min   : 2 Mbps                 Auto BW Max     : 20 Mbps
AB Up Thresh  : 10 percent              AB Down Thresh  : 5 percent
AB Up BW      : 1 Mbps                 AB Down BW      : 0 Mbps
AB Curr BW    : 7 Mbps                 AB Samp Intv    : 5 Mins
AB Adj Mul    : 288                    AB Samp Mul     : 1
AB Adj Time   : 1440 Mins              AB Samp Time    : 5 Mins
AB Adj Cnt    : 0                      AB Samp Cnt     : 0
AB Last Adj   : 03/19/2024 22:03:58    AB Next Adj     : 1440 Mins
ABMaxAvgRt    : 0 Mbps                 AB Lst AvgRt   : 7 Mbps
AB Ovfl Lmt   : 1                      AB Ovfl Cnt     : 0
ABOvflThres   : 10 percent             AB Ovfl BW      : 2 Mbps
AB UndflLmt   : 3                      AB Undrfl Cnt   : 0
ABUndflThrs   : 10 percent             AB Undrfl BW    : 2 Mbps
ABMaxUndflBW  : 0 Mbps
AB Adj Cause : overflow              AB Monitor BW   : False
Be Weight     : 100 percent             Af Weight       : 100 percent
L1 Weight     : 100 percent             L2 Weight      : 100 percent
Nc Weight     : 100 percent             Ef Weight       : 100 percent
H1 Weight     : 100 percent             H2 Weight      : 100 percent
Use Last Adj BW : Disabled
=====
```

In the following example, the current bandwidth is 7 Mbps, but the bandwidth dropped to 4 Mbps and the conditions for underflow are met. At the end of a sample interval, the sampled bandwidth is reduced by at least 10% and at least 2 Mbps, and this becomes an underflow sample. The conditions for an underflow sample are:

$$\{(1 - \text{sampledBW} / \text{currentBW}) \geq \text{threshold}\} \&\& \{(\text{currentBW} - \text{sampledBW}) \geq \text{thresholdBW}\}$$

$$\{(1 - 4 \text{ Mbps} / 7 \text{ Mbps}) \geq 0,1\} \&\& \{(7 \text{ Mbps} - 4 \text{ Mbps}) \geq 2 \text{ Mbps}\}$$

The underflow limit equals 3, so an auto-bandwidth adjustment can only take place after the third consecutive underflow sample. The new bandwidth will equal the **maximum sampled underflow bandwidth (ABMaxUndflBW)**. This is the maximum sampled bandwidth in the three consecutive underflow samples.

The signaled bandwidth is:

- if ( $\text{maxUnderflowBW} \geq \text{maxBW}$ ) then signaledBW = maxBW
- if ( $\text{maxUnderflowBW} < \text{minBW}$ ) then signaledBW = minBW  
else signaledBW = maxUnderflowBW
- if ( $4 \text{ Mbps} \geq 20 \text{ Mbps}$ ) then signaledBW = 20 Mbps;



- if (4 Mbps < 2 Mbps) then signaledBW = 2 Mbps;  
else signaledBW = 4 Mbps

The following output shows the auto-bandwidth data after two consecutive underflow samples (**AB Underfl Cnt: 2**). The maximum sampled underflow bandwidth equals 4 Mbps. No bandwidth adaptation can take place until there are three consecutive underflow samples (**AB UndflLmt: 3**).

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth

=====
MPLS LSP (Auto Bandwidth)
=====
Legend :
+ - Inherited
=====
-----
Type : Originating
-----
LSP Name      : LSP-PE-1-PE-2
Auto BW       : Enabled                AB OpState    : Up
Auto BW Min   : 2 Mbps                 Auto BW Max   : 20 Mbps
AB Up Thresh  : 10 percent              AB Down Thresh : 5 percent
AB Up BW     : 1 Mbps                  AB Down BW    : 0 Mbps
AB Curr BW    : 7 Mbps                 AB Samp Intv  : 5 Mins
AB Adj Mul    : 288                    AB Samp Mul   : 1
AB Adj Time   : 1440 Mins               AB Samp Time  : 5 Mins
AB Adj Cnt    : 3                      AB Samp Cnt   : 0
AB Last Adj   : 03/19/2024 22:03:58    AB Next Adj   : 1425 Mins
ABMaxAvgRt   : 7 Mbps                  AB Lst AvgRt  : 4 Mbps
AB Ovfl Lmt   : 1                      AB Ovfl Cnt   : 0
ABOvflThres  : 10 percent              AB Ovfl BW    : 2 Mbps
AB UndflLmt  : 3                       AB Undrfl Cnt : 2
ABUndflThrs  : 10 percent              AB Undrfl BW  : 2 Mbps
ABMaxUndflBW : 4 Mbps
AB Adj Cause  : overflow                AB Monitor BW : False
Be Weight    : 100 percent              Af Weight     : 100 percent
L1 Weight    : 100 percent              L2 Weight     : 100 percent
Nc Weight    : 100 percent              Ef Weight     : 100 percent
H1 Weight    : 100 percent              H2 Weight     : 100 percent
Use Last Adj BW : Disabled
=====
```

After a successful auto-bandwidth adjustment, the **ABMaxUndflBW** is reset, along with the **AB Adj Cnt**, **AB Underfl Cnt**, and **ABMaxAvgRt**.

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth

=====
MPLS LSP (Auto Bandwidth)
=====
Legend :
+ - Inherited
=====
-----
Type : Originating
-----
LSP Name      : LSP-PE-1-PE-2
Auto BW       : Enabled                AB OpState    : Up
Auto BW Min   : 2 Mbps                 Auto BW Max   : 20 Mbps
AB Up Thresh  : 10 percent              AB Down Thresh : 5 percent
```

```

AB Up BW      : 1 Mbps
AB Curr BW   : 4 Mbps
AB Adj Mul    : 288
AB Adj Time   : 1440 Mins
AB Adj Cnt    : 0
AB Last Adj   : 03/19/2024 22:23:58
ABMaxAvgRt   : 0 Mbps
AB Ovfl Lmt   : 1
ABOvflThres  : 10 percent
AB UndflLmt   : 3
ABUndflThrs  : 10 percent
ABMaxUndflBW : 0 Mbps
AB Adj Cause : underflow
Be Weight    : 100 percent
L1 Weight    : 100 percent
Nc Weight    : 100 percent
H1 Weight    : 100 percent
Use Last Adj BW : Disabled
=====
AB Down BW    : 0 Mbps
AB Samp Intv  : 5 Mins
AB Samp Mul   : 1
AB Samp Time  : 5 Mins
AB Samp Cnt   : 0
AB Next Adj   : 1440 Mins
AB Lst AvgRt  : 4 Mbps
AB Ovfl Cnt   : 0
AB Ovfl BW    : 2 Mbps
AB Undrfl Cnt : 0
AB Undrfl BW  : 2 Mbps
AB Monitor BW : False
Af Weight     : 100 percent
L2 Weight     : 100 percent
Ef Weight     : 100 percent
H2 Weight     : 100 percent

```

If the overload or underload trigger condition is met at the end of an adjust-interval, the auto-bandwidth adjustment is normal, based on the periodic trigger. Overflow and underflow auto-bandwidth adjustments only take place when the adjust-interval is not yet completed.

## Auto-Bandwidth – Manual Trigger Type

As before, auto-bandwidth adjustment is disabled to revert to a bandwidth of 2 Mbps, as follows:

```

# On PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-2" {
        delete auto-bandwidth
      }
    }
  }
}

```

Afterward, auto-bandwidth adjustment is configured on PE-1, as follows:

```

# On PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-2" {
        auto-bandwidth {
          max-bandwidth 20
          min-bandwidth 2
          adjust-up {
            percent 10
            bw 1
          }
        }
        multipliers {
          sample-multiplier 1
          adjust-multiplier 288
        }
      }
    }
  }
}

```

```
}
}
```

The auto-bandwidth adjustment can be triggered manually at all times by the following command (with or without the keyword **force**).

```
# On PE-1:
tools perform router "Base" mpls adjust-autobandwidth
tools perform router "Base" mpls adjust-autobandwidth lsp "LSP-PE-1-PE-2"
tools perform router "Base" mpls adjust-autobandwidth lsp "LSP-PE-1-PE-2" force
```

When no specific LSP is referred to, auto-bandwidth is attempted on all LSPs. If the LSP already has the requested bandwidth, the following output is returned.

```
# On PE-1:
tools perform router "Base" mpls adjust-autobandwidth lsp "LSP-PE-1-PE-2"
MINOR: CLI #2005: Error while processing command
- No Thresholds crossed for lsp LSP-PE-1-PE-2 .
```

By adding the keyword **force**, there is no check whether the thresholds are crossed. However, the granularity is 1 Mbps. In this case, it is not possible to signal a bandwidth that is at least 1 Mbps different, so the following error message is returned.

```
# On PE-1:
tools perform router "Base" mpls adjust-autobandwidth lsp "LSP-PE-1-PE-2" force
MINOR: CLI #2005: Error while processing command
- lsp LSP-PE-1-PE-2 active path is already at the requested value 10 Mbps.
```

If the first sample interval has not yet expired, the following error message is returned.

```
# On PE-1:
tools perform router "Base" mpls adjust-autobandwidth lsp "LSP-PE-1-PE-2"
MINOR: CLI #2005: Error while processing command
- No Autobandwidth Averages computed for lsp LSP-PE-1-PE-2.
```

If the **tools** command is launched after the first sample interval has expired (**ABMaxAvgRt** is filled in), the bandwidth can be adjusted. The **AB Adj Cause** is manual.

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth

=====
MPLS LSP (Auto Bandwidth)
=====
Legend :
+ - Inherited
=====
-----
Type : Originating
-----
LSP Name   : LSP-PE-1-PE-2
Auto BW    : Enabled                AB OpState      : Up
Auto BW Min : 2 Mbps                Auto BW Max     : 20 Mbps
AB Up Thresh : 10 percent            AB Down Thresh  : 5 percent
AB Up BW    : 1 Mbps                AB Down BW     : 0 Mbps
AB Curr BW  : 10 Mbps              AB Samp Intv   : 5 Mins
AB Adj Mul  : 288                   AB Samp Mul    : 1
AB Adj Time : 1440 Mins             AB Samp Time   : 5 Mins
AB Adj Cnt  : 1                     AB Samp Cnt    : 0
```

```

AB Last Adj      : 03/19/2024 22:46:43      AB Next Adj      : 1435 Mins
ABMaxAvgRt      : 10 Mbps                  AB Lst AvgRt    : 10 Mbps
AB Ovfl Lmt     : 0                        AB Ovfl Cnt     : 0
AB0vflThres    : 0 percent                 AB Ovfl BW      : 0 Mbps
AB UndflLmt     : 0                        AB Undrfl Cnt   : 0
ABUndflThrs    : 0 percent                 AB Undrfl BW    : 0 Mbps
ABMaxUndflBW   : 0 Mbps
AB Adj Cause   : manual                  AB Monitor BW   : False
Be Weight       : 100 percent              Af Weight       : 100 percent
L1 Weight       : 100 percent              L2 Weight       : 100 percent
Nc Weight       : 100 percent              Ef Weight       : 100 percent
H1 Weight       : 100 percent              H2 Weight       : 100 percent
Use Last Adj BW : Disabled
=====

```

The counters are not reset after a manually triggered auto-bandwidth adjustment. The adjust-interval is not interrupted, the measured bandwidth and the maximum underflow bandwidth are not reset, and the overflow and underflow count are not reset.

Launch the **tools** command with the keyword **force** and a bandwidth value. This will set the current bandwidth to this value, even if the value is not within the allowed range between the minimum and maximum bandwidth.

```

# On PE-1:
tools perform router "Base" mpls adjust-autobandwidth lsp "LSP-PE-1-PE-2" force bandwidth 30

```

```

[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth

=====
MPLS LSP (Auto Bandwidth)
=====
Legend :
+ - Inherited
=====
-----
Type : Originating
-----
LSP Name      : LSP-PE-1-PE-2
Auto BW       : Enabled                AB OpState    : Up
Auto BW Min   : 2 Mbps                 Auto BW Max   : 20 Mbps
AB Up Thresh  : 10 percent              AB Down Thresh : 5 percent
AB Up BW      : 1 Mbps                 AB Down BW    : 0 Mbps
AB Curr BW   : 30 Mbps                 AB Samp Intv  : 5 Mins
AB Adj Mul    : 288                    AB Samp Mul   : 1
AB Adj Time   : 1440 Mins               AB Samp Time  : 5 Mins
AB Adj Cnt    : 0                      AB Samp Cnt   : 0
AB Last Adj   : 03/19/2024 22:52:49    AB Next Adj   : 1440 Mins
ABMaxAvgRt   : 0 Mbps                  AB Lst AvgRt  : 0 Mbps
AB Ovfl Lmt   : 0                      AB Ovfl Cnt   : 0
AB0vflThres  : 0 percent                 AB Ovfl BW    : 0 Mbps
AB UndflLmt   : 0                      AB Undrfl Cnt : 0
ABUndflThrs  : 0 percent                 AB Undrfl BW  : 0 Mbps
ABMaxUndflBW : 0 Mbps
AB Adj Cause : manual                  AB Monitor BW : False
Be Weight     : 100 percent              Af Weight     : 100 percent
L1 Weight     : 100 percent              L2 Weight     : 100 percent
Nc Weight     : 100 percent              Ef Weight     : 100 percent
H1 Weight     : 100 percent              H2 Weight     : 100 percent
Use Last Adj BW : Disabled
=====

```

Manually triggered auto-bandwidth adjustments are also performed using MBB procedures.

## Auto-Bandwidth Adjustment Based on Forwarding Class Subset

With the configuration applied so far, there is no distinction between traffic from different forwarding classes (FCs). The average data rate is the sum of the traffic from all eight FCs. From SR OS Release 11.0.R4 onward, it is possible to provide a sampling weight for each Forwarding Class (FC) for each auto-bandwidth LSP. The average data rate is now the weighted sum of the traffic from all FCs.

```
# On PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-2" {
        auto-bandwidth {
          fc be {
            sampling-weight 50
          }
          fc af {
            sampling-weight 80
          }
        }
      }
    }
  }
}
```

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth

=====
MPLS LSP (Auto Bandwidth)
=====
Legend :
+ - Inherited
=====
-----
Type : Originating
-----
LSP Name   : LSP-PE-1-PE-2
Auto BW    : Enabled
Auto BW Min : 2 Mbps
AB Up Thresh : 10 percent
AB Up BW   : 1 Mbps
AB Curr BW : 30 Mbps
AB Adj Mul : 288
AB Adj Time : 1440 Mins
AB Adj Cnt : 0
AB Last Adj : 03/19/2024 22:52:49
ABMaxAvgRt : 0 Mbps
AB Ovfl Lmt : 0
ABOvflThres : 0 percent
AB UndflLmt : 0
ABUndflThrs : 0 percent
ABMaxUndflBW : 0 Mbps
AB Adj Cause : manual
Be Weight : 50 percent
L1 Weight : 100 percent
Nc Weight : 100 percent
H1 Weight : 100 percent
AB OpState : Up
Auto BW Max : 20 Mbps
AB Down Thresh : 5 percent
AB Down BW : 0 Mbps
AB Samp Intv : 5 Mins
AB Samp Mul : 1
AB Samp Time : 5 Mins
AB Samp Cnt : 0
AB Next Adj : 1440 Mins
AB Lst AvgRt : 0 Mbps
AB Ovfl Cnt : 0
AB Ovfl BW : 0 Mbps
AB Undrfl Cnt : 0
AB Undrfl BW : 0 Mbps
AB Monitor BW : False
Af Weight : 80 percent
L2 Weight : 100 percent
Ef Weight : 100 percent
H2 Weight : 100 percent
```

```
Use Last Adj BW : Disabled
=====
```

The sampling-weight values can be changed while auto-bandwidth is enabled. The auto-bandwidth algorithm is reset on the LSP at the end of the current collection interval. At that time, the current bandwidth will not be adjusted and the following counters are reset to 0: sample count, adjust count, overflow count, underflow count, max average data rate, and max average underflow data rate.

## Auto-Bandwidth on LSPs with Secondary Paths

When the active path goes down or becomes degraded, the bandwidth used to signal the auto-bandwidth MBB is the operational bandwidth of the previous active path. The parameter `current-bandwidth` requires a modified definition:

**current-bandwidth** — The last known reserved bandwidth for the LSP (this may be for a different path than the active one).

When the active path changes, the current bandwidth is updated to the operational bandwidth of the new active path. While the auto-bandwidth MBB on the active path is in progress, a statistics sample might be triggered because the intervals aren't reset when the active path changes. It is possible that an auto-adjustment is needed. The in-progress auto-bandwidth MBB is restarted with retry attempts to 0 and signaled bandwidth equal to the new measured bandwidth. If after five attempts, auto-bandwidth MBB fails, the current bandwidth and secondary **oper-bw** remain unchanged.

For a secondary/standby path, if the active path changes without the LSP going down, an auto-bandwidth MBB is triggered for the new active path. The bandwidth used to signal the MBB is the operational bandwidth of the previous active path (current bandwidth).

If the primary path is not currently active, but it has not gone down, then any MBB should use the configured bandwidth for the primary path.

Create two new strict paths and assign them to the LSP. The primary path is the direct strict path from PE-1 to PE-2. There are two secondary paths: *path-PE-1-PE-3-PE-2* and *loose*. The first one is standby, the latter is not.

```
# On PE-1:
configure {
  router "Base" {
    mpls {
      path "path-PE-1-PE-2" {
        admin-state enable
        hop 10 {
          ip-address 192.0.2.2
          type strict
        }
      }
      path "path-PE-1-PE-3-PE-2" {
        admin-state enable
        hop 10 {
          ip-address 192.0.2.3
          type strict
        }
        hop 20 {
          ip-address 192.0.2.2
          type strict
        }
      }
    }
  }
  lsp "LSP-PE-1-PE-2" {
```







```
port 1/1/c1/1 {
    admin-state disable
}
}
```

Verify that the secondary/standby path is now active.

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-2" activepath

=====
MPLS LSP: LSP-PE-1-PE-2 (active paths)
=====
Legend :
# - Manually switched path
#F - Manually forced switched path
=====
LSP Name      : LSP-PE-1-PE-2
LSP Id        : 10790
Path Name     : path-PE-1-PE-3-PE-2
Active Path   : Standby
To            : 192.0.2.2
LSP Type      : dynamic
=====
```

Check the auto-bandwidth data on the LSP. The current bandwidth for the LSP is the same as it used to be for the primary path. **AB Adj Cause** = activePathChange.

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth

=====
MPLS LSP (Auto Bandwidth)
=====
Legend :
+ - Inherited
=====
-----
Type : Originating
-----
LSP Name      : LSP-PE-1-PE-2
Auto BW       : Enabled                AB OpState    : Up
Auto BW Min   : 2 Mbps                 Auto BW Max   : 20 Mbps
AB Up Thresh  : 10 percent              AB Down Thresh : 5 percent
AB Up BW      : 1 Mbps                  AB Down BW    : 0 Mbps
AB Curr BW    : 5 Mbps                  AB Samp Intv  : 5 Mins
AB Adj Mul    : 288                     AB Samp Mul   : 1
AB Adj Time   : 1440 Mins               AB Samp Time  : 5 Mins
AB Adj Cnt    : 0                       AB Samp Cnt   : 0
AB Last Adj   : 03/19/2024 23:09:25    AB Next Adj   : 1440 Mins
ABMaxAvgRt   : 0 Mbps                  AB Lst AvgRt  : 5 Mbps
AB Ovfl Lmt   : 2                       AB Ovfl Cnt   : 0
ABOvflThres  : 10 percent              AB Ovfl BW    : 0 Mbps
AB UndflLmt   : 3                       AB Undrfl Cnt : 0
ABUndflThrs  : 10 percent              AB Undrfl BW  : 0 Mbps
ABMaxUndflBW : 0 Mbps
AB Adj Cause  : activePathChange        AB Monitor BW : False
Be Weight     : 50 percent              Af Weight     : 80 percent
L1 Weight     : 100 percent             L2 Weight     : 100 percent
Nc Weight     : 100 percent             Ef Weight     : 100 percent
H1 Weight     : 100 percent             H2 Weight     : 100 percent
```

```
Use Last Adj BW : Disabled
=====
```

The original situation is restored.

```
# On PE-1:
configure {
  port 1/1/c1/1 {
    admin-state enable
  }
}
```

The primary path comes up again.

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-2" activepath
```

```
=====
MPLS LSP: LSP-PE-1-PE-2 (active paths)
=====
```

```
Legend :
# - Manually switched path
#F - Manually forced switched path
=====
```

```
LSP Name      : LSP-PE-1-PE-2
LSP Id        : 10794
Path Name     : path-PE-1-PE-2
Active Path   : Primary
To            : 192.0.2.2
LSP Type     : dynamic
=====
```

The auto-bandwidth data again shows **AB Adj Cause**: activePathChange, but with a different timestamp.

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-2" auto-bandwidth
```

```
=====
MPLS LSP (Auto Bandwidth)
=====
```

```
Legend :
+ - Inherited
=====
```

```
Type : Originating
-----
```

```
LSP Name      : LSP-PE-1-PE-2
Auto BW       : Enabled                AB OpState    : Up
Auto BW Min   : 2 Mbps                 Auto BW Max   : 20 Mbps
AB Up Thresh  : 10 percent              AB Down Thresh : 5 percent
AB Up BW      : 1 Mbps                 AB Down BW    : 0 Mbps
AB Curr BW    : 5 Mbps                 AB Samp Intv  : 5 Mins
AB Adj Mul    : 288                    AB Samp Mul   : 1
AB Adj Time   : 1440 Mins              AB Samp Time  : 5 Mins
AB Adj Cnt    : 0                      AB Samp Cnt   : 0
AB Last Adj   : 03/19/2024 23:12:25    AB Next Adj   : 1440 Mins
ABMaxAvgRt   : 0 Mbps                  AB Lst AvgRt  : 5 Mbps
AB Ovfl Lmt   : 2                      AB Ovfl Cnt   : 0
ABOvflThres  : 10 percent              AB Ovfl BW    : 0 Mbps
AB UndflLmt  : 3                       AB Undrfl Cnt : 0
ABUndflThrs  : 10 percent              AB Undrfl BW  : 0 Mbps
```

```
ABMaxUndflBW      : 0 Mbps
AB Adj Cause      : activePathChange
Be Weight         : 50 percent
L1 Weight         : 100 percent
Nc Weight         : 100 percent
H1 Weight         : 100 percent
Use Last Adj BW   : Disabled
AB Monitor BW     : False
Af Weight         : 80 percent
L2 Weight         : 100 percent
Ef Weight         : 100 percent
H2 Weight         : 100 percent
=====
```

## Conclusion

Auto-bandwidth adjustment can be enabled on point-to-point LSPs in order to make a realistic bandwidth reservation, based on active iLER traffic monitoring. A user has control over how the bytes count for the different FCs by providing a sampling-weight factor. This can influence the average data rate over the sample interval.

The bandwidth is taken into account in the control plane when LSPs are established or when they change their bandwidth using MBB. The bandwidth in the data plane is not restricted by this setting.

Auto-bandwidth adjustment can be triggered in different ways: periodically, in case of overflow/underflow, manually, and in case of an active path change. It is also possible to have passive monitoring where no adjustment is done.

# Automatic Creation of RSVP-TE LSPs

This chapter provides information about automatic creation of Resource Reservation Protocol with Traffic Engineering (RSVP-TE) Label Switched Paths (LSPs).

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

## Applicability

This feature is applicable to SR OS with no hardware constraints because this is a control-plane feature only.

This chapter was originally written for SR OS Release 11.0.R6, but the configuration in this edition corresponds to MD-CLI in SR OS Release 21.2.R1.

## Overview

Automatic creation of RSVP-TE LSPs enables the automated creation of point-to-point RSVP-TE LSPs within a single Interior Gateway Protocol (IGP) Intermediate System to Intermediate System (IS-IS) level or Open Shortest Path First (OSPF) area that can subsequently be used by services and/or IGP shortcuts. The feature is divided into two components: creation of an RSVP-TE LSP mesh, and creation of single-hop RSVP-TE LSPs. Although both can be used simultaneously, it is likely that one or the other is used.

When creating an RSVP-TE LSP mesh, the mesh can be full or partial, the extent of which is governed by a prefix list containing the system addresses of all nodes that should form part of the mesh. When using single-hop RSVP-TE LSPs, point-to-point LSPs are established to all directly connected neighbors. The purpose of these single-hop LSPs is to allow for Equal Cost Multi-Path (ECMP) load balancing of traffic using LDP over RSVP, which is not possible using native RSVP LSPs.

The use of automatically created RSVP-TE LSPs avoids manual configuration of RSVP-TE LSP meshes. Even when provisioning tools—such as 5620 SAM—are used to automatically provision these LSPs, auto-mesh still provides a benefit by avoiding increased configuration file sizes.

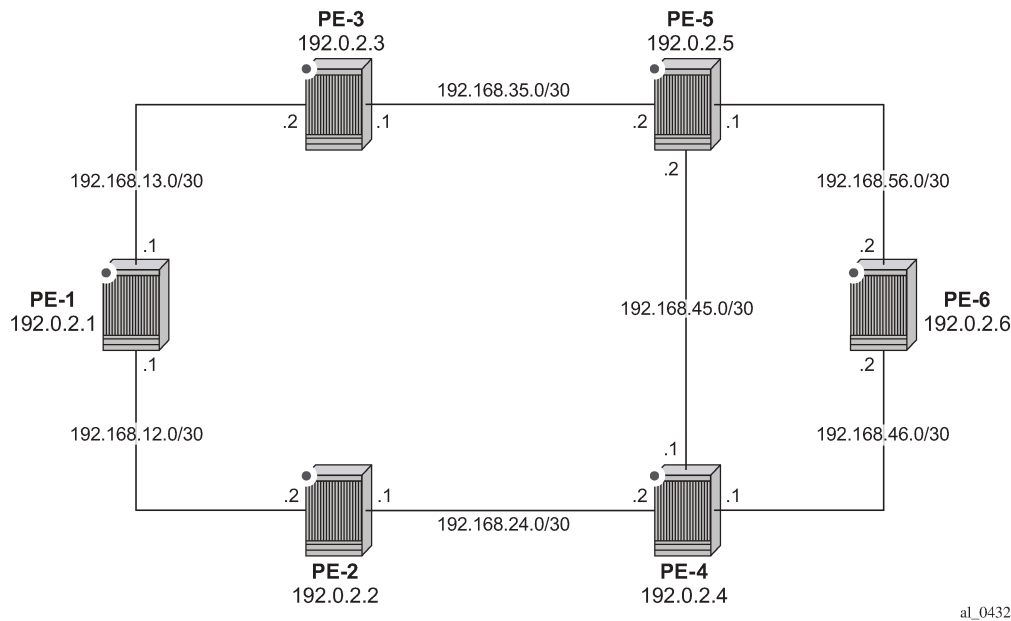
The use of automatically created Targeted Label Distribution Protocol (T-LDP) sessions is also described when using the automatically created RSVP LSPs for Layer 2 services.

## Configuration

### Example Topology

The example topology is shown in [Figure 4: Example Topology](#). All routers participate in a single IS-IS Level 2 area that has traffic engineering enabled. Multi-Protocol Label Switching (MPLS) and RSVP are enabled on every interface, but no LSPs are initially provisioned. All routers are Border Gateway Protocol (BGP) speakers and form part of Autonomous System (AS) 64496. PE-5 is a Route Reflector and the remaining routers are IBGP clients for the IPv4, VPN-IPv4, and L2-VPN address families. The objective of this example is to demonstrate how to automatically create transport LSPs using RSVP or LDP over RSVP, and then create services that utilize those LSPs. The exchange of BGP routes is needed for those services.

Figure 4: Example Topology



aL\_0432

### Automatic Creation of an RSVP-TE LSP Mesh

To start the process of automatically creating an RSVP-TE LSP mesh, the user must create a route policy referencing a prefix-list. This prefix-list contains the system addresses of all nodes that are required to be in the mesh, and can be entered as a series of /32 addresses, or simply as a range as follows. This range encompasses all of the system addresses of the nodes in the example topology because the requirement is to make a full mesh.

```
configure {
  policy-options {
    prefix-list "System-Addresses" {
      prefix 192.0.2.0/24 type range {
        start-length 32
      }
    }
  }
}
```

```

        end-length 32
    }
}
policy-statement "Remote-PEs-policy" {
    entry 10 {
        from {
            prefix-list ["System-Addresses"]
        }
        action {
            action-type accept
        }
    }
}
}}

```

After the route policy is created, the user must create an LSP template containing the common parameters which are used to establish all point-to-point LSPs within the mesh. For an RSVP-TE LSP mesh, the **lsp-template** must be configured with type **p2p-rsvp-mesh**. Upon creation of the template, CSPF is automatically enabled (and cannot be disabled), and the template must reference a **default-path** before it can be enabled. In the example contained in the following output, the template refers to a path named "loose-path" that has no strict or loose hops defined, meaning the system will dynamically calculate the path while considering other specified constraints. The LSP template in this output also stipulates frr-method facility bypass protection. The default behavior is no node-protect, so this configuration requests link protection only. FRR one-to-one protection is not supported for automatically created RSVP-TE LSPs; so facility bypass is the only form of protection supported.

Next, the user must associate the LSP template with the previously defined route policy, and this is accomplished using the **auto-lsp** command. In this example, the LSP template "Full-Mesh-template" is associated with the policy "Remote-PEs-policy" that in turn references a prefix-list containing all system addresses in the example topology. Up to five policies can be associated with an LSP template at the same time. If a policy associated with an LSP template is modified in order to add or remove prefixes, the system immediately re-evaluates the policy and the prefix-list to determine if one or more LSPs need to be established, or one or more LSPs need to be torn down.

```

configure {
    router "Base" {
        mpls {
            admin-state enable
            path "loose-path" {
                admin-state enable
            }
            lsp-template "Full-Mesh-template" {
                admin-state enable
                type p2p-rsvp-mesh
                default-path "loose-path"
                path-computation-method local-cspf
                fast-reroute {
                    frr-method facility
                }
            }
            auto-lsp "Full-Mesh-template" {
                policy ["Remote-PEs-policy"]
            }
        }
    }
}

```

When the **auto-lsp** command is entered, the system commences the process of establishing the point-to-point LSPs. The prefixes defined in the prefix list are checked, and if a prefix corresponds to a router

ID that is present in the Traffic Engineering Database (TED), the system instantiates a CSPF computed primary path to that prefix using the parameters specified in the LSP template. With the previously defined configuration applied on PE-6, the existence of point-to-point RSVP LSPs to every node in the example topology can be verified as shown in the following output. The LSP name is automatically constructed as TemplateName-DestIPv4Address-TunnelId. The LSP name signaled in the Session Attribute object concatenates the LSP name with the path name (for example Full-Mesh-template-192.0.2.1-61441::loose-path).

```
[/]
A:admin@PE-6# show router mpls lsp

=====
MPLS LSPs (Originating)
=====
LSP Name          Tun   Fastfail  Adm  Opr
  To              Id     Config
-----
Full-Mesh-template-192.0.2.1-61441
192.0.2.1         61441   Yes      Up   Up
Full-Mesh-template-192.0.2.2-61442
192.0.2.2         61442   Yes      Up   Up
Full-Mesh-template-192.0.2.3-61443
192.0.2.3         61443   Yes      Up   Up
Full-Mesh-template-192.0.2.4-61444
192.0.2.4         61444   Yes      Up   Up
Full-Mesh-template-192.0.2.5-61445
192.0.2.5         61445   Yes      Up   Up
-----
LSPs : 5
=====
```

The following output on PE-6 shows that the auto-created LSP to PE-1 has primary path "loose-path" and is operationally up:

```
[/]
A:admin@PE-6# show router mpls lsp "Full-Mesh-template-192.0.2.1-61441" path detail | match
"LSP Name" post-lines 5
LSP Name      : Full-Mesh-template-192.0.2.1-61441
From          : 192.0.2.6
To            : 192.0.2.1
Admin State   : Up                               Oper State    : Up
Path Name     : loose-path
Path LSP ID   : 60928                             Path Type     : Primary
```

The LSP template requests FRR link protection. On PE-6, this protection can be verified by querying each primary LSP. In the following output, the primary LSP to PE-1 (Full-Mesh-template-192.0.2.1-61441) is signaled through PE-5 (192.0.2.5) and PE-3 (192.0.2.3), and the presence of the @ indicator after each hop denotes that link protection is available to the primary path.

```
[/]
A:admin@PE-6# show router mpls lsp "Full-Mesh-template-192.0.2.1-61441" path detail | match
"Actual Hops" post-lines 4
Actual Hops   :
  192.168.56.2(192.0.2.6) @           Record Label  : N/A
-> 192.168.56.1(192.0.2.5) @         Record Label  : 524270
-> 192.168.35.1(192.0.2.3) @        Record Label  : 524274
-> 192.168.13.1(192.0.2.1) @        Record Label  : 524275
```

Finally, it can be verified that the signaled LSPs are placed in the tunnel table and made available to the tunnel table manager so they can be used by applications and services.

```
[/]
A:admin@PE-6# show router tunnel-table

=====
IPv4 Tunnel Table (Router: Base)
=====
Destination          Owner    Encap TunnelId  Pref  Nexthop        Metric
  Color
-----
192.0.2.1/32 [B]    rsvp    MPLS  61441    7    192.168.56.1   30
192.0.2.2/32 [B]    rsvp    MPLS  61442    7    192.168.46.1   20
192.0.2.3/32 [B]    rsvp    MPLS  61443    7    192.168.56.1   20
192.0.2.4/32 [B]    rsvp    MPLS  61444    7    192.168.46.1   10
192.0.2.5/32 [B]    rsvp    MPLS  61445    7    192.168.56.1   10
-----
Flags: B = BGP or MPLS backup hop available
      L = Loop-Free Alternate (LFA) hop available
      E = Inactive best-external BGP route
      k = RIB-API or Forwarding Policy backup hop
=====
```

When the LSP template is in use and LSPs are instantiated, it is necessary to disable the template to change any parameters that cannot be handled as a Make-Before-Break (MBB). This essentially includes all LSP parameters with the exception of bandwidth and FRR without node-protection. Modification of any other parameters requires disabling the LSP template and a re-signal of the LSP once the LSP template is re-enabled. MBB is supported for timer-based and manual re-signaling of the automatically created LSPs.

## Service and Application Verification

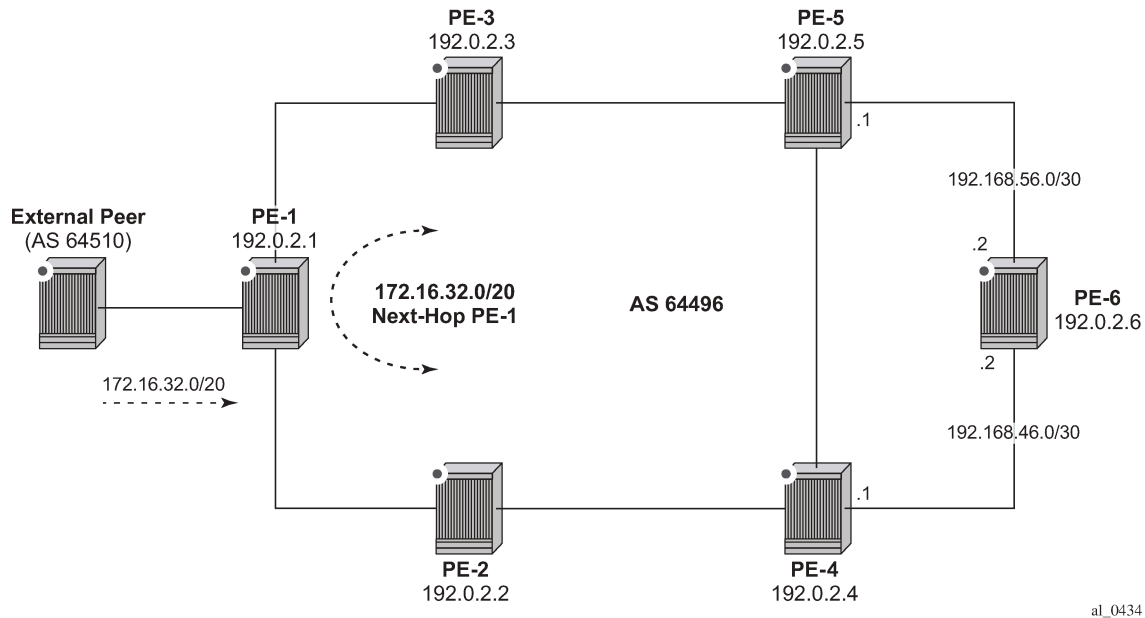
With the RSVP-TE LSP mesh in place, it is now possible to create services and applications to utilize those LSPs. These applications and services include Layer 2 and Layer 3 VPNs, resolution of BGP labeled routes and resolution of BGP, IGP, and static routes. However, the automatically created LSPs are not available for explicit binding in a statically provisioned service destination point (SDP).

## IGP Shortcuts

[Figure 5: IGP Shortcuts with RSVP-TE Auto-Mesh](#) demonstrates the use of IGP shortcuts. Prefix 172.16.32.0/20 is advertised to PE-1 from an external peer in AS 64510, which PE-1 subsequently advertises into IBGP, imposing Next-Hop-Self in the process. For more details on IGP shortcuts, see the [IGP Shortcuts](#) chapter.



Figure 5: IGP Shortcuts with RSVP-TE Auto-Mesh



The objective is for PE-6 to use the automatically created LSP to PE-1 as an IGP shortcut (typically implemented in order to maintain a "BGP-free" core). IGP shortcuts for BGP are enabled under the main **bgp** context using the command **next-hop-resolution shortcut-tunnel** with options for **rsvp**, **ldp**, or **bgp**. Because the example topology only has (automatically created) RSVP-TE LSPs, this option is selected. Besides **rsvp**, **ldp**, and **bgp**, there are other options, but these are beyond the scope of this chapter.

```
[/]
A:admin@PE-6# configure {
  router "Base" {
    bgp {
      next-hop-resolution {
        shortcut-tunnel {
          family ipv4 {
            resolution-filter ?
          }
        }
      }
    }
  }
}

resolution-filter

bgp          - Use BGP tunneling for next-hop resolution
ldp          - Use LDP tunneling for next-hop resolution
mpls-fwd-policy - Use MPLS forwarding policy for next-hop resolution
rib-api      - Use RIB API gRPC service for next-hop resolution
rsvp        - Use RSVP tunneling for next-hop resolution
sr-isis     - Use IS-IS SR tunneling for next-hop resolution
sr-ospf     - Use OSPF SR tunneling for next-hop resolution
sr-ospf3    - Use OSPFv3 SR tunneling for next-hop resolution
sr-policy   - Use SR policies for next-hop resolution
sr-te       - Use SR-TE
```

On PE-6, the following BGP next -hop resolution configuration defines that IGP shortcuts are resolved using RSVP:

```
[/]
A:admin@PE-6# configure {
```

```

router "Base" {
  bgp {
    next-hop-resolution {
      shortcut-tunnel {
        family ipv4 {
          resolution filter
          resolution-filter {
            rsvp true
          }
        }
      }
    }
  }
}

```

When the shortcuts are enabled, the route-table (and FIB) can be validated to ensure that the programmed next hop is the advertising BGP speaker (as opposed to the IGP next hop), and that traffic is tunneled to that next hop through an RSVP LSP. In this case, the RSVP LSP is the LSP with tunnel ID 61441, which is the LSP to PE-1.

```

[/]
A:admin@PE-6# show router route-table 172.16.32.0/20
=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]                Type  Proto  Age      Pref
  Next Hop[Interface Name]                Metric
-----
172.16.32.0/20                    Remote BGP    00h01m11s 170
  192.0.2.1 (tunneled:RSVP:61441)        30
-----
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
=====

```

### Layer 3 VPN

Layer 3 VPNs can utilize the automatically created LSPs by using the **auto-bind-tunnel** feature configured with the **rsvp** option (possibly in combination with LDP). A Virtual Private Routed Network (VPRN) is configured on PE-1 and PE-6. PE-1 is configured with a loopback address of 172.16.1.1/24 and advertises the VPN-IPv4 prefix 172.16.1.0/24 into IBGP, while PE-6 is configured with a loopback address of 172.16.6.1/24 and advertises the VPN-IPv4 prefix 172.16.6.0/24 into IBGP. The VPRN configuration on PE-6 is as follows. The only difference on PE-1 is the IP address assigned to the loopback interface.

```

[/]
A:admin@PE-6# configure {
  service {
    vprn "VPRN1"
    admin-state enable
    service-id 1
    customer "1"
    bgp-ipvpn {
      mpls {

```



requirement to adopt a model within the service construct that permits automatic creation of SDP bindings, and this is achieved using a pseudowire template dictating the characteristics of the SDP. The secondary effect of using pseudowire templates to dynamically create SDPs is that these automatically created SDPs can currently only use LDP or BGP as a transport tunnel, not RSVP. The solution is to enable LDP over RSVP.

This can be implemented using static provisioning of peers as shown in the next output, or it can be done using automatic creation of T-LDP sessions. Regardless of the method, a reciprocal configuration must exist at both peer endpoints. The static per-peer configuration is applied in the **targeted-session** context specifying the remote peer system IP address, and the keyword **tunneling**, which enables tunneling of LDP FECs over RSVP LSPs with a far-end address matching that of the T-LDP peer. At a global level, the **prefer-tunnel-in-tunnel false** command is shown, but is only required when a next hop router advertises a FEC over link-level LDP and T-LDP. In this case, by default, the system would prefer the link-level LDP tunnel, so the **prefer-tunnel-in-tunnel false** instructs the system to prefer an LDP over RSVP tunnel if it is available. Although link-layer LDP is not present in the example topology, the command is included because the presence of link-layer LDP is common.

```
[/]
A:admin@PE-6# configure {
  router "Base" {
    ldp {
      prefer-tunnel-in-tunnel false
      targeted-session {
        peer 192.0.2.1 {
          tunneling
        }
      }
    }
  }
}
```

The following output provides an example demonstrating the automatic creation of T-LDP sessions. No explicit reference is made to specific peers, but rather a **peer-template** is configured containing the parameters which apply to all T-LDP sessions spawned by this template. In this example, only the **tunneling** command is required. A **peer-template-map** is then used to create a mapping between the peer-template (TLDP-Mesh-template) and a **policy** defining the IP addresses of remote nodes to which T-LDP sessions should be established. In this example, the policy "Remote-PEs-policy" is the same policy previously used by the auto-created RSVP LSP mesh.

```
[/]
A:admin@PE-6# configure {
  router "Base" {
    ldp {
      prefer-tunnel-in-tunnel false
      targeted-session {
        peer-template "TLDP-Mesh-template" {
          admin-state enable
          tunneling true
        }
        peer-template-map "TLDP-Mesh-template" {
          policy-map ["Remote-PEs-policy"]
        }
      }
    }
  }
}
```

Regardless of whether T-LDP sessions are explicitly provisioned, or dynamically created using a peer-template, the result is that a targeted LDP session is established which can be used for advertising address and service FECs, and which is capable of tunneling LDP over RSVP.

```
[/]
A:admin@PE-6# show router ldp targ-peer 192.0.2.1 detail

=====
LDP IPv4 Targeted Peers
=====
-----
192.0.2.1
-----
Admin State      : Up          Oper State      : Up
Last Oper Chg   : 0d 00:01:11
Hold Time       : 45          Hello Factor    : 3
Oper Hold Time  : 45
Hello Reduction : Disabled    Hello Reduction Fctr : 3
Keepalive Timeout : 40       Keepalive Factor : 4
Active Adjacencies : 1      Last Modified   : Never
Auto Created    : Yes
Creator         : template   Template Name    : TLDP-Mesh-template
Tunneling       : Enabled
Lsp Name        : None
Mcast-Tunneling : Disabled
Lsp Name        : None
Local LSR       : None       32-BitLocalLsr  : Disabled
Local-LSR ID adv. : Disabled
Community       :
BFDD Status     : Disabled

=====
No. of IPv4 Targeted Peers: 1
=====
```

To create VPLS services using dynamically-created SDPs, BGP Auto-Discovery (BGP-AD) must be used together with LDP (or BGP) pseudowire signaling, for more details see the LDP VPLS Using BGP-Auto Discovery chapter.

In the following output, PE-6 uses BGP-AD and LDP signaling. The same configuration is applied on PE-1. The `vpls-id` is configured in the `bgp-ad` context. The `vpls-id` is a network-wide identifier assigned to all VPLS Switch Instances (VSIs) belonging to the same VPLS, and is carried in VPLS Network Layer Reachability Information (NLRI) as an extended community attribute. A second parameter used for BGP-AD and carried in the VPLS NLRI is the VSI-ID, which uniquely identifies each VSI. The VSI-ID is automatically derived from the global ASN, the VPLS service ID, and the system IP address. To automatically create SDPs, the `bgp` context of the VPLS service refers to a `pw-template` defining the parameters of the pseudowire. In this example, the use of the hash (entropy) label is enabled in the pseudowire template, and a `split-horizon-group`, SHG, is applied.

```
[/]
A:admin@PE-6# configure {
  service {
    pw-template "PW2-template" {
      pw-template-id 2
      hash-label {
      }
      split-horizon-group {
        name "SHG"
      }
    }
  }
  vpls "VPLS-2" {
```

```

        admin-state enable
        service-id 2
        customer "1"
        bgp 1 {
            pw-template-binding "PW2-template" {
            }
        }
        bgp-ad {
            admin-state enable
            vpls-id "64496:2"
        }
        sap 1/1/4:2 {
        }
    }
}

```

The following output shows the BGP operational parameters for vpls 2:

```

[/]
A:admin@PE-6# show service id 2 bgp
=====
BGP Information
=====
Bgp Instance           : 1
Vsi-Import             : None
Vsi-Export             : None
Route Dist             : None
Oper Route Dist        : 64496:2
Oper RD Type           : derivedVpls
Rte-Target Import     : None           Rte-Target Export: None
Oper RT Imp Origin    : derivedVpls    Oper RT Import   : 64496:2
Oper RT Exp Origin    : derivedVpls    Oper RT Export   : 64496:2

PW-Template Id        : 2             PW-Template SHG  : None
Oper Group            : None
Mon Oper Group        : None
BFD Template         : None
BFD-Enabled          : no             BFD-Encap       : ipv4
Import Rte-Tgt       : None
=====

```

The BGP-AD operational parameters for VPLS 2 are as follows:

```

[/]
A:admin@PE-6# show service id 2 bgp-ad
-----
BGP Auto-discovery Information
-----
Admin State           : Up
Vpls Id               : 64496:2
Prefix                : 192.0.2.6
-----

```

The following shows that an SDP with ID 32767:4294967295 of type BgpAd has been automatically created. Both the SDP and the SAP are operationally up.

```

[/]
A:admin@PE-6# show service id 2 base | match "Service Access" post-lines 10

```

Service Access & Destination Points

Identifier	Type	AdmMTU	OprMTU	Adm	Opr
sap:1/1/4:2	q-tag	1518	1518	Up	Up
<b>sdp:32767:4294967295 SB(192.0.2.1)</b>	<b>BgpAd</b>	0	1548	Up	Up

\* indicates that the corresponding row element may have been truncated.

To create Epipe services using dynamically created SDPs, two options exist. Either LDP FEC 129 signaling can be used, which in turn dictates the presence of pseudowire routing information, or BGP-VPWS based signaling can be used, for more details, see the "BGP Virtual Private Wire Services" chapter in the *7450 ESS, 7750 SR, 7950 XRS, and VSR Layer 2 Services and EVPN Advanced Configuration Guide for Classic CLI*. This example illustrates the use of BGP VPWS, but in either case, only single-segment pseudowires are supported. The following output shows the configuration requirements for a basic BGP-based Epipe service on PE-6. Once again, a **pw-template** is used to define the characteristics of the pseudowire, and this template is referenced in the **bgp** context of the Epipe service. The **bgp** context is also where the **route-distinguisher** and **route-target** values are configured, which are carried in the VPWS NLRI and extended communities respectively. The **ve-name**, **ve-id**, and **remote-ve-name** are all configured in the **bgp-vpws** context. The **ve-id** is carried in the VPWS NLRI, and when a PE router receives a VPWS NLRI to try to establish an Epipe service, the **ve-id** from the NLRI is validated against the **ve-id** configured in the **remote-ve-name**. These must match before the Epipe becomes operational.

```
[/]
A:admin@PE-6# configure {
  service {
    pw-template "PW3-template" {
      pw-template-id 3
      hash-label {
      }
    }
  }
  epipe "Epipe-3" {
    admin-state enable
    service-id 3
    customer "1"
    bgp 1 {
      route-distinguisher "64496:3"
      route-target {
        export "target:64496:3"
        import "target:64496:3"
      }
      pw-template-binding "PW3-template" {
      }
    }
  }
  bgp-vpws {
    admin-state enable
    local-ve {
      name "PE-6"
      id 6
    }
    remote-ve "PE-1" {
      id 1
    }
  }
  sap 1/1/4:3 {
  }
}
}
```

The basic service information is truncated to show only the relevant information in order to verify that the service is operational. SDP (32766:4294967294) has been automatically created and is of type **BgpVpws**. Both the SDP and the SAP are operationally up.

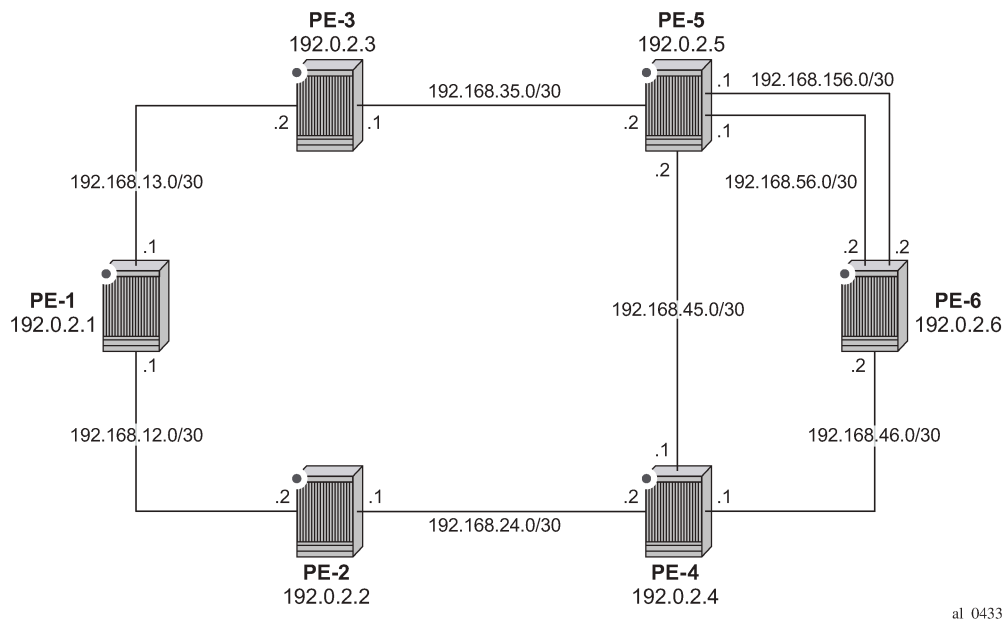
```
[/]
A:admin@PE-6# show service id 3 base | match "Service Access" post-lines 10
Service Access & Destination Points
-----
Identifier                                     Type           AdmMTU  OprMTU  Adm  Opr
-----
sap:1/1/4:3                                  q-tag          1518    1518    Up   Up
sdp:32766:4294967294 SB(192.0.2.1)          BgpVpws      0       1548    Up   Up
=====
```

### Automatic Creation of RSVP Single-Hop LSPs

As previously discussed, the purpose of a single-hop LSP mesh is to allow for ECMP load balancing of traffic using LDP over RSVP. ECMP load balancing could be implemented using LDP over a partial or full mesh of RSVP-TE LSPs, but the use of single-hop LSPs additionally allows for load balancing across a number of parallel RSVP LSPs between nodes. To illustrate ECMP load balancing over multiple parallel RSVP LSPs, the example topology of [Figure 4: Example Topology](#) is modified to include a parallel link between PE-6 and PE-5 as shown in [Figure 6: Example Topology for Single-Hop LDP over RSVP with ECMP](#). In addition, all routers are enabled for ECMP=2, as follows.

```
[/]
configure {
  router "Base" {
    ecmp 2
  }
}
```

Figure 6: Example Topology for Single-Hop LDP over RSVP with ECMP





Unlike the automatically created RSVP-TE LSP mesh previously described, the automatically created single-hop RSVP-TE LSPs have no requirement for a prefix-list to be referenced containing the prefixes of the remote nodes that form part of the mesh. In the case of automatically created single-hop LSPs, the TE database keeps track of each TE link which comes up to a directly connected IGP neighbor. The system then establishes a single-hop LSP with a destination address matching the router ID of the neighbor and with a strict hop consisting of the address of the interface used by the TE link.

The first requirement is to create an LSP template containing the common parameters used to establish each single-hop LSP. For a single-hop LSP mesh, the **lsp-template** must be configured with type **p2p-one-hop**. Upon creation of the template, **cspf** is automatically enabled (and cannot be disabled), and the **hop-limit** is set to a value of **2**. The hop limit defines the number of nodes the LSP may traverse, and, because these are single-hop LSPs to adjacent neighbors, a limit of 2 is sufficient. The template must also reference a **default-path** before it can be enabled. The following example references a path named "loose-path" that has no strict or loose hops defined. When the RSVP PATH message is actually generated to create the one-hop LSP, it contains one strict-hop to the interface address of the neighbor; and as destination the system address of the adjacent node.

The next requirement is to trigger the creation of single-hop LSPs, and this is achieved using the **auto-lsp** command. In this example, the LSP template "Single-Hop-template" is referenced, and the command is completed with the keyword one-hop to indicate the creation of single-hop LSPs. Unlike an RSVP-TE mesh, there is no requirement to reference a route policy. In the example, the auto-lsp "Full-Mesh-template" and LSP template "Full-Mesh-template" are removed on all PEs.

```
[/]
A:admin@a PE># configure {
  router "Base" {
    mpls {
      delete auto-lsp "Full-Mesh-template"
      delete lsp-template "Full-Mesh-template"
    }
  }
}
```

The following one-hop LSP template is created on all nodes:

```
[/]
A:admin@a PE># configure {
  router "Base" {
    mpls {
      admin-state enable
      path "loose-path" {
        admin-state enable
      }
      lsp-template "Single-Hop-template" {
        admin-state enable
        type p2p-rsvp-one-hop
        default-path "loose-path"
        hop-limit 2
        path-computation-method local-cspf
      }
      auto-lsp "Single-Hop-template" {
        one-hop true
      }
    }
  }
}
```

When the **auto-lsp** command is entered, the system starts the process of establishing the single-hop LSPs. A check is made of the TE database for every TE link to a directly connected IGP neighbor, and a single-hop LSP is established across each TE link. The following output is taken from PE-6 and shows the automatically created single-hop LSPs. The LSP names are automatically constructed as TemplateName-DestIPv4Address-TunnelId. The LSP name signaled in the session attribute object concatenates the LSP name with the path name (for example Single-Hop-template-192.0.2.4-61449::loose-path). Recall from [Figure 6: Example Topology for Single-Hop LDP over RSVP with ECMP](#) that PE-6 has a single TE-enabled link to PE-4, and two TE-enabled links to PE-5, so with ECMP=2, there is one LSP to PE-4 (192.0.2.4) and two LSPs to PE-5 (192.0.2.5). However, if ECMP=1, only one single-hop LSP would be signaled to PE-5.

```
[/]
A:admin@PE-6# show router mpls lsp

=====
MPLS LSPs (Originating)
=====
LSP Name                               Tun   Fastfail  Adm  Opr
To                                     Id    Config
-----
Single-Hop-template-192.0.2.4-61448    61448  No        Up   Up
192.0.2.4
Single-Hop-template-192.0.2.5-61449    61449 No       Up Up
192.0.2.5
Single-Hop-template-192.0.2.5-61450    61450 No       Up Up
192.0.2.5
-----
LSPs : 3
=====
```

The purpose of single-hop LSPs is to enable ECMP load balancing using LDP over RSVP, so there is a requirement to configure T-LDP sessions between RSVP LSP endpoints. This can be implemented using static peer provisioning, or it can be done using automatic creation of T-LDP sessions, both of which have been previously described and they are therefore not repeated. In this example, the automatic creation of T-LDP sessions approach is used, and T-LDP sessions are created to adjacent neighbors that are capable of tunneling inside RSVP.

```
[/]
A:admin@PE-6# show router ldp session family ipv4

=====
LDP IPv4 Sessions
=====
Peer LDP Id      Adj Type  State      Msg Sent  Msg Recv  Up Time
-----
192.0.2.4:0     Targeted  Established 42        43        0d 00:02:55
192.0.2.5:0     Targeted  Established 43        45        0d 00:03:04
-----
No. of IPv4 Sessions: 2
=====
```

To validate the ECMP load balancing capability, PE-5 is configured to advertise prefix 172.16.5.0/24 to PE-6. In turn, PE-6 is configured for **ibgp-multipath** to enable load balancing over IGP links to the BGP next hop address, **next-hop-resolution shortcut-tunnel resolution-filter ldp** to enable tunneling of traffic destined toward the BGP next hop in MPLS, and **ecmp 2**. For more information about BGP multipath, see the "BGP Multipath" chapter in the *7450 ESS, 7750 SR, 7950 XRS, and VSR Unicast Routing Protocols Advanced Configuration Guide for MD CLI*.

```
[/]
```

```
A:admin@PE-6# configure {
  router "Base" {
    bgp {
      ibgp-multipath true
      next-hop-resolution {
        shortcut-tunnel {
          family ipv4 {
            resolution filter
            resolution-filter {
              ldp true
              delete rsvp
            }
          }
        }
      }
    }
  }
}
```

The prefix 172.16.5.0/24 advertised by PE-5 is learned on PE-6 and installed in the RIB/FIB with PE-5's system address (192.0.2.5) as next hop.

```
[/]
A:admin@PE-6# show router bgp routes 172.16.5.0/24
```

```
=====
BGP Router ID:192.0.2.6      AS:64496      Local AS:64496
=====
Legend -
Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                  l - leaked, x - stale, > - best, b - backup, p - purge
Origin codes  : i - IGP, e - EGP, ? - incomplete
=====
BGP IPv4 Routes
=====
```

Flag	Network Nexthop (Router) As-Path	LocalPref Path-Id	MED IGP Cost Label
u*>i	172.16.5.0/24 192.0.2.5 No As-Path	100 None	None 10 -

```
-----
Routes : 1
=====
```

Checking the FIB for the next hop address 192.0.2.5, it can be verified that both links are installed as next hop addresses, meaning that ECMP load balancing is active.

```
[/]
A:admin@PE-6# show router fib 1 ip-prefix-prefix-length 192.0.2.5/32
```

```
=====
FIB Display
=====
```

Prefix [Flags] NextHop	Protocol
192.0.2.5/32 192.168.56.1 (int-PE-6-PE-5) 192.168.156.1 (int-PE-6-PE-5-2nd)	ISIS

```
-----
```

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

## Conclusion

Automatic creation of RSVP-TE LSPs provides a good solution for reducing the amount of provisioning activity required when configuring RSVP LSPs. However, there are some constraints with regard to the way that services are deployed on top of those LSPs. SDPs cannot explicitly reference automatically-created RSVP LSPs, which means that automatically created SDPs need to be used for Layer 2 services. In turn, automatically-created SDPs can only use LDP or BGP as a transport tunnel (not RSVP), so, to use the automatically created RSVP mesh, LDP over RSVP must be used. These restrictions need to be fully understood before considering deployment of automatically created RSVP-TE LSPs.

## BFD for RSVP-TE and LDP LSPs

This chapter provides information about BFD for RSVP-TE LSPs and LDP LSPs.

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

### Applicability

This chapter was initially written based on SR OS Release 15.0.R7, but the MD-CLI in the current edition corresponds to SR OS Release 23.3.R3.

### Overview

SR OS supports RFC 5884 and enables LSPs to be monitored between the ingress and egress LERs, regardless of the number of LSRs that the LSP traverses. For continuity checks in MPLS LSPs, BFD packets are transmitted using the MPLS encapsulation, so they share fate with the LSP data path. When enabled, faults to individual LSPs can be detected quickly, so BFD for MPLS LSPs is ideal for monitoring LSPs carrying high-value services, where detecting forwarding failures in a minimum amount of time is critical. The LSPs can be established through RSVP-TE or LDP.

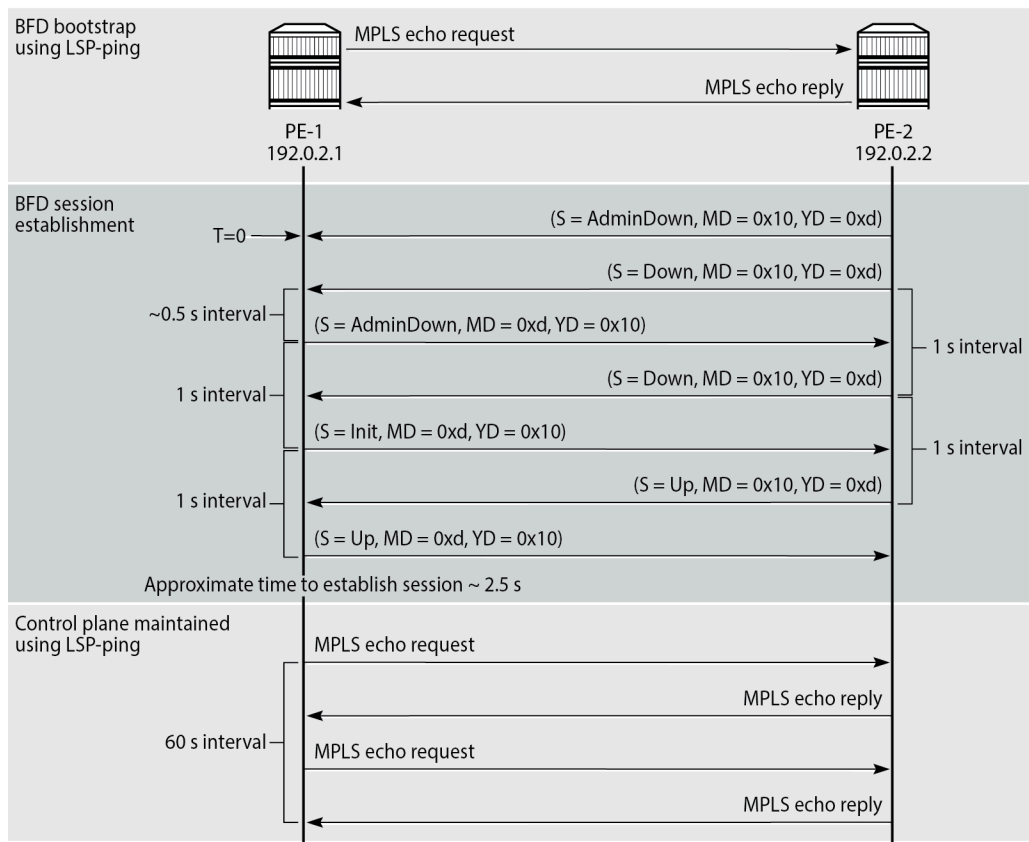
Enabling BFD for LSPs avoids manual hop-by-hop troubleshooting of each element along the LSP. BFD sessions are created and run end-to-end, from ingress to egress, so BFD session state is maintained in the ingress LER and egress LER, but not in intermediate LSRs. If an LSP BFD session changes state, an SNMP trap is generated. Because LSPs are unidirectional, a routed return path is used for the BFD control packets from the egress LER toward the ingress LER.

BFD is only used for fault detection, and will not redirect traffic to an alternate path. On detection of a failure, BFD informs other software components, which then can redirect traffic to avoid faulty links.

BFD is bootstrapped using an LSP ping. An MPLS echo request packet is transmitted along the LSP path, including a BFD discriminator TLV containing the head-end BFD discriminator value. The tail end responds with an echo reply packet, using the IP forwarding path, including the tail-end BFD discriminator value.

Afterward, BFD control packets establish a BFD session between the head end and tail end using the discriminator values from the bootstrap session. The egress LER will send a BFD control packet upon receipt. Every 60 s, the head end transmits an LSP ping for control plane verification. The minimum value for the LSP BFD control transmit interval is 100 ms. [Figure 7: MPLS LSP BFD session establishment: BFD handshake](#) shows the MPLS LSP BFD session establishment with a BFD control transmit interval of 1 second.

Figure 7: MPLS LSP BFD session establishment: BFD handshake



Legend:  
MD - My Discriminator  
YD - Your Discriminator

35627

An MPLS LSP is created from head-end PE-1 to tail-end PE-2 with BFD enabled. A BFD template is created with transmit and receive interval each equal to 1000 ms in this example. This BFD template is applied to the MPLS LSP.

When the MPLS LSP is enabled, an LSP ping is used to bootstrap BFD to LSP with a default ping interval of 60 s. The BFD discriminators (MD = 0x10 and YD = 0xd) are negotiated using LSP ping, so the BFD session establishment starts with a message containing both discriminators. The MPLS echo reply is sent using IP routing instead of using another MPLS LSP.

Each session has its own pair of discriminators, so multiple discriminators are allocated by the system. The two-way handshake to establish a BFD session between network elements takes a few seconds in this example: 2.5 s to establish the BFD session after the BFD is bootstrapped. If the LSP BFD control transmit interval is as low as 100 ms, the handshake takes less than a second.

BFD can be used for RSVP-TE and LDP LSPs. If BFD is applied to RSVP-TE LSPs, it only runs on the currently active path. It cannot determine if any non-active paths (for example, a secondary path or primary path during reversion) that the system might switch to are up and forwarding. If BFD is applied to LDP LSPs, the session runs on the path defined by the underlying IGP.

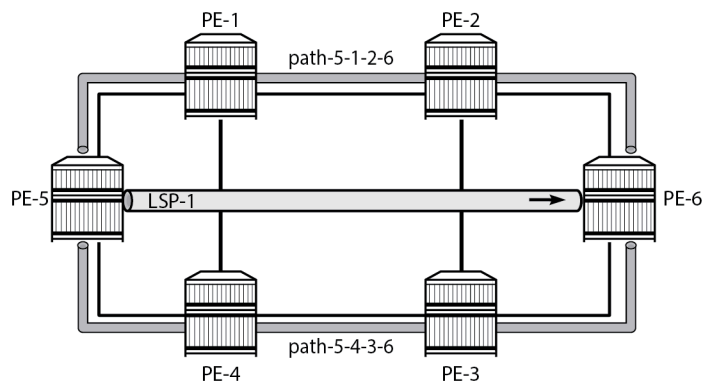
BFD for LSPs can be combined with a failure action. For RSVP-TE LSPs, the failure action can be *down* or *failover*; see the [BFD for RSVP-TE LSPs with Failure Action](#) chapter for more information. For LDP LSPs, the failure action can only be *down*. LSP BFD does not affect the operational state of an LSP, because the operational state is controlled by the control plane. Therefore, the failure action **down** will mark the LSP as unavailable for use by services.

## Configuration

### BFD for RSVP-TE LSPs

[Figure 8: BFD for RSVP-TE LSPs - topology](#) shows the example topology for BFD for RSVP-TE LSPs.

Figure 8: BFD for RSVP-TE LSPs - topology



27613

The initial configuration includes:

- Cards, MDAs, and ports
- Router interfaces
- IS-IS as IGP on all interfaces (alternatively, OSPF can be used), with traffic engineering enabled
- MPLS and RSVP enabled on all interfaces

### Base configuration

The example topology from [Figure 8: BFD for RSVP-TE LSPs - topology](#) has an LSP defined on PE-5 using two strict paths, where *path-5-1-2-6* is taking the upper path and used as the primary path, and *path-5-4-3-6* is the lower path and used as the secondary, as follows:

```
# on PE-5:
configure {
  router "Base" {
    mpls {
      path "path-5-1-2-6" {
        admin-state enable
        hop 10 {
```

```

        ip-address 192.168.15.1
        type strict
    }
    hop 20 {
        ip-address 192.168.12.2
        type strict
    }
    hop 30 {
        ip-address 192.168.26.2
        type strict
    }
}
path "path-5-4-3-6" {
    admin-state enable
    hop 10 {
        ip-address 192.168.45.1
        type strict
    }
    hop 20 {
        ip-address 192.168.34.1
        type strict
    }
    hop 30 {
        ip-address 192.168.36.2
        type strict
    }
}
lsp "lsp-1" {
    admin-state enable
    type p2p-rsvp
    to 192.0.2.6
    path-computation-method local-cspf
    primary "path-5-1-2-6" {
    }
    secondary "path-5-4-3-6" {
    }
}
}

```

## BFD for RSVP-TE LSPs configuration

There are four steps to configure BFD for RSVP-TE LSPs:

1. Configure a BFD template.
2. Enable LSP BFD on the tail node.
3. Apply the BFD template to the LSP or LSP path.
4. Enable BFD on the LSP or LSP path.

### Step 1: Configure a BGP template

The BFD template provides the control packet timer values for the BFD session to use at the LSP head end. The general command to define a BFD template is as follows:

```

configure {
    bfd {
        bfd-template <[32 chars max]> {
            transmit-interval <[10..100000] milli-seconds>
            receive-interval <[10..100000] milli-seconds>
        }
    }
}

```



```

    echo-receive <[100..100000] milli-seconds>
    multiplier <[1..20]>
    type {cpm-np}
}

```



**Note:**

The minimum transmit and receive interval for MPLS LSPs equals 100 ms. Intervals smaller than 100 ms require BFD type CPM-NP and can be used for SR-TE LSPs, but not for RSVP-TE LSPs.

Network processor BFD is not supported for RSVP-TE LSPs. An error is generated if a user tries to apply a BFD template of **type cpm-np** to an LSP, as follows.

```

# on PE-5:
configure {
  bfd {
    bfd-template "bfd-cpm-np-Tx100" {
      type cpm-np
    }
  }
}

```

```

*[ex:/configure router "Base" mpls lsp "lsp-1" bfd]
A:admin@PE-5# bfd-template "bfd-cpm-np-Tx100"

*[ex:/configure router "Base" mpls lsp "lsp-1" bfd]
A:admin@PE-5# commit
MINOR: MGMT_CORE #4001: configure router "Base" mpls lsp "lsp-1"
bfd bfd-template - bfd-template type is not valid for this application -
configure bfd bfd-template "bfd-cpm-np-Tx100" type

```

BFD templates may be used by different BFD applications (for example, LSPs or pseudowires). If the BFD timer values are changed in a template, the BFD sessions on LSPs or spoke-SDPs to which that template applies try to renegotiate their timers to the new values. An error will be generated when the user attempts to commit changes to a BFD template that is already applied to an LSP where the new values are invalid for LSP BFD.

In this example, the BFD template used is configured as follows:

```

# on PE-5:
configure {
  bfd {
    bfd-template "bfdt-1" {
      multiplier 5
      receive-interval 2000
      transmit-interval 2000
    }
  }
}

```

## Step 2: Enable LSP BFD on the tail node

The BFD state machine at the tail end initially uses system-wide parameters, because an LSP is unidirectional so no configuration for the LSP exists at the tail end. The head end then attempts to adjust the control packet timer values when it transitions to the INIT state.

LSP BFD is enabled or disabled on a node-wide basis with the **bfd-sessions <maxlimit>** command in the **configure router lsp-bfd** context. The **maxlimit** parameter configures the maximum number of LSP BFD sessions that can be established. This is required at the tail end of the LSP. The BFD state machine at the

tail end initially uses system-wide parameters, such as the transmit and receive intervals, which are both 1000 ms by default, but the minimum value is 100 ms.

In this example, the tail node is configured as follows:

```
# on PE-6:
configure {
  router "Base" {
    lsp-bfd {
      bfd-sessions 10                # must be set to non-zero value
      # tail-end {
      #   transmit-interval 1000    # default
      #   receive-interval 1000    # default
      #   multiplier 3              # default
      # }
    }
  }
}
```

Because BFD resources are shared by different BFD applications, the limit defined here must provide sufficient resources for other applications.

### Steps 3 and 4: Apply the BGP template to the LSP or LSP path and enable BFD

LSP BFD is applicable to configured RSVP LSPs as well as to mesh point-to-point and one-hop point-to-point auto-created LSPs. It is configured on an RSVP-TE LSP, or on the path of an RSVP-TE LSP, under the **bfd** context at the LSP head end.

A BFD template must always be configured first. BFD is then enabled using the **bfd-liveness** command.

To apply and enable the BFD template at RSVP-TE LSP level, the command is as follows:

```
configure {
  router "Base" {
    mpls {
      lsp <lsp-name> {
        bfd {
          bfd-liveness <boolean>
          bfd-template <name>
          # lsp-ping-interval 60      # default: 60 s
          # failure-action none      # default: none
          # wait-for-up-timer 4       # default: 4 s
        }
      }
    }
  }
}
```

When BFD is configured at the LSP level, BFD packets follow the currently active path of the LSP.

To apply and enable the BFD template at primary path level, the command is as follows:

```
configure {
  router "Base" {
    mpls {
      lsp <lsp-name> {
        primary <path-name> {
          bfd {
            bfd-liveness true
            bfd-template <name>
          }
        }
      }
    }
  }
}
```

It is possible to configure LSP BFD on a secondary path, but the corresponding BFD session will only be established when the secondary path becomes the active path after failover. It is not possible to configure LSP BFD on point-to-multipoint LSPs.

LSP BFD at the LSP level and the path level are mutually exclusive. That is, if LSP BFD is already configured for the LSP, its configuration for the path is blocked. Likewise, it cannot be configured on the LSP if it is already configured at the path level.

LSP BFD is supported on auto-created LSPs. In that case, LSP BFD is configured on mesh point-to-point and one-hop point-to-point auto-created LSPs using the LSP template.

In this example, on the head-end node, the BFD template is applied to the LSP and BFD is enabled, as follows:

```
# on PE-5:
configure {
  router "Base" {
    mpls {
      lsp "lsp-1" {
        bfd {
          bfd-liveness true
          bfd-template "bfdt-1"
        }
      }
    }
  }
}
```

## BFD verification

The details of the MPLS LSP show that BFD is enabled and using BFD template *bfdt-1*, as follows:

```
[/]
A:admin@PE-5# show router mpls lsp detail

=====
MPLS LSPs (Originating) (Detail)
=====
Legend :
+ - Inherited
=====
-----
Type : Originating
-----
LSP Name      : lsp-1
LSP Type      : RegularLsp          LSP Tunnel ID      : 1
LSP Index     : 1                  TTM Tunnel Id      : 1
From          : 192.0.2.5
To            : 192.0.2.6
Adm State     : Up                  Oper State          : Up
LSP Up Time   : 0d 00:26:45         LSP Down Time      : 0d 00:00:00
Transitions   : 1                  Path Changes       : 1
Retry Limit   : 0                  Retry Timer         : 30 sec
Signaling     : RSVP               Resv. Style        : SE
Hop Limit     : 255                Negotiated MTU     : 8982
Adaptive      : Enabled            ClassType           : 0
FastReroute   : Disabled           Oper FR             : Disabled
PathCompMethod : local-cspf        ADSPEC              : Disabled
FallbkPathComp : not-applicable
Metric        : N/A                Metric Type         : igp
Load Bal Wt   : N/A                ClassForwarding    : Disabled
Include Grps  :                    Exclude Grps        :
None
Least Fill    : Disabled           Soft Preemption    : Enabled
BFD Template : bfdt-1          BFD Ping Intvl     : 60
BFD Enable   : True           BFD Failure-action : None
WaitForUpTimer : 4
```

```

---snip---
Primary(a)      : path-5-1-2-6
Bandwidth      : 0 Mbps
Secondary      : path-5-4-3-6
Bandwidth      : 0 Mbps
Up Time        : 0d 00:26:45
Down Time      : 0d 00:26:45
=====

```

Initially, the BFD session is running over path *path-5-1-2-6*, as follows:

```

[/]
A:admin@PE-5# 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 Multipl   Tx Intvl  Rx Intvl
Protocols           Type     LAG Port  LAG ID
Loc Addr            LAG name
-----
lsp-1::path-5-1-2-6      Up        795       732
192.0.2.6           5        2000      2000
rsvpLsp             central   N/A       N/A
192.0.2.5
-----
No. of BFD sessions: 1
=====

```

At the head end, the BFD session details are as follows:

```

[/]
A:admin@PE-5# show router bfd session detail lsp-rsvp head

=====
BFD On LSP Session
=====
Rsvp Session      : lsp-1::path-5-1-2-6
Remote Address    : 192.0.2.6
Lsp Id           : 52224
Oper State       : Up
Recd Msgs       : 786
Up Time         : 0d 00:25:00
Last Down Time  : 0d 00:00:04
Tunnel Id        : 1
Protocols        : rsvpLsp
Sent Msgs       : 855
Up Transitions   : 1
Down Transitions : 0
Version Mismatch : 0

Forwarding Information
Local Discr      : 1
Local Diag      : 0 (None)
Local Mode       : Async
Local Min Tx    : 2000
Last Sent       : 07/10/2023 09:04:18
Type            : central
Remote Discr    : 1
Remote Diag     : 0 (None)
Remote Min Tx   : 1000
Local State     : Up
Local Mult     : 5
Local Min Rx   : 2000
Remote State    : Up
Remote Mode     : Async
Remote Mult     : 3

```

```
Remote C-flag : 1
Last Recv    : 07/10/2023 09:04:18      Remote Min Rx   : 1000
=====
=====
```

At the tail end, the BFD session details are as follows:

```
[/]
A:admin@PE-6# show router bfd session detail lsp-rsvp tail

=====
BFD On LSP Session
=====
Rsvp Session   : lsp-1::path-5-1-2-6
Remote Address : 192.0.2.5
Lsp Id        : 52224
Oper State    : Up
Up Time       : 0d 00:29:55
Last Down Time : 0d 00:00:04
Tunnel Id     : 1
Protocols     : rsvpLsp
Up Transitions : 1
Down Transitions : 0
Version Mismatch : 0

Forwarding Information

Local Discr    : 1
Local Diag     : 0 (None)
Local Mode     : Async
Local Min Tx   : 1000
Last Sent (ms) : 1
Type          : cpm-np
Remote Discr   : 1
Remote Diag    : 0 (None)
Remote Min Tx  : 2000
Remote C-flag  : 0
Last Recv (ms) : 1
Local State    : Up
Local Mult     : 3
Local Min Rx   : 1000
Remote State   : Up
Remote Mode    : Async
Remote Mult    : 5
Remote Min Rx  : 2000
=====
=====
```

A failure is emulated by bringing down the link between PE-1 and PE-2. BFD detects the failure in the upper path quickly, which results in the BFD session being re-established on *path-5-4-3-6*, as follows:

```
[/]
A:admin@PE-5# 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 Multipl   Tx Intvl Rx Intvl
Protocols           Type     LAG Port  LAG ID
Loc Addr                               LAG name
-----
lsp-1::path-5-4-3-6 Up         6         8
192.0.2.6           5         2000     2000
rsvpLsp             central   N/A      N/A
192.0.2.5
-----
No. of BFD sessions: 1
=====
```

Bringing up the link between PE-1 and PE-2 will result in the primary path becoming active again.

The ping bootstrap and periodic verification information for BFD on LSPs can be displayed at the head end, as follows:

```
[/]  
A:admin@PE-5# show test-oam lsp-bfd lsp-name "lsp-1"  
  
-----  
LSP Ping Bootstrap and Periodic Verification Information for BFD on LSPs  
-----  
OAM Operational State      : Bootstrapped - Sending Periodic Verification  
FEC Type                   : RSVP  
LSP Name : lsp-1  
LSP Path Status           : active  
Source Address             : 192.0.2.5  
Replying Node              : 192.0.2.6  
Latest Return Code        : EgressRtr (3)  
Latest Return Subcode     : 1  
Local BFD Discriminator   : 2          Remote BFD Discriminator : 2  
LSP Ping Tx Interval (s) : 60         Bootstrap Retry Count   : 0  
Tx LSP Ping Requests     : 1          Rx LSP Ping Replies     : 1  
-----  
No. of matching BFD on LSP sessions: 1  
-----
```

BFD sessions changing state are trapped, so these are logged in log 99, as follows:

```
120 2023/07/10 09:10:21.071 UTC MINOR: BFD #2002 Base 192.0.2.6  
"The lspHead BFD session with Local Discriminator 2 on 192.0.2.6 is up (Tunnel Id 1, Path LSP  
ID 52226)"  
  
121 2023/07/10 09:11:23.500 UTC WARNING: MPLS #2011 Base VR 1:  
"LSP path lsp-1::path-5-1-2-6 is operationally enabled ('no shutdown')"  
  
122 2023/07/10 09:11:24.494 UTC MINOR: BFD #2004 Base 192.0.2.6  
"The protocol (RSVP LSP) using BFD session on node 192.0.2.6 has been added"  
  
123 2023/07/10 09:11:25.494 UTC MINOR: MPLS #2027 Base VR 1:  
"LSP lsp-1 active path lsp-1::path-5-4-3-6 has changed to active path lsp-1::path-5-1-2-6"  
  
124 2023/07/10 09:11:25.494 UTC MINOR: BFD #2004 Base 192.0.2.6  
"The protocol (RSVP LSP) using BFD session on node 192.0.2.6 has been cleared"  
  
125 2023/07/10 09:11:25.494 UTC MINOR: BFD #2003 Base 192.0.2.6  
"The lspHead BFD Session with Local Discriminator 2 on 192.0.2.6 has been deleted"  
  
126 2023/07/10 09:11:25.495 UTC MAJOR: SVCMGR #2316 Base  
"Processing of a SDP state change event is finished and the status of all affected SDP Bindings  
on SDP 56 has been updated."  
  
127 2023/07/10 09:11:28.494 UTC WARNING: MPLS #2012 Base VR 1:  
"LSP path lsp-1::path-5-4-3-6 is operationally disabled ('shutdown') because nonActive  
Secondary"  
  
128 2023/07/10 09:11:29.231 UTC MINOR: BFD #2002 Base 192.0.2.6  
"The lspHead BFD session with Local Discriminator 3 on 192.0.2.6 is up (Tunnel Id 1, Path LSP  
ID 52228)"
```

The **tools** command for displaying LSP details at the head end also includes BFD related information, if applicable, as follows:

```
[/]
A:admin@PE-5# tools dump router mpls lspinfo "lsp-1" detail
LSP "lsp-1" LspIdx 1 LspType Dynamic State LSPS_UP Flags 0x2000
AdminState Up OperState Up RowStatus Active
From N/A To 192.0.2.6
NumPaths 2 NumSdps 1 NumCBFSdps 0 NumFltrEntries 0
ActivePath lsp-1::path-5-1-2-6(LspId 52228)
HoldTimeRemaining 0secs ClassType 0 SoftPreemption TRUE Metric 0 OperMetric 30
LDPoRsvp Include VprnAutoBind Include IgpShortCut Include BgpShortCut Include
BgpTransTunnel Include IpShCutTtlPropLocal TRUE IpShCutTtlPropTans TRUE
RelativeMetricOffset 2147483647 MTU 8982 InUseByLdp FALSE TTMPref 7
EntropyLabel inherit OperEntropyLabel enable NegEntropyLabel disable
ClassForwarding: Disabled
BFD Enabled Template bfdt-1 PingInterval 60 FailureAction None WaitForUp 4sec
PCE Report: Disabled PCE Control: Disabled
Path Profile:
None
Admin Tags:
None
Lsp-self-ping: Config: inherit, Oper: Disabled, TimedOutCnt: 0 OamNoRsc: 0
Path "path-5-1-2-6" LspId 52228 LspPathIndex 2 PathType Primary ActivePath Yes
RowStatus Active LastChange 000 00:51:09.090
AdminState Up OperState Up OperStateChange 000 00:14:35.230
TE Computed Hop List:
Hop[1] IngIp 192.0.2.5 IngLnkId 0 EgrIp 192.168.15.2 EgrLnkId 0 RtrId 192.0.2.5 Flag 0x0
Hop[2] IngIp 192.168.15.1 IngLnkId 0 EgrIp 192.168.12.1 EgrLnkId 0 RtrId 192.0.2.1 Flag
0x0
Hop[3] IngIp 192.168.12.2 IngLnkId 0 EgrIp 192.168.26.1 EgrLnkId 0 RtrId 192.0.2.2 Flag
0x0
Hop[4] IngIp 192.168.26.2 IngLnkId 0 EgrIp 192.0.2.6 EgrLnkId 0 RtrId 192.0.2.6 Flag 0x0
LspPath FsmState LSP_PATH_S_UP Flags 0x0 miscFlags 0x2
RetryAttempts 0 RetryInterval 30 NextRetryIn 0secs
FailNode 0.0.0.0 FailCode noError
Class Type 0 SetupPri 7 HoldPri 0 Pref 0 HopLimit 255 BW 0Mbps
TotIgpCost 30 OperMetric 30 MTU 8982
BFD Disabled Template n/a PingInterval 60 WaitForUp 4sec
Degraded No
Oper Values:
Class Type 0 SetupPri 7 HoldPri 0 HopLimit 255 BW 0Mbps
RecordRoute RecordLabel No Adspec
No PropagateAdminGroup Exclude 0x00000000 Include 0x00000000
No FRR
Metric 30 CSPF No Least Fill Intra-area
NegotiatedEntropyLabel Disabled
PCE-Computed No PCE-Reported No PCE-Controlled No
BFD State Up InitTime 0d 00:14:34 UpTime 0d 00:14:30
OldMBBPathsCleanedUp Yes
Path "path-5-4-3-6" LspId 52226 LspPathIndex 3 PathType Secondary ActivePath No
RowStatus Active LastChange 000 00:51:09.090
AdminState Up OperState Down OperStateChange 000 00:14:30.230
LspPath FsmState LSP_PATH_S_DOWN Flags 0x40000 miscFlags 0x2
RetryAttempts 0 RetryInterval 30 NextRetryIn 0secs
FailNode 192.0.2.5 FailCode nonActiveSecondary
Class Type 0 SetupPri 7 HoldPri 0 Pref 255 HopLimit 255 BW 0Mbps
TotIgpCost 0 OperMetric 16777215 MTU 0
SRLG Disabled SRLGDisjoint No
BFD Disabled Template n/a PingInterval 60 WaitForUp 4sec
Degraded No
OldMBBPathsCleanedUp Yes
```

```
Total Ingress LSP Count      : 1
```

The current BFD session information for RSVP LSPs can be displayed using the following **tools** command at the head end:

```
[/]
A:admin@PE-5# tools dump router bfd lsp-rsvp head
-----
FEC: (PTR 0x10eb60f50)
RSVP : vrId: 1 (To: 192.0.2.6 - 1 - 192.0.2.5), Sender (192.0.2.5 - 52228)

Session: lsp-1::path-5-1-2-6    refCnt = 1
PingIntvl: 60 Flags: 0x6 ProtNhidx: 13 NumNextHop: 1
TemplName: bfdt-1 LspName: lsp-1 TunnelId: 1 NumLspUser: 0
NextHop: 192.168.15.1 IfIndex: 1 Flags: 0x0 isBackup: N
PGId: 0 [State: N/A] NhIdx: 13
Label:- [0]524286
      BFD Handle: 3 State: UP LastEvent: UP
      BFD UserId: 24 TmrActive: N [0] NumRetry: 0
      DstAddr: 127.0.0.3 LocalDiscr: 3 RemoteDiscr: 0
-----
Total FEC Count in Head: 1
```

Other **tools** commands can display BFD LSP information at the tail end, as follows:

```
[/]
A:admin@PE-6# tools dump test-oam lsp-bfd tail
-----
Total Number of Active Tail Cache Sessions : 1
-----

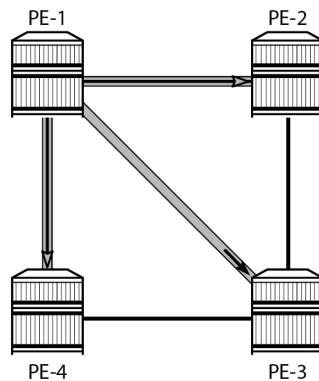
VrId          : 1
RemoteBfdDisc : 3
LocalBfdDisc  : 3
FecType       : rsvp_ipv4(3)
LspId         : 52228
TunnelId      : 1
SenderIp      : 192.0.2.5
TunnEndIp     : 192.0.2.6
ExtTunnId     : 192.0.2.5
Bootstrap Echo Rx : rcvd 2023/07/10 09:11:24.00 UTC
                  handle 3 seqNum 2 rc 3 rsc 1
Last Echo Req Rx  : rcvd 2023/07/10 09:28:30.00 UTC
                  handle 3 seqNum 19 rc 3 rsc 1
-----
Number of Matched Tail Cache Sessions : 1
-----
```

## BFD for LDP LSPs

Figure 9: BFD for LDP LSPs - topology shows the example topology for BFD for LDP LSPs.



Figure 9: BFD for LDP LSPs - topology



27614

The initial configuration includes:

- Cards, MDAs, and ports
- Router interfaces
- IS-IS as IGP on all interfaces (alternatively, OSPF can be used)

## Base configuration

The example topology from [Figure 9: BFD for LDP LSPs - topology](#) has LDP configured on all interfaces. LDP automatically generates and distributes labels across the network, so for the topology in [Figure 9: BFD for LDP LSPs - topology](#), LDP tunnels are created so that every node can reach any other node; only the tunnels originating in PE-1 are shown. The LDP configuration for PE-1 is as follows; the LDP configuration for PE-2, PE-3, and PE-4 is similar.

```
# on PE-1:
configure {
  router "Base" {
    ldp {
      interface-parameters {
        interface "int-PE-1-PE-2" {
          ipv4 {
          }
        }
        interface "int-PE-1-PE-4" {
          ipv4 {
          }
        }
      }
    }
  }
}
```

## BFD for LDP LSPs configuration

There are six steps to configure BFD for LDP LSPs:

1. Create a BFD template.
2. Enable LSP BFD on the tail node.

3. Create a prefix list.
4. Configure LSP BFD for LDP.
5. Apply the BFD template to the LDP LSP.
6. Enable BFD on the LDP LSP.

### Step 1: Create a BFD template

The command to define a BFD template is the same as for RSVP-TE LSPs. The BFD template used for the LDP LSPs is configured as follows:

```
# on PE-1:
configure
  bfd {
    bfd-template "bfdt-2" {
      receive-interval 1000
      transmit-interval 1000
    }
  }
```

### Step 2: Enable LSP BFD on the tail node

The command to enable or disable LSP BFD on a node-wide basis at the tail end of the tunnels is the same as for RSVP-TE LSPs. In this example, the tail nodes PE-2 and PE-3 are configured as follows:

```
# on PE-2 and PE-3:
configure {
  router "Base" {
    lsp-bfd {
      bfd-sessions 5                # must be set to non-zero value
      # tail-end {
        # transmit-interval 1000    # default
        # receive-interval 1000    # default
        # multiplier 3              # default
      # }
    }
  }
}
```

### Step 3: Create a prefix list

When high-value services are relying on the LDP tunnels between PE-1, PE-2, and PE-3, a prefix list with the system IP addresses (or other routable loopback addresses) of PE-2 and PE-3 can be used in PE-1 for monitoring these tunnels. In this example, the *pfx-lst-1* prefix list is defined as follows:

```
# on PE-1
configure {
  policy-options {
    prefix-list "pfx-lst-1" {
      prefix 192.0.2.2/32 type exact {
      }
      prefix 192.0.2.3/32 type exact {
      }
    }
  }
}
```

## Steps 4, 5, and 6: Configure LSP BFD for LDP, apply BFD template, enable BFD

LSP BFD is configured for LDP using the following commands:

```
configure {
  router "Base" {
    ldp {
      lsp-bfd <prefix-list-name> {
        priority <priority-level>           # default: 1
        bfd-template <bfd-template-name>
        source-address <ip-address>
        bfd-liveness <boolean>
        lsp-ping-interval <seconds>        # default: 60
        failure-action <down>              # default: none
      }
    }
  }
}
```

The priority level is set to one, by default, and is used in case a prefix appears in multiple prefix lists; see the *7450 ESS*, *7750 SR*, *7950 XRS*, and *VSR MPLS Guide* for more information. The source address can be any local address routable by the other nodes in the network; by default, the system IP address is used. The LSP ping interval defines how frequently ping messages must be sent on the LSP. The only possible failure action for LDP LSPs is *down*, which makes the LSP unavailable for user traffic. By default, no failure action is configured.

In this example, the BFD template is applied to the head-end node in the **lsp-bfd** context, as follows. The **lsp-bfd** command takes the prefix list name defined in [step 3](#) as its argument. BFD template must always be applied first. BFD is then enabled using the **bfd-liveness** command.

```
# on PE-1:
configure {
  router "Base" {
    ldp {
      lsp-bfd "pfx-lst-1" {
        bfd-template "bfdt-2"
        source-address 192.0.2.1
        bfd-liveness true
      }
    }
  }
}
```

## BFD verification

The prefix lists applied to LDP BFD are the following:

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

=====
BFD on LDP LSP Configuration Summary
=====
Prio  Prefix List Name                Enabled  Prefixes
-----
1     pfx-lst-1                       Yes      2
-----
No. of prefix lists: 1
=====
```

The LDP BFD information for prefix list *pfx-lst-1* is as follows:

```
[/]
A:admin@PE-1# show router ldp lsp-bfd "pfx-lst-1"

=====
BFD on LDP LSP Configuration Detail
=====
Prefix List       : pfx-lst-1
Prefix Count     : 2
BFD Template     : bfdt-2
Source Address   : 192.0.2.1
BFD Enable      : Yes
LSP Ping Interval : 60 seconds
Failure Action   : none
Priority         : 1
=====
```

The prefixes of prefix list *pfx-lst-1* to which the system tries to establish BFD sessions are the following:

```
[/]
A:admin@PE-1# show router ldp lsp-bfd "pfx-lst-1" prefixes

=====
BFD on LDP LSP Prefix List "pfx-lst-1" (Enabled)
=====
Prefix                               Operational State
-----
192.0.2.2/32                          Up
192.0.2.3/32                          Up
-----
No. of prefixes: 2
=====
```

The LDP BFD session data created and maintained at the head end PE-1 is as follows:

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

=====
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
-----
192.0.2.2/32         Up        469      427
N/A                  3        1000     1000
ldpLsp              central   N/A      N/A
192.0.2.1
192.0.2.3/32         Up        471      426
N/A                  3        1000     1000
ldpLsp              central   N/A      N/A
192.0.2.1
-----
No. of BFD sessions: 2
=====
```

The following command shows the LDP BFD session information at the tail end PE-3:

```
[/]
A:admin@PE-3# show router bfd session lsp-ldp tail

=====
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
-----
192.0.2.3/32        Up         N/A       N/A
192.0.2.1           3         1000      1000
  ldpLsp             cpm-np    N/A       N/A
192.0.2.3
-----
No. of BFD sessions: 1
=====
```

The ping bootstrap and periodic verification information for BFD on LSPs can be displayed at the head end, as follows:

```
[/]
A:admin@PE-1# show test-oam lsp-bfd

-----
LSP Ping Bootstrap and Periodic Verification Information for BFD on LSPs
-----
OAM Operational State : Bootstrapped - Sending Periodic Verification
FEC Type               : LDP
Prefix                 : 192.0.2.2/32
Source Address         : 192.0.2.1
Replying Node          : 192.0.2.2
Latest Return Code     : EgressRtr (3)
Latest Return Subcode  : 1
Local BFD Discriminator : 1           Remote BFD Discriminator : 1
LSP Ping Tx Interval (s) : 60       Bootstrap Retry Count   : 0
Tx LSP Ping Requests   : 10         Rx LSP Ping Replies     : 10
-----
OAM Operational State : Bootstrapped - Sending Periodic Verification
FEC Type               : LDP
Prefix                 : 192.0.2.3/32
Source Address         : 192.0.2.1
Replying Node          : 192.0.2.3
Latest Return Code     : EgressRtr (3)
Latest Return Subcode  : 1
Local BFD Discriminator : 2           Remote BFD Discriminator : 1
LSP Ping Tx Interval (s) : 60       Bootstrap Retry Count   : 0
Tx LSP Ping Requests   : 10         Rx LSP Ping Replies     : 10
-----
No. of matching BFD on LSP sessions: 2
-----
```

BFD sessions changing state are trapped, so these are logged to log 99. When the link between PE-1 and PE-2 is restored, the following messages are logged in log 99:

```

93 2023/07/10 14:13:21.913 UTC WARNING: ISIS #2045 Base VR: 1 ISIS (0) Adjacency state
"Adjacency status changed to initializing for interface: int-PE-1-PE-2, for level: l1l2, LSP-
id: 1920.0000.2002.00-00 "

94 2023/07/10 14:13:21.915 UTC WARNING: ISIS #2045 Base VR: 1 ISIS (0) Adjacency state
"Adjacency status changed to up for interface: int-PE-1-PE-2, for level: l1l2, LSP-id:
1920.0000.2002.00-00 "

95 2023/07/10 14:13:22.583 UTC WARNING: RSVP #2003 Base VR 1:
"Neighbor 192.168.12.2 on interface int-PE-1-PE-2 changed to active state"

96 2023/07/10 14:13:22.942 UTC MINOR: BFD #2004 Base 192.0.2.2
"The protocol (LDP LSP) using BFD session on node 192.0.2.2 has been cleared"

97 2023/07/10 14:13:22.942 UTC MINOR: BFD #2004 Base 192.0.2.3
"The protocol (LDP LSP) using BFD session on node 192.0.2.3 has been cleared"

98 2023/07/10 14:13:22.943 UTC MINOR: BFD #2003 Base 192.0.2.2
"The lspHead BFD Session with Local Discriminator 1 on 192.0.2.2 has been deleted"

99 2023/07/10 14:13:22.943 UTC MINOR: BFD #2003 Base 192.0.2.3
"The lspHead BFD Session with Local Discriminator 2 on 192.0.2.3 has been deleted"

100 2023/07/10 14:13:23.382 UTC MINOR: BFD #2004 Base 192.0.2.2
"The protocol (LDP LSP) using BFD session on node 192.0.2.2 has been added"

101 2023/07/10 14:13:23.414 UTC MINOR: BFD #2004 Base 192.0.2.3
"The protocol (LDP LSP) using BFD session on node 192.0.2.3 has been added"

102 2023/07/10 14:13:26.874 UTC MINOR: BFD #2002 Base 192.0.2.3
"The lspHead BFD session with Local Discriminator 4 on 192.0.2.3 is up"

103 2023/07/10 14:13:33.155 UTC MINOR: BFD #2002 Base 192.0.2.2
"The lspHead BFD session with Local Discriminator 3 on 192.0.2.2 is up"

```

The current BFD session information for LDP LSPs can be displayed using a **tools** command, as follows:

```

[/]
A:admin@PE-1# tools dump router bfd lsp-ldp prefix 192.0.2.3/32
-----
FEC: (PTR 0x10eb60ff0)
LDP: vrId: 1 (To: 192.0.2.3/32), Sender (192.0.2.1)
  PingIntvl: 60 Flags: 0x6 ProtNhidx: 16 NumNextHop: 1
  TemplName: bfd-2 LspName: TunnelId: 65538 NumLspUser: 0
  NextHop: 192.168.12.2 IfIndex: 1 Flags: 0x0 isBackup: N
  PGId: 9 [State: UP] NhIdx: 16
  Label: - [0]524283
    BFD Handle: 4 State: UP LastEvent: UP
    BFD UserId: 25 TmrActive: N [0] NumRetry: 0
    DstAddr: 127.0.0.4 LocalDiscr: 4 RemoteDiscr: 0
-----
Total FEC Count in Head: 1

Total FEC Count in Tail: 0

```

The BFD templates used by LDP can also be listed using a **tools** command, as follows:

```

[/]
A:admin@PE-1# tools dump router ldp lsp-bfd bfd-templates-in-use

```

```
=====
BFD on LDP LSP BFD Template Summary
=====
Prefix List Name          BFD Template Name
-----
pfx-lst-1                 bfdt-2
-----
No. of prefix lists: 1
=====
```

## Conclusion

BFD is supported for RSVP-TE and LDP LSPs and is ideal for monitoring LSPs carrying high-value services, where detecting failures in a minimum amount of time is critical.

# BFD for RSVP-TE LSPs with Failure Action

This chapter describes BFD for RSVP-TE LSPs with failure action.

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

## Applicability

This chapter was initially written based on SR OS Release 15.0.R7, but the MD-CLI in the current edition corresponds to SR OS Release 23.7.R1.

The chapter [BFD for RSVP-TE and LDP LSPs](#) is prerequisite reading.

## Overview

Using the **failure-action** command, the operator can configure the action taken by the system if a BFD session fails for an RSVP LSP or LDP prefix list.

When **failure-action failover** is configured, and the LSP BFD session goes down on the currently active path, the LSP switches from the primary path to a secondary path, or from the currently active secondary path to the next best preference secondary path if the currently active path was a secondary.

When **failure-action down** is configured, the LSP is registered as unusable in the tunnel table manager (TTM) when BFD on the LSP goes down. A tunnel being registered as unusable in TTM is not available to the routing table manager (RTM) and all routes using that tunnel are withdrawn. SDP auto-bind will not use an LSP until it is registered as usable. Traffic cannot pass through that LSP, even when secondary paths are available for that LSP.

In either case, SNMP traps are raised when the BFD state machine for the LSP transitions.

Nokia recommends configuring the BFD control packet timer intervals long enough to deal with transient data path disruptions that may occur when the underlying transport network recovers following a failure.

LSP BFD only runs on the currently active path. It cannot determine if any non-active paths (for example, a secondary path or primary path during reversion) that the system might switch to are up and forwarding.

When BFD failure action is configured on an RSVP-TE LSP directly, the action can be failover or down. When BFD failure action is configured on an RSVP-TE LSP indirectly, through an LSP template, the only action available is down. This chapter only covers the direct configuration of a failure action.



## Configuration

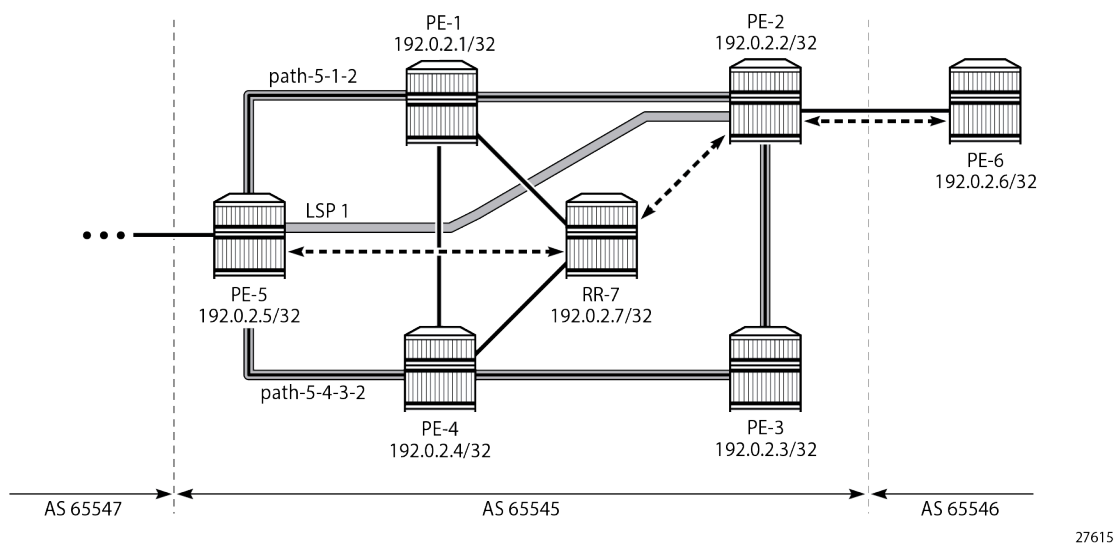
The following scenarios are described in this section:

- Failure action failover
- Failure action down

### Failure action failover

Figure 10: [Topology for failure action failover](#) shows the topology used for failure action failover. A BGP shortcut is defined in AS 65545 running between the autonomous system border routers (ASBRs) PE-5 and PE-2. That shortcut is RSVP-TE LSP *lsp-1* composed of two paths, where the first path is the upper path from PE-5 via PE-1 to PE-2, and the second path is the lower path from PE-5 via PE-4 and PE-3 to PE-2.

Figure 10: *Topology for failure action failover*



The initial configuration includes:

- Cards, MDAs, and ports
- Router interfaces
- IS-IS as IGP on all interfaces (alternatively, OSPF can be used), with traffic engineering enabled
- MPLS and RSVP enabled on all interfaces
- BGP configured, with RR-7 being the route reflector in AS 65545 for clients PE-2 and PE-5, and PE-6 located in AS 65546 and connected to PE-2. PE-6 advertises its prefix 192.0.2.6/32 to AS 65545.

The *lsp-1* LSP is defined with primary path *path-5-1-2* and secondary path *path-5-4-3-2*. The two paths are established when the LSP is brought up, to minimize traffic loss in case of a failure. BFD template *bfdt-1* with a failure action failover is applied to *lsp-1* at the LSP level.

```
# on ASBR PE-5:
```

```

configure {
  router "Base" {
    mpls {
      path "path-5-1-2" {
        admin-state enable
        hop 10 {
          ip-address 192.168.15.1
          type strict
        }
        hop 20 {
          ip-address 192.168.12.2
          type strict
        }
      }
      path "path-5-4-3-2" {
        admin-state enable
        hop 10 {
          ip-address 192.168.45.1
          type strict
        }
        hop 20 {
          ip-address 192.168.34.1
          type strict
        }
        hop 30 {
          ip-address 192.168.23.1
          type strict
        }
      }
    }
    lsp "lsp-1" {
      admin-state enable
      type p2p-rsvp
      to 192.0.2.2
      path-computation-method local-cspf
      bfd {
        bfd-liveness true
        bfd-template "bfdt-1"
        failure-action failover
      }
      primary "path-5-1-2" {
      }
      secondary "path-5-4-3-2" {
        standby true
      }
    }
  }
}

```

The details of the LSP show the configured failure action, as follows:

```

[/]
A:admin@PE-5# show router mpls lsp "lsp-1" detail

=====
MPLS LSPs (Originating) (Detail)
=====
Legend :
+ - Inherited
=====
Type : Originating
-----
LSP Name      : lsp-1
LSP Type      : RegularLsp                LSP Tunnel ID      : 1
LSP Index     : 1                        TTM Tunnel Id     : 1

```

```

From       : 192.0.2.5
To         : 192.0.2.2
Adm State  : Up
LSP Up Time : 0d 00:13:02
Transitions : 1
Retry Limit : 0
Signaling  : RSVP
---snip---

Oper State : Up
LSP Down Time : 0d 00:00:00
Path Changes : 1
Retry Timer  : 30 sec
Resv. Style  : SE

BFD Template : bfdt-1
BFD Enable   : True
WaitForUpTimer : 4
---snip---

BFD Ping Intvl : 60
BFD Failure-action : Failover

Primary(a) : path-5-1-2
Up Time    : 0d 00:13:02

Bandwidth  : 0 Mbps
Standby    : path-5-4-3-2
Up Time    : 0d 00:13:02

Bandwidth  : 0 Mbps
=====

```

With this configuration, the BFD session is running over the upper path, as follows:

```

[/]
A:admin@PE-5# 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 Multipl   Tx Intvl Rx Intvl
Protocols          Type     LAG Port  LAG ID
Loc Addr                               LAG name
-----
lsp-1::path-5-1-2  Up        461      421
192.0.2.2          5         2000     2000
rsvpLsp            central   N/A      N/A
192.0.2.5
-----
No. of BFD sessions: 1
=====

```

BGP route 192.0.2.6/32 is advertised by PE-6 out of AS 65546, as follows. This route has next hop 192.0.2.2, which is the system address of PE-2.

```

[/]
A:admin@PE-5# show router bgp routes

=====
BGP Router ID:192.0.2.5      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    IGP Cost
As-Path                                     Label
-----
u*>i 192.0.2.6/32                            100       None
     192.0.2.2                              1         20
     65546                                    -         -
-----
Routes : 1
=====

```

To keep the core of AS 65545 BGP free, traffic is tunneled through the *lsp-1* LSP, as follows:

```

[/]
A:admin@PE-5# show router route-table 192.0.2.6/32

=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]                        Type  Proto  Age           Pref
Next Hop[Interface Name]                  Metric
-----
192.0.2.6/32                               Remote BGP     00h04m31s  170
    192.0.2.2 (tunneled:RSVP:1)             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
=====

```

The LSP and path details can also be shown using a **tools dump** command, as follows. LSP *lsp-1* is up, *path-5-1-2* is the active path taking three hops, so the operational metric is 20, and *path-5-4-3-2* is the standby path, but not active.

```

[/]
A:admin@PE-5# tools dump router mpls lspinfo "lsp-1" detail
LSP "lsp-1" LspIdx 1 LspType Dynamic State LSPS_UP Flags 0x2000
AdminState Up OperState Up RowStatus Active
From N/A To 192.0.2.2
NumPaths 2 NumSdps 0 NumCBFSdps 0 NumFltrEntries 0
ActivePath lsp-1: path-5-1-2(LspId 56320)
HoldTimeRemaining 0secs ClassType 0 SoftPreemption TRUE Metric 0 OperMetric 20
LDPoRsvp Include VprnAutoBind Include IgpShortCut Include BgpShortCut Include
BgpTransTunnel Include IpShCutTtlPropLocal TRUE IpShCutTtlPropTans TRUE
RelativeMetricOffset 2147483647 MTU 8982 InUseByLdp FALSE TTMPref 7
EntropyLabel inherit OperEntropyLabel enable NegEntropyLabel disable
ClassForwarding: Disabled
BFD Enabled Template bfdt-1 PingInterval 60 FailureAction failover WaitForUp 4sec
PCE Report: Disabled PCE Control: Disabled
Path Profile:
None
Admin Tags:
None
Lsp-self-ping: Config: inherit, Oper: Disabled, TimedOutCnt: 0 OamNoRsc: 0
Path "path-5-1-2" LspId 56320 LspPathIndex 1 PathType Primary ActivePath Yes
RowStatus Active LastChange 000 00:14:05.560
AdminState Up OperState Up OperStateChange 000 00:14:05.550
TE Computed Hop List:
Hop[1] IngIp 192.0.2.5 IngLnkId 0 EgrIp 192.168.15.2 EgrLnkId 0 RtrId 192.0.2.5 Flag 0x0

```

```

Hop[2] IngIp 192.168.15.1 IngLnkId 0 EgrIp 192.168.12.1 EgrLnkId 0 RtrId 192.0.2.1 Flag
0x0
Hop[3] IngIp 192.168.12.2 IngLnkId 0 EgrIp 192.0.2.2 EgrLnkId 0 RtrId 192.0.2.2 Flag 0x0
LspPath FsmState LSP_PATH_S_UP Flags 0x0 miscFlags 0x2
RetryAttempts 0 RetryInterval 30 NextRetryIn 0secs
FailNode 0.0.0.0 FailCode noError
Class Type 0 SetupPri 7 HoldPri 0 Pref 0 HopLimit 255 BW 0Mbps
TotIgpCost 20 OperMetric 20 MTU 8982
BFD Disabled Template n/a PingInterval 60 WaitForUp 4sec
Degraded No
Oper Values:
Class Type 0 SetupPri 7 HoldPri 0 HopLimit 255 BW 0Mbps
RecordRoute RecordLabel No Adspec
No PropagateAdminGroup Exclude 0x00000000 Include 0x00000000
No FRR
Metric 20 CSPF No Least Fill Intra-area
NegotiatedEntropyLabel Disabled
PCE-Computed No PCE-Reported No PCE-Controlled No
BFD State Up InitTime 0d 00:14:06 UpTime 0d 00:14:01
OldMBBPathsCleanedUp Yes
Path "path-5-4-3-2" LspId 56322 LspPathIndex 2 PathType Standby ActivePath No
RowStatus Active LastChange 000 00:14:05.560
AdminState Up OperState Up OperStateChange 000 00:14:05.550
TE Computed Hop List:
Hop[1] IngIp 192.0.2.5 IngLnkId 0 EgrIp 192.168.45.2 EgrLnkId 0 RtrId 192.0.2.5 Flag 0x0
Hop[2] IngIp 192.168.45.1 IngLnkId 0 EgrIp 192.168.34.2 EgrLnkId 0 RtrId 192.0.2.4 Flag
0x0
Hop[3] IngIp 192.168.34.1 IngLnkId 0 EgrIp 192.168.23.2 EgrLnkId 0 RtrId 192.0.2.3 Flag
0x0
Hop[4] IngIp 192.168.23.1 IngLnkId 0 EgrIp 192.0.2.2 EgrLnkId 0 RtrId 192.0.2.2 Flag 0x0
LspPath FsmState LSP_PATH_S_UP Flags 0x0 miscFlags 0x2
RetryAttempts 0 RetryInterval 30 NextRetryIn 0secs
FailNode 0.0.0.0 FailCode noError
Class Type 0 SetupPri 7 HoldPri 0 Pref 255 HopLimit 255 BW 0Mbps
TotIgpCost 30 OperMetric 30 MTU 8982
SRLG Disabled SRLGDisjoint No
BFD Disabled Template n/a PingInterval 60 WaitForUp 4sec
Degraded No
Oper Values:
Class Type 0 SetupPri 7 HoldPri 0 HopLimit 255 BW 0Mbps
RecordRoute RecordLabel No Adspec
No PropagateAdminGroup Exclude 0x00000000 Include 0x00000000
No FRR
Metric 30 CSPF No Least Fill Intra-area
NegotiatedEntropyLabel Disabled
PCE-Computed No PCE-Reported No PCE-Controlled No
BFD State N/A
OldMBBPathsCleanedUp Yes

Total Ingress LSP Count          : 1
    
```

Bringing down the link between PE-1 and PE-2 results in the secondary path, *path-5-4-3-2* of LSP *lsp-1*, becoming active, and the BFD session is re-established on that path, as follows:

```

[/]
A:admin@PE-5# 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
-----
lsp-1::path-5-4-3-2                      Up        4         6
192.0.2.2                                5        2000     2000
rsvpLsp                                   central   N/A      N/A
192.0.2.5
-----
No. of BFD sessions: 1
=====

```

BFD sessions changing state are logged in the trap log, as follows:

```

91 2023/08/11 11:50:08.101 CEST WARNING: MPLS #2012 Base VR 1:
"LSP path lsp-1::path-5-1-2 is operationally disabled ('shutdown') because resvTear"

92 2023/08/11 11:50:08.101 CEST MINOR: MPLS #2027 Base VR 1:
"LSP lsp-1 active path lsp-1::path-5-1-2 has changed to active path lsp-1::path-5-4-3-2"

93 2023/08/11 11:50:08.101 CEST MINOR: BFD #2004 Base 192.0.2.2
"The protocol (RSVP LSP) using BFD session on node 192.0.2.2 has been cleared"

94 2023/08/11 11:50:08.101 CEST MINOR: BFD #2003 Base 192.0.2.2
"The lspHead BFD Session with Local Discriminator 1 on 192.0.2.2 has been deleted"

95 2023/08/11 11:50:08.102 CEST MINOR: BFD #2004 Base 192.0.2.2
"The protocol (RSVP LSP) using BFD session on node 192.0.2.2 has been added"

96 2023/08/11 11:50:12.944 CEST MINOR: BFD #2002 Base 192.0.2.2
"The lspHead BFD session with Local Discriminator 2 on 192.0.2.2 is up (Tunnel Id 1, Path LSP
ID 56322)"

```

The **tools dump** command shows that *lsp-1* is still up, and *path-5-4-3-2* is active with four hops, so the LSP operational metric is 30, as follows:

```

[/]
A:admin@PE-5# tools dump router mpls lspinfo "lsp-1" detail
LSP "lsp-1" LspIdx 1 LspType Dynamic State LSPS_UP Flags 0x2000
AdminState Up OperState Up RowStatus Active
From N/A To 192.0.2.2
NumPaths 2 NumSdps 0 NumCBFSdps 0 NumFltrEntries 0
ActivePath lsp-1::path-5-4-3-2(LspId 56322)
HoldTimeRemaining 0secs ClassType 0 SoftPreemption TRUE Metric 0 OperMetric 30
LDPoRsvp Include VprnAutoBind Include IgpShortCut Include BgpShortCut Include
BgpTransTunnel Include IpShCutTtlPropLocal TRUE IpShCutTtlPropTans TRUE
RelativeMetricOffset 2147483647 MTU 8982 InUseByLdp FALSE TTMPref 7
EntropyLabel inherit OperEntropyLabel enable NegEntropyLabel disable
ClassForwarding: Disabled
BFD Enabled Template bfdt-1 PingInterval 60 FailureAction failover WaitForUp 4sec
PCE Report: Disabled PCE Control: Disabled
Path Profile:
None
Admin Tags:
None
Lsp-self-ping: Config: inherit, Oper: Disabled, TimedOutCnt: 0 OamNoRsc: 0
Path "path-5-1-2" LspId 56324 LspPathIndex 1 PathType Primary ActivePath No
RowStatus Active LastChange 000 00:15:50.820
AdminState Up OperState Down OperStateChange 000 00:01:07.400
LspPath FsmState LSP_PATH_S_DOWN Flags 0x0 miscFlags 0x2

```

```

RetryAttempts 2 RetryInterval 30 NextRetryIn 32secs
FailNode 192.0.2.5 FailCode noCspfRouteToDestination
Class Type 0 SetupPri 7 HoldPri 0 Pref 0 HopLimit 255 BW 0Mbps
TotIgpCost 0 OperMetric 16777215 MTU 0
BFD Disabled Template n/a PingInterval 60 WaitForUp 4sec
Degraded No
OldMBBPathsCleanedUp Yes
Path "path-5-4-3-2" LspId 56322 LspPathIndex 2 PathType Standby ActivePath Yes
RowStatus Active LastChange 000 00:15:50.820
AdminState Up OperState Up OperStateChange 000 00:15:50.810
TE Computed Hop List:
Hop[1] IngIp 192.0.2.5 IngLnkId 0 EgrIp 192.168.45.2 EgrLnkId 0 RtrId 192.0.2.5 Flag 0x0
Hop[2] IngIp 192.168.45.1 IngLnkId 0 EgrIp 192.168.34.2 EgrLnkId 0 RtrId 192.0.2.4 Flag
0x0
Hop[3] IngIp 192.168.34.1 IngLnkId 0 EgrIp 192.168.23.2 EgrLnkId 0 RtrId 192.0.2.3 Flag
0x0
Hop[4] IngIp 192.168.23.1 IngLnkId 0 EgrIp 192.0.2.2 EgrLnkId 0 RtrId 192.0.2.2 Flag 0x0
LspPath FsmState LSP_PATH_S_UP Flags 0x0 miscFlags 0x2
RetryAttempts 0 RetryInterval 30 NextRetryIn 0secs
FailNode 0.0.0.0 FailCode noError
Class Type 0 SetupPri 7 HoldPri 0 Pref 255 HopLimit 255 BW 0Mbps
TotIgpCost 30 OperMetric 30 MTU 8982
SRLG Disabled SRLGDisjoint No
BFD Disabled Template n/a PingInterval 60 WaitForUp 4sec
Degraded No
Oper Values:
Class Type 0 SetupPri 7 HoldPri 0 HopLimit 255 BW 0Mbps
RecordRoute RecordLabel No Adspec
No PropagateAdminGroup Exclude 0x00000000 Include 0x00000000
No FRR
Metric 30 CSPP No Least Fill Intra-area
NegotiatedEntropyLabel Disabled
PCE-Computed No PCE-Reported No PCE-Controlled No
BFD State Up InitTime 0d 00:01:07 UpTime 0d 00:01:03
OldMBBPathsCleanedUp Yes

Total Ingress LSP Count : 1

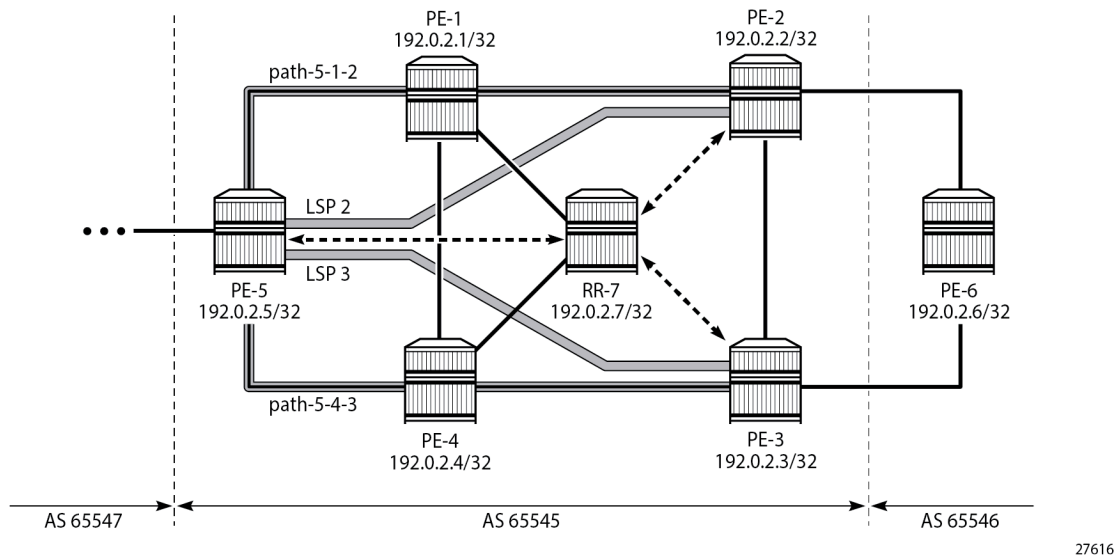
```

The secondary path, *path-5-4-3-2*, becoming active does not result in a change of the BGP next-hop. Traffic continues to flow from PE-5 to PE-6 via LSP *lsp-1*, but now via the lower path. The BFD failure action failover combined with standby secondary paths can help detect failures faster, with minimal traffic loss, which is especially useful in larger domains, or when the LSP passes through multiple domains.

## Failure action down

[Figure 11: Topology for failure action down](#) shows the topology used for failure action down. BGP shortcuts are defined in AS 65545 running between the ASBRs PE-5, PE-2, and PE-3. A first shortcut is offered through an RSVP-TE LSP called *lsp-2*, with a single path from PE-5 via PE-1 to PE-2; the second shortcut is offered through another RSVP-TE LSP called *lsp-3*, with a single path from PE-5 via PE-4 to PE-3. When LSP *lsp-2* fails and the failure gets detected by BFD, LSP *lsp-2* is unavailable in the TTM and traffic starts using *lsp-3*, implying a change of the BGP next hop; this scenario being an edge prefix-independent convergence (PIC) scenario. See the "BGP Fast Reroute" chapter in *7450 ESS, 7750 SR, 7950 XRS, and VSR Unicast Routing Protocols Advanced Configuration Guide for MD CLI* for more information about edge PIC.

Figure 11: Topology for failure action down



The initial configuration includes:

- Cards, MDAs, and ports
- Router interfaces
- IS-IS as IGP on all interfaces (alternatively, OSPF can be used), with traffic engineering enabled
- MPLS and RSVP-TE enabled on all interfaces
- BGP configured, with RR-7 being the route reflector for clients PE-2, PE-3, and PE-5 in AS 65545, and PE-6 located in AS 65546 and connected to PE-2 and PE-3. PE-6 exports prefix 192.0.2.6/32 to PE-2 and PE-3.

The LSPs from [Figure 11: Topology for failure action down](#) are configured with a single path, as follows. The paths referred to from these LSPs are fully strict paths, using interface IP addresses. Both *lsp-2* and *lsp-3* have BFD enabled with failure action down.

```
# on PE-5
configure {
  router "Base" {
    mpls {
      path "path-5-1-2" {
        admin-state enable
        hop 10 {
          ip-address 192.168.15.1
          type strict
        }
        hop 20 {
          ip-address 192.168.12.2
          type strict
        }
      }
      path "path-5-4-3" {
        admin-state enable
        hop 10 {
          ip-address 192.168.45.1
          type strict
        }
      }
    }
  }
}
```



```

    }
    hop 20 {
        ip-address 192.168.34.1
        type strict
    }
}
lsp "lsp-2" {
    admin-state enable
    type p2p-rsvp
    to 192.0.2.2
    path-computation-method local-cspf
    bfd {
        bfd-liveness true
        bfd-template "bfdt-1"
        failure-action down
    }
    primary "path-5-1-2" {
    }
}
lsp "lsp-3" {
    admin-state enable
    type p2p-rsvp
    to 192.0.2.3
    path-computation-method local-cspf
    bfd {
        bfd-liveness true
        bfd-template "bfdt-1"
        failure-action down
    }
    primary "path-5-4-3" {
    }
}
}

```

The details of the LSP show the configured failure action, as follows:

```

[/]
A:admin@PE-5# show router mpls lsp "lsp-2" detail

=====
MPLS LSPs (Originating) (Detail)
=====
Legend :
+ - Inherited
=====
-----
Type : Originating
-----
LSP Name      : lsp-2
LSP Type      : RegularLsp          LSP Tunnel ID      : 2
LSP Index     : 2                  TTM Tunnel Id      : 2
From          : 192.0.2.5
To            : 192.0.2.2
Adm State     : Up                  Oper State          : Up
LSP Up Time   : 0d 00:03:42         LSP Down Time      : 0d 00:00:00
Transitions   : 1                  Path Changes        : 1
Retry Limit    : 0                  Retry Timer         : 30 sec
Signaling     : RSVP               Resv. Style         : SE
Hop Limit     : 255                 Negotiated MTU      : 8982
Adaptive      : Enabled             ClassType           : 0
FastReroute   : Disabled            Oper FR              : Disabled
PathCompMethod : local-cspf         ADSPEC              : Disabled
FallbkPathComp : not-applicable
Metric        : N/A                 Metric Type         : igp

```

```

Load Bal Wt      : N/A
Include Grps    :
None
Least Fill      : Disabled
BFD Template    : bfdt-1
BFD Enable      : True
WaitForUpTimer  : 4
ClassForwarding : Disabled
Exclude Grps    :
None
Soft Preemption : Enabled
BFD Ping Intvl  : 60
BFD Failure-action : Down

Revert Timer    : Disabled
Entropy Label   : Enabled+
Negotiated EL   : Disabled
Auto BW         : Disabled
LdpOverRsvp    : Enabled
VprnAutoBind    : Enabled
IGP Shortcut    : Enabled
IGP LFA         : Disabled
AllowSrOverSrte : Disabled
BGPTransTun    : Enabled
Oper Metric     : 20
Prop Adm Grp    : Disabled
PCE Report      : Disabled+
PCE Control     : Disabled
Path Profile    : None
Admin Tags      : None
Lsp Self Ping   : Disabled+
SelfPingOAMFail* : 0

Primary(a)      : path-5-1-2
Bandwidth       : 0 Mbps
Up Time         : 0d 00:03:42

BGP Shortcut    : Enabled
IGP Rel Metric  : Disabled

Self Ping Timeouts : 0

=====
* indicates that the corresponding row element may have been truncated.
  
```

Multiple BGP paths are available out of PE-5 to reach PE-6, as follows. The path via PE-2 is the currently active path, the path via PE-3 is the backup path.

```

[/]
A:admin@PE-5# show router route-table protocol bgp alternative

=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]
Next Hop[Interface Name]
Alt-NextHop
Type Proto Age Metric Pref
Alt-Metric
-----
192.0.2.6/32 Remote BGP 00h03m19s 170
192.0.2.6/32 (Backup) Remote BGP 00h03m19s 170
192.0.2.2 (tunneled:RSVP:2) 20
192.0.2.3 (tunneled:RSVP:3) 20
-----
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
Backup = BGP backup route
LFA = Loop-Free Alternate nexthop
S = Sticky ECMP requested
=====
  
```

The first of the following BFD sessions is running over the active path:

```
[/]
```

```
A:admin@PE-5# 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
-----
lsp-2::path-5-1-2          Up        143       135
192.0.2.2                5          2000      2000
  rsvpLsp              central    N/A       N/A
  192.0.2.5
lsp-3::path-5-4-3          Up         145       135
  192.0.2.3            5          2000      2000
  rsvpLsp              central    N/A       N/A
  192.0.2.5
-----
No. of BFD sessions: 2
=====
```

The BFD session running over the active path of the *lsp-2* LSP is also indicated in the output of the following **tools** command:

```
[/]
A:admin@PE-5# tools dump router mpls lspinfo "lsp-2" detail
LSP "lsp-2" LspIdx 2 LspType Dynamic State LSPS_UP Flags 0x2000
AdminState Up OperState Up RowStatus Active
From N/A To 192.0.2.2
NumPaths 1 NumSdps 0 NumCBFSdps 0 NumFltrEntries 0
ActivePath lsp-2::path-5-1-2(LspId 56832)
HoldTimeRemaining 0secs ClassType 0 SoftPreemption TRUE Metric 0 OperMetric 20
LDPoRsvp Include VprnAutoBind Include IgpShortCut Include BgpShortCut Include
BgpTransTunnel Include IpShCutTtlPropLocal TRUE IpShCutTtlPropTans TRUE
RelativeMetricOffset 2147483647 MTU 8982 InUseByLdp FALSE TTMPref 7
EntropyLabel inherit OperEntropyLabel enable NegEntropyLabel disable
ClassForwarding: Disabled
BFD Enabled Template bfdt-1 PingInterval 60 FailureAction down WaitForUp 4sec
PCE Report: Disabled PCE Control: Disabled
Path Profile:
  None
Admin Tags:
  None
Lsp-self-ping: Config: inherit, Oper: Disabled, TimedOutCnt: 0 OamNoRsc: 0
Path "path-5-1-2" LspId 56832 LspPathIndex 1 PathType Primary ActivePath Yes
  RowStatus Active LastChange 000 00:04:25.220
  AdminState Up OperState Up OperStateChange 000 00:04:25.220
  TE Computed Hop List:
    Hop[1] IngIp 192.0.2.5 IngLnkId 0 EgrIp 192.168.15.2 EgrLnkId 0 RtrId 192.0.2.5 Flag 0x0
    Hop[2] IngIp 192.168.15.1 IngLnkId 0 EgrIp 192.168.12.1 EgrLnkId 0 RtrId 192.0.2.1 Flag
0x0
    Hop[3] IngIp 192.168.12.2 IngLnkId 0 EgrIp 192.0.2.2 EgrLnkId 0 RtrId 192.0.2.2 Flag 0x0
  LspPath FsmState LSP_PATH_S_UP Flags 0x0 miscFlags 0x2
  RetryAttempts 0 RetryInterval 30 NextRetryIn 0secs
  FailNode 0.0.0.0 FailCode noError
  Class Type 0 SetupPri 7 HoldPri 0 Pref 0 HopLimit 255 BW 0Mbps
  TotIgpCost 20 OperMetric 20 MTU 8982
  BFD Disabled Template n/a PingInterval 60 WaitForUp 4sec
```

```

Degraded No
Oper Values:
  Class Type 0 SetupPri 7 HoldPri 0 HopLimit 255 BW 0Mbps
  RecordRoute RecordLabel No Adspec
  No PropagateAdminGroup Exclude 0x00000000 Include 0x00000000
  No FRR
  Metric 20 CSPF No Least Fill Intra-area
  NegotiatedEntropyLabel Disabled
  PCE-Computed No PCE-Reported No PCE-Controlled No
  BFD State Up InitTime 0d 00:04:25 UpTime 0d 00:04:20
  OldMBBPathsCleanedUp Yes

Total Ingress LSP Count      : 1
    
```

Emulating a path failure by bringing down port 1/1/c2/1 on PE-2 brings the primary path in the *lsp-2* LSP down, as follows:

```

[/]
A:admin@PE-5# show router mpls lsp "lsp-2" path

=====
MPLS LSP lsp-2 Path
=====
-----
LSP Name      : lsp-2
From          : 192.0.2.5
To            : 192.0.2.2
Adm State     : Up
Oper State    : Down
-----
Path Name      Next Hop      Type      Out I/F  Adm  Opr
-----
path-5-1-2    n/a                Primary   n/a      Up   Dwn
=====
    
```

The only remaining BFD session is now the BFD session for the *lsp-3* LSP with next hop 192.0.2.3, as follows:

```

[/]
A:admin@PE-5# 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
-----
lsp-3::path-5-4-3      Up        188        173
192.0.2.3              5          2000       2000
  rsvpLsp          central    N/A        N/A
  192.0.2.5
-----
No. of BFD sessions: 1
=====
    
```

BGP traffic is diverted into *lsp-3*, and the BGP next hop is 192.0.2.3, as follows:

```
[/]
A:admin@PE-5# show router route-table protocol bgp alternative
=====
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 BGP     00h00m49s 170
   192.0.2.3 (tunneled:RSVP:3)          20
-----
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
=====
```



**Note:** Though it is possible to configure **failure-action down** for LSPs with secondary paths, no failover to any secondary path will take place. When the active path goes down, the entire LSP is registered as unusable in the TTM, regardless of any secondary paths. Because a BFD session running on a secondary unused path can be confusing to operators and is taking up resources, Nokia recommends defining the LSPs to only use a single path when failure action down is configured, as in the example.

## Conclusion

The BFD failure action failover or down can help detect failures faster with minimal traffic loss on switchover, which is especially useful in larger domains or when the LSP passes through multiple domains.

# DiffServ Traffic Engineering

This chapter provides information about DiffServ traffic engineering.

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

## Applicability

This chapter was initially written for SR OS Release 14.0.R2. The CLI in the current edition corresponds to SR OS Release 23.3.R1.

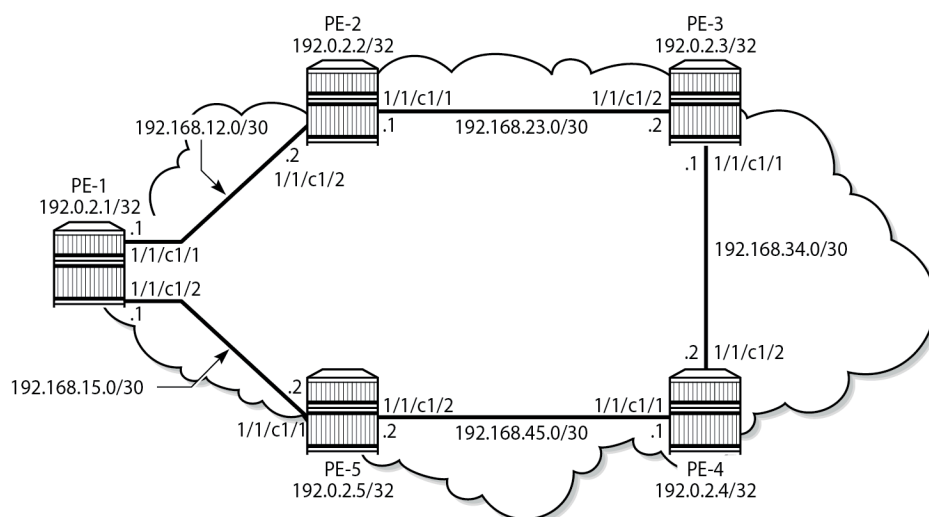
## Overview

Differentiated Services (DiffServ) is a mechanism to classify and manage network traffic to provide Quality of Service (QoS). DiffServ Traffic Engineering (DiffServ TE) reserves bandwidth for Label Switched Paths (LSPs) on ReSource reserVation Protocol (RSVP) interfaces on a per TE class basis.

## Example Topology

[Figure 12: Example topology](#) shows the example topology that contains five 7750 SRs in a ring topology.

Figure 12: Example topology



38584

## Definitions

The following definitions are used in this chapter:

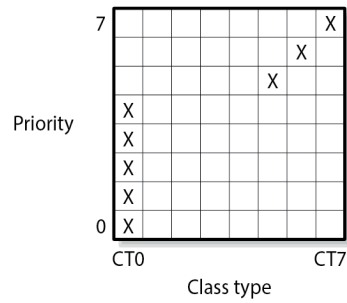
- Forwarding Classes (FCs) classify micro-flows into macro-flows. FCs can be mapped to Class Types (CTs).
- A CT is macro-flow crossing a link governed by a specific Bandwidth Constraint (BC). The BC is defined on a per-link and per-CT basis. A CT can be considered as a network-wide FC, advertised by the IGP (OSPF opaque link state advertisement (LSA), IS-IS TE Type Length Value (TLV)).
  - IGP-TE can reserve bandwidth per CT on a link (BC).
  - RSVP-TE can reserve bandwidth per LSP path, based on TE class.
- A TE class is a combination of a CT and a preemption priority.

There are eight FCs that can be mapped to CTs. The CTs range from CT0 (lowest) to CT7 (highest) and each gets a percentage of the bandwidth of the link. Each CT has eight different priority levels that are used for preemption. Even though there are 64 different potential combinations of CT and priority, only eight different combinations can be defined for TE classes. All CTs and priorities must be manually configured.

The system allows up to eight TE classes to be configured. The more TE classes are defined, the more RSVP LSPs need to be configured for each service. TE classes are consistently configured on all TE-aware Label Switching Routers (LSRs) throughout the network and advertised through the IGP.

The following shows a DiffServ TE configuration where each CT can reserve up to 10% of the maximum reservable bandwidth meaning that 20% of the bandwidth is not allocated to any CT. MPLS needs to be shut down when DiffServ TE is configured. [Figure 13: Mapping of TE classes](#) shows the mapping of the TE classes.

Figure 13: Mapping of TE classes



25847

```
# on PE-1:
configure {
  router "Base" {
    rsvp {
      preemption-timer 5
      diffserv-te {
        admission-control-model mam
        class-type-bw {
          ct0 10
          ct1 10
          ct2 10
          ct3 10
          ct4 10
          ct5 10
          ct6 10
          ct7 10
        }
        te-class 0 {
          class-type 0
          priority 0
        }
        te-class 1 {
          class-type 0
          priority 1
        }
        te-class 2 {
          class-type 0
          priority 2
        }
        te-class 3 {
          class-type 0
          priority 3
        }
        te-class 4 {
          class-type 0
          priority 4
        }
        te-class 5 {
          class-type 5
          priority 5
        }
        te-class 6 {
          class-type 6
          priority 6
        }
        te-class 7 {
```



```

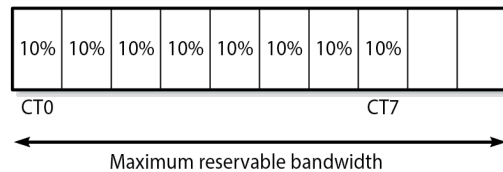
class-type 7
priority 7
}
fc af {
class-type 0
}
fc be {
class-type 0
}
fc ef {
class-type 5
}
fc h1 {
class-type 6
}
fc h2 {
class-type 0
}
fc l1 {
class-type 0
}
fc l2 {
class-type 0
}
fc nc {
class-type 7
}
}

```

This configuration is applied on all TE-aware LSRs for consistency.

Figure 14: Bandwidth reservation for the CTs shows the bandwidth reservation for the CTs.

Figure 14: Bandwidth reservation for the CTs



25848

Eight TE classes are defined, each with a different priority, using four CTs (0, 5, 6 and 7). There is no need to assign any bandwidth to the unused CTs. The configuration in this example is not a recommendation. In this example, the BC for each CT is 10% of the maximum reservable bandwidth.

On each node, the RSVP status shows whether DiffServ TE is enabled and how it is configured, as follows:

```

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

=====
RSVP Status
=====
Admin Status      : Up                Oper Status      : Up
Keep Multiplier   : 3                  Refresh Time     : 30 sec
Message Pacing    : Disabled          Pacing Period    : 100 msec
Max Packet Burst  : 650 msgs         Refresh Bypass   : Disabled
Rapid Retransmit  : 5 hmsec          Rapid Retry Limit: 3
Graceful Shutdown : Disabled          SoftPreemptionTimer: 5 sec

```

```

GR Max Recovery      : 300 sec          GR Max Restart      : 120 sec
Implicit Null Label  : Disabled          Node-id in RR0     : Exclude
P2P Merge Point Ab* : Disabled          P2MP Merge Point A* : Disabled
Auth Over Bypass    : Disabled
DiffServTE AdmModel: Mam
Percent Link Bw CT0 : 10                Entropy Label      : Disabled
Percent Link Bw CT1 : 10                Percent Link Bw CT4 : 10
Percent Link Bw CT2 : 10                Percent Link Bw CT5 : 10
Percent Link Bw CT3 : 10                Percent Link Bw CT6 : 10
TE0 -> Class Type   : 0                Percent Link Bw CT7 : 10
TE1 -> Class Type   : 0                Priority            : 0
TE2 -> Class Type   : 0                Priority            : 1
TE3 -> Class Type   : 0                Priority            : 2
TE4 -> Class Type   : 0                Priority            : 3
TE5 -> Class Type   : 5                Priority            : 4
TE6 -> Class Type   : 6                Priority            : 5
TE7 -> Class Type   : 7                Priority            : 6
FCName              : af                Priority            : 7
FCName              : be                Class Type         : 0
FCName              : ef                Class Type         : 0
FCName              : h1                Class Type         : 5
FCName              : h2                Class Type         : 6
FCName              : l1                Class Type         : 0
FCName              : l2                Class Type         : 0
FCName              : nc                Class Type         : 7
IgpThresholdUpdate  : Disabled
Up Thresholds(%)    : 0 15 30 45 60 75 80 85 90 95 96 97 98 99 100
Down Thresholds(%)  : 100 99 98 97 96 95 90 85 80 75 60 45 30 15 0
Update Timer        : N/A
Update on CAC Fail  : Disabled

```

=====  
\* indicates that the corresponding row element may have been truncated.

The OSPF LSAs contain BC information, as shown in the following output taken from PE-1:

```

[/]
A:admin@PE-1# show router ospf opaque-database adv-router 192.0.2.3 detail
=====
Rtr Base OSPFv2 Instance 0 Opaque Link State Database (type: All) (detail)
=====
-----
Opaque LSA
-----
Area Id          : 0.0.0.0                Adv Router Id    : 192.0.2.3
Link State Id    : 1.0.0.1                LSA Type        : Area Opaque
Sequence No      : 0x80000002             Checksum        : 0x9a28
Age              : 232                    Length          : 28
Options         : E
Advertisement    : Traffic Engineering
ROUTER-ID TLV   (0001) Len 4 : 192.0.2.3
-----
Opaque LSA
-----
Area Id          : 0.0.0.0                Adv Router Id    : 192.0.2.3
Link State Id    : 1.0.0.2                LSA Type        : Area Opaque
Sequence No      : 0x8000000a             Checksum        : 0xe962
Age              : 57                     Length          : 164
Options         : E
Advertisement    : Traffic Engineering
LINK INFO TLV   (0002) Len 140 :
  Sub-TLV: 1     Len: 1     LINK_TYPE : 1
  Sub-TLV: 2     Len: 4     LINK_ID  : 192.0.2.2

```

```

Sub-TLV: 3      Len: 4      LOC_IP_ADDR  : 192.168.23.2
Sub-TLV: 4      Len: 4      REM_IP_ADDR  : 192.168.23.1
Sub-TLV: 5      Len: 4      TE_METRIC    : 10
Sub-TLV: 6      Len: 4      MAX_BDWTH   : 10000000 Kbps
Sub-TLV: 7      Len: 4      RSRVBL_BDWTH : 10000000 Kbps
Sub-TLV: 8      Len: 32     UNRSRVD_CLS0 :
P0: 1000000 Kbps P1: 1000000 Kbps P2: 1000000 Kbps P3: 1000000 Kbps
P4: 1000000 Kbps P5: 1000000 Kbps P6: 1000000 Kbps P7: 1000000 Kbps
Sub-TLV: 9      Len: 4      ADMIN_GROUP  : 0 None
Sub-TLV: 17     Len: 36     TELK_BW_CONST:
  BW Model : MAM
  BC0: 1000000 Kbps BC1: 1000000 Kbps BC2: 1000000 Kbps BC3: 1000000 Kbps
  BC4: 1000000 Kbps BC5: 1000000 Kbps BC6: 1000000 Kbps BC7: 1000000 Kbps
---snip---
=====

```

In the preceding output, only the output for the interface between PE-2 and PE-3 is shown; the output is similar for the other interfaces. On each interface between nodes, there are eight BCs for the eight CTs: from BC0 to BC7. In this example, each of the BCs has the same constraint of 1 Gb/s, which corresponds to 10% of the 10 Gb/s interfaces. As long as no LSP is configured with a CT and a bandwidth, no bandwidth is reserved. The BCs for an interface, such as the interface between PE-1 and PE-2, can be shown as follows:

```

[/]
A:admin@PE-1# show router rsvp interface "int-PE-1-PE-2" detail

=====
RSVP Interface (Detailed) : int-PE-1-PE-2
=====
-----
Interface : int-PE-1-PE-2
-----
Interface      : int-PE-1-PE-2
Port ID       : 1/1/cl/1
Admin State   : Up
Active Sessions : 0
Total Sessions : 0
Subscription  : 100 %
Total BW      : 10000 Mbps
Hello Interval : 3000 ms
Key Type Auth : Disabled
Keychain Auth : Disabled
Auth Rx Seq Num : n/a
Auth Tx Seq Num : n/a
Refresh Reduc. : Disabled
Bfd Enabled    : No
ImplicitNullLabel : Disabled*
Oper State    : Up
Active Resvs  : 0
Port Speed    : 10000 Mbps
Aggregate     : Dsabl
Hello Timeouts : 0

Percent Link Bandwidth for Class Types*
Link Bw CT0   : 10
Link Bw CT1   : 10
Link Bw CT2   : 10
Link Bw CT3   : 10
Link Bw CT4   : 10
Link Bw CT5   : 10
Link Bw CT6   : 10
Link Bw CT7   : 10

Bandwidth Constraints for Class Types (Kbps)
BC0           : 1000000
BC1           : 1000000
BC2           : 1000000
BC3           : 1000000
BC4           : 1000000
BC5           : 1000000
BC6           : 1000000
BC7           : 1000000

Bandwidth for TE Class Types (Kbps)
TE0-> Resv. Bw : 0
Unresv. Bw    : 1000000

```

```

TE1-> Resv. Bw : 0           Unresv. Bw : 1000000
TE2-> Resv. Bw : 0           Unresv. Bw : 1000000
TE3-> Resv. Bw : 0           Unresv. Bw : 1000000
TE4-> Resv. Bw : 0           Unresv. Bw : 1000000
TE5-> Resv. Bw : 0           Unresv. Bw : 1000000
TE6-> Resv. Bw : 0           Unresv. Bw : 1000000
TE7-> Resv. Bw : 0           Unresv. Bw : 1000000

IGP Update
Up Thresholds(%) : 0 15 30 45 60 75 80 85 90 95 96 97 98 99 100 *
Down Thresholds(%) : 100 99 98 97 96 95 90 85 80 75 60 45 30 15 0 *
IGP Update Pending : No
Next Update : N/A

No Neighbors.
* indicates inherited values
=====

```

In this example, all BCs for the CTs are equal to 1 Gb/s, which is 10% of the maximum reservable bandwidth of 10 Gb/s. Currently no bandwidth is reserved for any of the TE classes. The unreserved bandwidth equals 1 Gb/s for TE0 through TE7.

The maximum bandwidth that can be allocated depends on the bandwidth of the link and the subscription percentage. When the subscription percentage is doubled to 200%, the BCs are doubled too, as follows:

```

# on PE-1:
configure {
  router "Base" {
    rsvp {
      interface "int-PE-1-PE-2" {
        subscription 200
      }
    }
  }
}

```

```

[/]
A:admin@PE-1# show router rsvp interface "int-PE-1-PE-2" detail

```

```

=====
RSVP Interface (Detailed) : int-PE-1-PE-2
=====
-----
Interface : int-PE-1-PE-2
-----
---snip---
Percent Link Bandwidth for Class Types*
Link Bw CT0      : 10           Link Bw CT4      : 10
Link Bw CT1      : 10           Link Bw CT5      : 10
Link Bw CT2      : 10           Link Bw CT6      : 10
Link Bw CT3      : 10           Link Bw CT7      : 10

Bandwidth Constraints for Class Types (Kbps)
BC0      : 2000000           BC4      : 2000000
BC1      : 2000000           BC5      : 2000000
BC2      : 2000000           BC6      : 2000000
BC3      : 2000000           BC7      : 2000000

Bandwidth for TE Class Types (Kbps)
TE0-> Resv. Bw : 0           Unresv. Bw : 2000000
TE1-> Resv. Bw : 0           Unresv. Bw : 2000000
TE2-> Resv. Bw : 0           Unresv. Bw : 2000000
TE3-> Resv. Bw : 0           Unresv. Bw : 2000000
TE4-> Resv. Bw : 0           Unresv. Bw : 2000000
TE5-> Resv. Bw : 0           Unresv. Bw : 2000000

```

```
TE6-> Resv. Bw : 0 Unresv. Bw : 2000000
TE7-> Resv. Bw : 0 Unresv. Bw : 2000000
---snip---
```

The subscription percentage is restored to its default value of 100% as follows:

```
# on PE-1:
configure {
  router "Base" {
    rsvp {
      interface "int-PE-1-PE-2" {
        delete subscription
      }
    }
  }
}
```

## Bandwidth constraint models

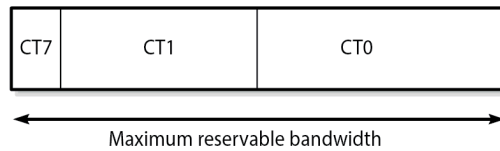
Two models are available for the bandwidth calculation that is required during the LSP setup: the Maximum Allocation Model (MAM) and the Russian Doll Model (RDM). [Table 1: Comparison bandwidth constraint models](#) shows a comparison between the two models.

Table 1: Comparison bandwidth constraint models

Maximum Allocation Model (MAM)	Russian Doll Model (RDM)
Fixed BC per CT. No bandwidth sharing between CTs.	Maps one BC to one or more CTs. Lower CTs are allowed to reserve from the unused bandwidth of the pools defined for higher CTs.
Achieves isolation between CTs and guaranteed bandwidth to CTs without the need for preemption.	No isolation between CTs. Requires preemption to guarantee bandwidth to CTs other than the premium.
Bandwidth may be wasted.	Efficient use of bandwidth.
Easy to manage.	More complex.

[Figure 15: Bandwidth reservation in Maximum Allocation Model for three CTs](#) shows the reserved bandwidth for the different class types according to the MAM model. In this example, there is only bandwidth reserved for CT0, CT1, and CT2.

Figure 15: Bandwidth reservation in Maximum Allocation Model for three CTs

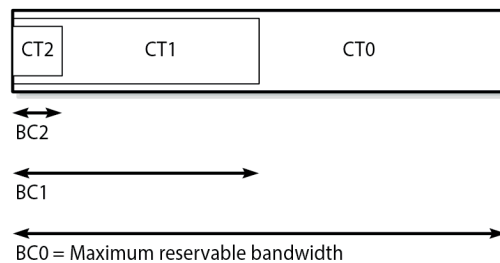


25849

Bandwidth that is reserved for a specific CT cannot be used by any other CT. Therefore, bandwidth may be wasted.

The Russian Doll Model is more flexible: when CT1 has some spare bandwidth that may be used by CT0, this is allowed. Depending on the configured setup priority and hold priority, this may be reversed when CT1 requires the bandwidth. The bandwidth reservation in the Russian Doll Model is shown in [Figure 16: Bandwidth reservation in Russian Doll Model for three CTs](#).

Figure 16: Bandwidth reservation in Russian Doll Model for three CTs



25850

## Backup class types

The main CT is defined at LSP level or primary path level. The main CT is used at the first attempt for the initial establishment and re-signal Make-Before-Break (MBB) of the LSP primary path. Re-signaling of the LSP path can be triggered manually or timer-based. Subsequent retries use the backup CT, which is configured on the primary path level. Secondary paths are always signaled with the main CT. There is no verification whether the backup CT is lower than the main CT. This applies to CSPF and non-CSPF LSPs. An example of an LSP with main CT1 and backup CT0 is as follows:

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      path "dyn" {
        admin-state enable
      }
    }
    lsp "LSP-PE-1-PE-3-withBackupCT" {
      admin-state enable
      type p2p-rsvp
      to 192.0.2.3
      path-computation-method local-cspf
      class-type 1
      primary "dyn" {
        bandwidth 50
      }
    }
  }
}
```

```

        backup-class-type 0
        priority {
            setup-priority 4
            hold-priority 4
        }
    }
}

```

Possible triggers for using the backup CT are:

- Local interface failure or control plane failure (hello timeout)
- Received Resv message with the local-protection-in-use flag set (global revertive trigger)
- Received Patherr message with Fast ReRoute (FRR) protection active notification (global revertive trigger)
- Received Patherr message with error code 34 (Reroute) and value 1 (soft preemption trigger)
- Received Patherr message with Preemption pending flag set (soft preemption trigger)
- Received ResvTear message

When the reservable bandwidth for a CT (including the bandwidth for the inner dolls in case of RDM) is insufficient, this does not trigger the backup CT to be used. If possible, an alternate path is used for the LSP requiring this bandwidth.

## Priorities

Two different priorities are linked to an LSP in a range from 0 to 7, where 0 is the highest priority and 7 the lowest. These values are important when preemption occurs, as follows:

```

# on PE-1:
A:admin@PE-1# configure {
    router "Base" {
        mpls {
            lsp "LSP-PE-1-PE-3" {
                primary "dyn" {
                    priority ?

priority

                hold-priority          - Priority of an LSP session at preemption action
                setup-priority         - Priority when insufficient bandwidth for LSP setup

where:
hold-priority <number>
<number> - <0..7>
Default - 0
and
setup-priority <number>
<number> - <0..7>
Default - 7

```

The following shows an LSP with both priorities equal to 4:

```

# on PE-1:
configure {
    router "Base" {
        mpls {
            lsp "LSP-PE-1-PE-3" {

```

```

admin-state enable
type p2p-rsvp
to 192.0.2.3
path-computation-method local-cspf
primary "dyn" {
    bandwidth 50
    priority {
        setup-priority 4
        hold-priority 4
    }
}

```

- The first priority in the configuration is the setup priority. When an LSP is signaled and there is not enough bandwidth available on the egress Label Edge Router (eLER) or LSR, the LSP can preempt an established LSP with a hold priority lower than this setup priority. For a setup priority of 4, existing LSPs with a hold priority of 5, 6, or 7 can be preempted in case of insufficient bandwidth.
- The second priority in the configuration is the hold priority. When this LSP is established and a new LSP needs to be established, and there is insufficient bandwidth, this LSP can only be preempted by an LSP with a higher setup priority than this hold priority. For a hold priority of 4, the LSP can be preempted by any LSP with a setup priority of 0, 1, 2, or 3.

The default values are a setup priority of 7 and a hold priority of 0. A low setup priority of 7 means the LSP cannot preempt any LSP. A high hold priority of 0 implies that the LSP cannot be preempted by any other LSP.

The setup priority needs to be lower than or equal to the hold priority to avoid preemption loops. Nokia recommends that the setup priority and the hold priority are set to equal values.

Bandwidth, CT information, and priorities are shown as follows:

```

[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-3" path detail
=====
MPLS LSP LSP-PE-1-PE-3 Path (Detail)
=====
Legend :
  @ - Detour Available          # - Detour In Use
  b - Bandwidth Protected      n - Node Protected
  s - Soft Preemption
  S - Strict                   L - Loose
  A - ABR                      + - Inherited
=====
-----
LSP LSP-PE-1-PE-3
Path dyn
-----
LSP Name      : LSP-PE-1-PE-3
From          : 192.0.2.1
To            : 192.0.2.3
Admin State   : Up
Path Name     : dyn
Path LSP ID   : 43008
Path Admin    : Up
Out Interface : 1/1/c1/1
---snip---
Neg MTU       : 1564
Bandwidth     : 50 Mbps
Hop Limit     : 255
Record Route  : Record
Record Label  : Record
Setup Priority : 4
Oper State    : Up
Path Type     : Primary
Path Oper     : Up
Out Label     : 524287
Oper MTU      : 1564
Oper Bandwidth : 50 Mbps
Oper HopLimit : 255
Oper Record Route : Record
Oper Record Label : Record
Oper SetupPriority: 4

```



```

Hold Priority      : 4
Class Type         : 0
Backup CT         : None
MainCT Retry      : n/a
  Rem              :
MainCT Retry      : 0
  Limit           :
---snip---
=====

```

When the LSP is being established, the path message contains the setup and hold priorities, and the required bandwidth, as follows:

```

# on PE-1:
debug
  router
    rsvp
      packet
        path detail
      exit

```

```

1 2023/04/12 21:29:24.744 CEST MINOR: DEBUG #2001 Base RSVP
"RSVP: PATH Msg
Send PATH From:192.0.2.1, To:192.0.2.3
      TTL:255, Checksum:0xe791, Flags:0x0
Session  - EndPt:192.0.2.3, TunnId:2, ExtTunnId:192.0.2.1
SessAttr - Name:LSP-PE-1-PE-3::dyn
          SetupPri:4, HoldPri:4, Flags:0x46
RSVPHop  - Ctype:1, Addr:192.168.12.1, LIH:2
TimeValue - RefreshPeriod:30
SendTempl - Sender:192.0.2.1, LspId:43008
SendTSpec - Ctype:QOS, CDR:50.000 Mbps, PBS:50.000 Mbps, PDR:infinity
          MPU:20, MTU:1564
LabelReq - IfType:General, L3ProtID:2048
RR0      - IpAddr:192.168.12.1, Flags:0x0
ERO      - IPv4Prefix 192.168.12.2/32, Strict
          IPv4Prefix 192.168.23.2/32, Strict
"

```

As soon as the LSP is established, the bandwidth is reserved on the interface int-PE-1-PE-2 on PE-1 in TE class 4 (configured as a combination of CT0 and priority 4), as follows:

```

[/]
A:admin@PE-1# show router rsvp interface "int-PE-1-PE-2" detail

=====
RSVP Interface (Detailed) : int-PE-1-PE-2
=====
-----
Interface : int-PE-1-PE-2
-----
Interface      : int-PE-1-PE-2
Port ID       : 1/1/cl/1
Admin State   : Up
Active Sessions : 1
Total Sessions : 1
Subscription  : 100 %
Total BW      : 10000 Mbps
---snip---
Percent Link Bandwidth for Class Types*
Link Bw CT0   : 10
Link Bw CT4   : 10

```

```

Link Bw CT1      : 10          Link Bw CT5      : 10
Link Bw CT2      : 10          Link Bw CT6      : 10
Link Bw CT3      : 10          Link Bw CT7      : 10

Bandwidth Constraints for Class Types (Kbps)
BC0              : 1000000      BC4              : 1000000
BC1              : 1000000      BC5              : 1000000
BC2              : 1000000      BC6              : 1000000
BC3              : 1000000      BC7              : 1000000

Bandwidth for TE Class Types (Kbps)
TE0-> Resv. Bw   : 0            Unresv. Bw      : 1000000
TE1-> Resv. Bw   : 0            Unresv. Bw      : 1000000
TE2-> Resv. Bw   : 0            Unresv. Bw      : 1000000
TE3-> Resv. Bw   : 0            Unresv. Bw      : 1000000
TE4-> Resv. Bw   : 50000       Unresv. Bw      : 950000
TE5-> Resv. Bw   : 0            Unresv. Bw      : 1000000
TE6-> Resv. Bw   : 0            Unresv. Bw      : 1000000
TE7-> Resv. Bw   : 0            Unresv. Bw      : 1000000
---snip---
=====

```

## Configuration

The example topology consists of five 7750 SRs in a ring topology, as shown in [Figure 12: Example topology](#).

### Initial Configuration

The nodes have the following initial configuration:

- Cards, MDAs, ports
- Router interfaces. For PE-1:

```

# on PE-1:
configure {
  router "Base" {
    interface "int-PE-1-PE-2" {
      port 1/1/c1/1
      ipv4 {
        primary {
          address 192.168.12.1
          prefix-length 30
        }
      }
    }
  }
  interface "int-PE-1-PE-5" {
    port 1/1/c1/2
    ipv4 {
      primary {
        address 192.168.15.1
        prefix-length 30
      }
    }
  }
}
interface "system" {
  ipv4 {

```

```
        primary {
            address 192.0.2.1
            prefix-length 32
        }
    }
}
```

- IGP: OSPF (alternatively, IS-IS could have been used) with TE enabled. For PE-1:

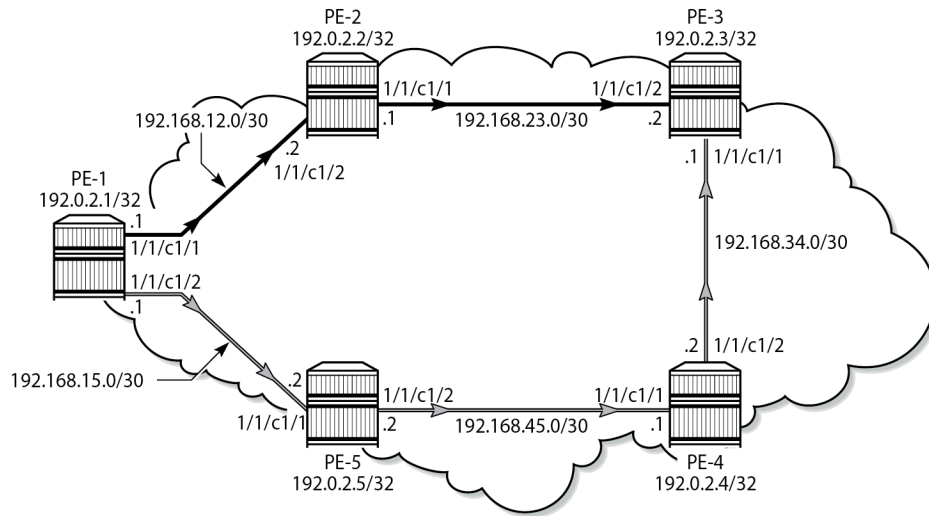
```
# on PE-1:
configure {
    router "Base" {
        ospf 0 {
            admin-state enable
            traffic-engineering true
            area 0.0.0.0 {
                interface "int-PE-1-PE-2" {
                    interface-type point-to-point
                }
                interface "int-PE-1-PE-5" {
                    interface-type point-to-point
                }
                interface "system" {
                }
            }
        }
    }
}
```

- MPLS and RSVP enabled on all interfaces. For PE-1:

```
# on PE-1:
configure {
    router "Base" {
        mpls {
            admin-state enable
            interface "int-PE-1-PE-2" {
            }
            interface "int-PE-1-PE-5" {
            }
        }
        rsvp {
            admin-state enable
            interface "int-PE-1-PE-2" {
            }
            interface "int-PE-1-PE-5" {
            }
        }
    }
}
```

LSPs are established from PE-1 to PE-3 with the short path via PE-2 as preferred. If insufficient bandwidth is available on the short path via PE-2, the longer path via PE-5 and PE-4 is taken, as shown in [Figure 17: Paths from PE-1 to PE-3](#).

Figure 17: Paths from PE-1 to PE-3



38585

Initially, the default BC model, which is MAM, is enabled. Different LSPs are created with different class type and priority. When an LSP is established, the bandwidth reservation is verified on the interfaces of PE-1. The same LSPs are later established for the second BC model, RDM, where the bandwidth reservation is more efficient. For simplicity, FRR is not enabled on the LSPs.

## Maximum Allocation Model

### Enable DiffServ MAM

The DiffServ TE configuration must be consistent on all the nodes in the setup. In this example, DiffServ TE is configured as follows:

```
# on PE-1:
configure {
  router "Base" {
    rsvp {
      diffserv-te {
        admission-control-model mam
        class-type-bw {
          ct0 50
          ct1 40
          ct7 10
        }
        te-class 0 {
          class-type 0
          priority 7
        }
        te-class 1 {
          class-type 0
          priority 4
        }
        te-class 2 {
          class-type 1
        }
      }
    }
  }
}
```

```

        priority 7
    }
    te-class 3 {
        class-type 1
        priority 4
    }
    te-class 4 {
        class-type 2
        priority 7
    }
    te-class 5 {
        class-type 2
        priority 2
    }
    fc af {
        class-type 1
    }
    fc be {
        class-type 0
    }
    fc nc {
        class-type 2
    }
}

```

The sum of bandwidth percentages can be lower than, but must not exceed 100%, as follows:

```

A:admin@PE-1# configure {
  router "Base" {
    rsvp {
      diffserv-te {
        admission-control-model mam
        class-type-bw {
          ct0 50
          ct1 50
          ct7 10
        }
      }
    }
  }
}
MINOR: RSVP #1005: configure router "Base" rsvp diffserv-te class-type-bw ct0 - Invalid
operation for RSVP instance - Total CT percent exceeds 100

```

Fewer than eight classes can be configured, as in this example.

In the example, only three CTs are used by the TE classes: CT0, CT1, and CT2. However, 10% of the maximum reservable bandwidth is allocated to CT7. Because the MAM model does not allow bandwidth allocated to a CT to be used by other CTs, only 90% of the bandwidth can be reserved: 50% to be divided between TE0 and TE1, and 40% to be divided between TE2 and TE3. TE4 and TE5 do not have any bandwidth allocated. The bandwidth allocated to CT7 is completely wasted. This is just an example, not a recommendation.

The same settings are repeated in the RDM model, where the bandwidth is not wasted. The following bandwidth information can be seen on any interface:

```

[/]
A:admin@PE-1# show router rsvp interface "int-PE-1-PE-2" detail

=====
RSVP Interface (Detailed) : int-PE-1-PE-2
=====
Interface : int-PE-1-PE-2
-----
---snip---

```

```

Percent Link Bandwidth for Class Types*
Link Bw CT0      : 50          Link Bw CT4      : 0
Link Bw CT1      : 40          Link Bw CT5      : 0
Link Bw CT2      : 0           Link Bw CT6      : 0
Link Bw CT3      : 0           Link Bw CT7      : 10

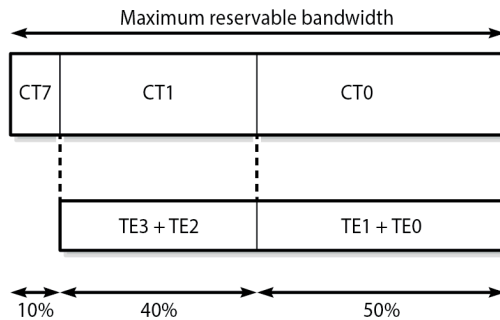
Bandwidth Constraints for Class Types (Kbps)
BC0              : 5000000      BC4              : 0
BC1              : 4000000      BC5              : 0
BC2              : 0            BC6              : 0
BC3              : 0            BC7              : 1000000

Bandwidth for TE Class Types (Kbps)
TE0-> Resv. Bw   : 0           Unresv. Bw      : 5000000
TE1-> Resv. Bw   : 0           Unresv. Bw      : 5000000
TE2-> Resv. Bw   : 0           Unresv. Bw      : 4000000
TE3-> Resv. Bw   : 0           Unresv. Bw      : 4000000
TE4-> Resv. Bw   : 0           Unresv. Bw      : 0
TE5-> Resv. Bw   : 0           Unresv. Bw      : 0
TE6-> Resv. Bw   : 0           Unresv. Bw      : 0
TE7-> Resv. Bw   : 0           Unresv. Bw      : 0
---snip---
=====

```

Figure 18: MAM bandwidth allocation shows the bandwidth allocation for the CTs and TE classes.

Figure 18: MAM bandwidth allocation



25852

## Establishing LSPs

TE class 5 corresponds to CT2 and priority 2. No bandwidth can be reserved for an RSVP LSP with CT2, setup priority 2, and hold priority 2. This can be verified by configuring an empty path and an LSP, as follows:

```

# on PE-1:
configure {
  router "Base" {
    mpls {
      path "dyn" {
        admin-state enable
      }
      lsp "LSP-PE-1-PE-3-TE5" {
        admin-state enable
        type p2p-rsvp
        to 192.0.2.3
      }
    }
  }
}

```

```
class-type 2
path-computation-method local-cspf
primary "dyn" {
    bandwidth 1000
    priority {
        setup-priority 2
        hold-priority 2
    }
}
```

The path computation method must be local CSPF. The class type is by default CT0, but can be changed to CT2 by configuration. The class type can be configured in the **lsp** context, as shown here, or in the **primary path** context. The setup priority and hold priority are configured in the **primary path** context.

The LSP cannot be established, because no bandwidth is allocated to TE class 5, as follows:

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-3-TE5" path detail | match "Failure Code"
Failure Code      : noCspfRouteToDestination
```

No bandwidth is reserved on the RSVP interfaces, as follows:

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

=====
RSVP Interfaces
=====
Interface                Total    Active    Total BW  Resv BW  Adm Opr
                        Sessions Sessions (Mbps)   (Mbps)
-----
system                   -        -         -         -         Up  Up
int-PE-1-PE-2            0        0       10000     0         Up  Up
int-PE-1-PE-5            0        0       10000     0         Up  Up
-----
Interfaces : 3
=====
```

Bandwidth can be reserved for an LSP with CT1 and priorities 4, as for the following LSP:

```
# on PE-1:
configure {
    router "Base" {
        mpls {
            lsp "LSP-PE-1-PE-3-TE3" {
                admin-state enable
                type p2p-rsvp
                to 192.0.2.3
                path-computation-method local-cspf
                primary "dyn" {
                    bandwidth 2000
                    class-type 1
                    priority {
                        setup-priority 4
                        hold-priority 4
                    }
                }
            }
        }
    }
}
```

In the example, the CT is configured in the **primary path** context whereas the CT in the previous example was configured in the **lsp** context.

The path is set up via PE-2 to PE-3, as follows:

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-3-TE3" path detail

=====
MPLS LSP LSP-PE-1-PE-3-TE3 Path (Detail)
=====
Legend :
  @ - Detour Available          # - Detour In Use
  b - Bandwidth Protected      n - Node Protected
  s - Soft Preemption
  S - Strict                    L - Loose
  A - ABR                       + - Inherited
=====

-----
LSP LSP-PE-1-PE-3-TE3
Path dyn
-----
LSP Name      : LSP-PE-1-PE-3-TE3
From          : 192.0.2.1
To            : 192.0.2.3
Admin State   : Up                Oper State    : Up
Path Name     : dyn
Path LSP ID   : 55296              Path Type     : Primary
Path Admin    : Up                Path Oper     : Up
Out Interface : 1/1/c1/1           Out Label     : 524287
---snip---
Actual Hops   :
  192.168.12.1(192.0.2.1)         Record Label  : N/A
-> 192.168.12.2(192.0.2.2)         Record Label  : 524287
-> 192.168.23.2(192.0.2.3)        Record Label  : 524287
---snip---
=====
```

Bandwidth is reserved in TE class 3, as follows:

```
[/]
A:admin@PE-1# show router rsvp interface "int-PE-1-PE-2" detail

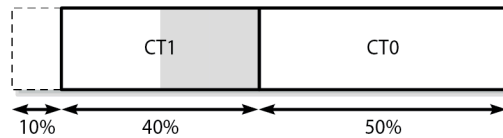
=====
RSVP Interface (Detailed) : int-PE-1-PE-2
=====

-----
Interface : int-PE-1-PE-2
-----
---snip---
Bandwidth for TE Class Types (Kbps)
TE0-> Resv. Bw : 0           Unresv. Bw   : 5000000
TE1-> Resv. Bw : 0           Unresv. Bw   : 5000000
TE2-> Resv. Bw : 0           Unresv. Bw   : 2000000
TE3-> Resv. Bw : 2000000     Unresv. Bw   : 2000000
TE4-> Resv. Bw : 0           Unresv. Bw   : 0
TE5-> Resv. Bw : 0           Unresv. Bw   : 0
TE6-> Resv. Bw : 0           Unresv. Bw   : 0
TE7-> Resv. Bw : 0           Unresv. Bw   : 0
---snip---
=====
```

Figure 19: Reserved and unreserved bandwidth shows the bandwidth reservation for CT1 on interface int-PE-1-PE-2 on PE-1 and on interface int-PE-2-PE-3 on PE-2.



Figure 19: Reserved and unreserved bandwidth



25853

An additional LSP is configured with CT1 and priority 4 and with CT0 as backup CT. The backup CT is not used when the amount of unreserved bandwidth for CT1 is insufficient, as in the following case where int-PE-1-PE-2 and int-PE-1-PE-5 have insufficient unreserved bandwidth for CT1:

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-3-TE3-backupTE1" {
        admin-state enable
        type p2p-rsvp
        to 192.0.2.3
        path-computation-method local-cspf
        primary "dyn" {
          bandwidth 5000
          class-type 1
          backup-class-type 0
          priority {
            setup-priority 4
            hold-priority 4
          }
        }
      }
    }
  }
}
```

The LSP does not come up, as follows:

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

=====
MPLS LSPs (Originating)
=====
LSP Name                               Tun   Fastfail  Adm  Opr
To                                       Id    Config
-----
LSP-PE-1-PE-3-TE5                       3     No        Up   Dwn
  192.0.2.3
LSP-PE-1-PE-3-TE3                       4     No        Up   Up
  192.0.2.3
LSP-PE-1-PE-3-TE3-backupTE1           5     No        Up   Dwn
  192.0.2.3
-----
LSPs : 3
=====
```

The bandwidth requirement is lowered, as follows:

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-3-TE3-backupTE1" {
        primary "dyn" {
```

```

        bandwidth 2500
    }
}

```

Interface int-PE-1-PE-2 does not have sufficient bandwidth for CT1, but the longer path via PE-5 and PE-4 has sufficient unreserved bandwidth for CT1. The LSP is operationally up, as follows:

```

[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-3-TE3-backupTE1" path detail
=====
MPLS LSP LSP-PE-1-PE-3-TE3-backupTE1 Path (Detail)
=====
Legend :
  @ - Detour Available          # - Detour In Use
  b - Bandwidth Protected       n - Node Protected
  s - Soft Preemption
  S - Strict                    L - Loose
  A - ABR                      + - Inherited
=====
LSP LSP-PE-1-PE-3-TE3-backupTE1
Path dyn
-----
LSP Name      : LSP-PE-1-PE-3-TE3-backupTE1
From          : 192.0.2.1
To            : 192.0.2.3
Admin State   : Up                Oper State    : Up
Path Name     : dyn
Path LSP ID   : 32256              Path Type     : Primary
Path Admin    : Up                Path Oper     : Up
Out Interface : 1/1/c1/2          Out Label     : 524287
---snip---
Setup Priority : 4                 Oper SetupPriority: 4
Hold Priority  : 4                 Oper HoldPriority : 4
Class Type     : 1                 Oper CT        : 1
Backup CT      : 0
---snip---
Actual Hops    :
  192.168.15.1(192.0.2.1)          Record Label  : N/A
  -> 192.168.15.2(192.0.2.5)       Record Label  : 524287
  -> 192.168.45.1(192.0.2.4)       Record Label  : 524287
  -> 192.168.34.1(192.0.2.3)       Record Label  : 524286
---snip---
=====

```

The bandwidth reservation on RSVP interface int-PE-1-PE-2 remains unchanged, because the bandwidth is reserved on int-PE-1-PE-5, as follows:

```

[/]
A:admin@PE-1# show router rsvp interface "int-PE-1-PE-5" detail
=====
RSVP Interface (Detailed) : int-PE-1-PE-5
=====
Interface : int-PE-1-PE-5
-----
Interface      : int-PE-1-PE-5
Port ID       : 1/1/c1/2
---snip---
Percent Link Bandwidth for Class Types*

```

```

Link Bw CT0      : 50          Link Bw CT4      : 0
Link Bw CT1      : 40          Link Bw CT5      : 0
Link Bw CT2      : 0           Link Bw CT6      : 0
Link Bw CT3      : 0           Link Bw CT7      : 10

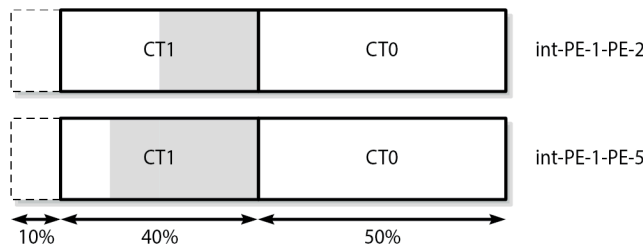
Bandwidth Constraints for Class Types (Kbps)
BC0              : 5000000     BC4              : 0
BC1              : 4000000     BC5              : 0
BC2              : 0           BC6              : 0
BC3              : 0           BC7              : 1000000

Bandwidth for TE Class Types (Kbps)
TE0-> Resv. Bw   : 0           Unresv. Bw      : 5000000
TE1-> Resv. Bw   : 0           Unresv. Bw      : 5000000
TE2-> Resv. Bw   : 0           Unresv. Bw      : 1500000
TE3-> Resv. Bw   : 2500000     Unresv. Bw      : 1500000
TE4-> Resv. Bw   : 0           Unresv. Bw      : 0
TE5-> Resv. Bw   : 0           Unresv. Bw      : 0
TE6-> Resv. Bw   : 0           Unresv. Bw      : 0
TE7-> Resv. Bw   : 0           Unresv. Bw      : 0
---snip---
=====

```

Figure 20: Reserved and unreserved bandwidth on PE-1 shows the reserved and unreserved bandwidth on the RSVP interfaces on PE-1.

Figure 20: Reserved and unreserved bandwidth on PE-1



25854

## Trigger backup class-type

This mechanism is described for MAM, but it is also supported in RDM.

On PE-1, port 1/1/c1/2 is shut down. The long path via PE-5 and PE-4 can no longer be used. However, the LSP has a backup CT (CT0), which is triggered by the port being down, as follows:

```

# on PE-1:
configure {
  port 1/1/c1/2 {
    admin-state disable
  }
}

```

```

[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-3-TE3-backupTE1" path detail

```

```

=====
MPLS LSP LSP-PE-1-PE-3-TE3-backupTE1 Path (Detail)
=====

```

```

Legend :
  @ - Detour Available      # - Detour In Use
  b - Bandwidth Protected  n - Node Protected
  s - Soft Preemption
  S - Strict                L - Loose
  A - ABR                  + - Inherited
=====
-----
LSP LSP-PE-1-PE-3-TE3-backupTE1
Path dyn
-----
LSP Name      : LSP-PE-1-PE-3-TE3-backupTE1
From          : 192.0.2.1
To           : 192.0.2.3
Admin State   : Up                      Oper State    : Up
Path Name     : dyn
Path LSP ID   : 32258                   Path Type     : Primary
Path Admin    : Up                      Path Oper     : Up
Out Interface : 1/1/c1/1                 Out Label    : 524286
---snip---
Setup Priority : 4                      Oper SetupPriority: 4
Hold Priority  : 4                      Oper HoldPriority : 4
Class Type    : 1                      Oper CT        : 0
Backup CT     : 0
---snip---
Actual Hops   :
  192.168.12.1(192.0.2.1)              Record Label  : N/A
-> 192.168.12.2(192.0.2.2)              Record Label  : 524286
-> 192.168.23.2(192.0.2.3)             Record Label  : 524286
---snip---
=====

```

The bandwidth for this LSP is reserved in TE class 1, as follows:

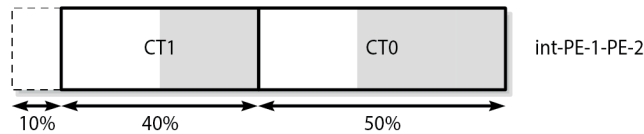
```

[/]
A:admin@PE-1# show router rsvp interface "int-PE-1-PE-2" detail
=====
RSVP Interface (Detailed) : int-PE-1-PE-2
=====
-----
Interface : int-PE-1-PE-2
-----
Interface      : int-PE-1-PE-2
Port ID       : 1/1/c1/1
Admin State    : Up                      Oper State    : Up
Active Sessions : 2                      Active Resvs  : 2
Total Sessions : 2
---snip---
Bandwidth for TE Class Types (Kbps)
TE0-> Resv. Bw : 0                      Unresv. Bw   : 2500000
TE1-> Resv. Bw : 2500000                 Unresv. Bw   : 2500000
TE2-> Resv. Bw : 0                      Unresv. Bw   : 2000000
TE3-> Resv. Bw : 2000000                 Unresv. Bw   : 2000000
TE4-> Resv. Bw : 0                      Unresv. Bw   : 0
TE5-> Resv. Bw : 0                      Unresv. Bw   : 0
TE6-> Resv. Bw : 0                      Unresv. Bw   : 0
TE7-> Resv. Bw : 0                      Unresv. Bw   : 0
---snip---
=====

```

**Figure 21: Bandwidth reservation** shows the bandwidth reservation on interface int-PE-1-PE-2. The bandwidth reservation on interface int-PE-2-PE-3 on PE-2 is identical.

Figure 21: Bandwidth reservation



25855

The preceding examples illustrate that bandwidth can be wasted in the MAM model. The bandwidth allocated to CT7 cannot be used because there is no TE class configured with CT7. The bandwidth cannot be shared between CTs. The next section describes how the same LSPs are used in the RDM model. They are established one-by-one and, therefore, they are shut down, as follows:

```
# on PE-1:
configure {
  port 1/1/c1/2 {
    admin-state enable
  }
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-3-TE5" {
        admin-state disable
      }
      lsp "LSP-PE-1-PE-3-TE3" {
        admin-state disable
      }
      lsp "LSP-PE-1-PE-3-TE3-backupTE1" {
        admin-state disable
      }
    }
  }
}
```

## Russian Doll Model

### Enable DiffServ RDM

The DiffServ TE configuration needs to be consistent on all the nodes in the network, as follows:

```
# on PE-1:
configure {
  router "Base" {
    rsvp {
      diffserv-te {
        admission-control-model rdm
        class-type-bw {
          ct0 50
          ct1 40
          ct7 10
        }
        te-class 0 {
          class-type 0
          priority 7
        }
        te-class 1 {
          class-type 0
          priority 4
        }
      }
    }
  }
}
```

```

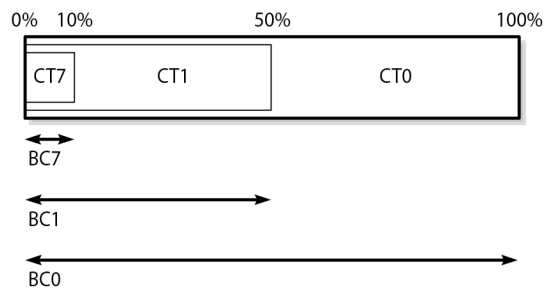
te-class 2 {
  class-type 1
  priority 7
}
te-class 3 {
  class-type 1
  priority 4
}
te-class 4 {
  class-type 2
  priority 7
}
te-class 5 {
  class-type 2
  priority 2
}
fc af {
  class-type 1
}
fc be {
  class-type 0
}
fc nc {
  class-type 2
}
}

```

In this example, three FCs are mapped to CTs. FC BE corresponds to CT0, FC AF to CT1, and FC NC to CT2. CT0 can be mapped to TE class 0 for priority 7, and to TE class 1 for priority 4. The mapping is similar for CT1 (with priorities 7 and 4), and CT2 (with priorities 7 or 2).

The RDM model allows the outer dolls (lower CT) to use the unused bandwidth allocated to the inner dolls (higher CT), as shown in [Figure 22: Russian Doll Model for three class types](#).

Figure 22: Russian Doll Model for three class types



25856

The calculation of the BCs takes into account the BCs of the inner dolls, as shown in the OSPF LSAs in the opaque database, as follows:

```

[/]
A:admin@PE-1# show router ospf opaque-database adv-router 192.0.2.1 detail

=====
Rtr Base OSPFv2 Instance 0 Opaque Link State Database (type: All) (detail)
=====
-----
Opaque LSA
-----

```

```

Area Id       : 0.0.0.0           Adv Router Id   : 192.0.2.1
Link State Id : 1.0.0.1           LSA Type        : Area Opaque
Sequence No   : 0x80000003       Checksum        : 0x9035
Age           : 23                Length          : 28
Options       : E
Advertisement : Traffic Engineering
              ROUTER-ID TLV (0001) Len 4 : 192.0.2.1
-----
Opaque LSA
-----
Area Id       : 0.0.0.0           Adv Router Id   : 192.0.2.1
Link State Id : 1.0.0.2           LSA Type        : Area Opaque
Sequence No   : 0x80000023       Checksum        : 0x302
Age           : 71                Length          : 164
Options       : E
Advertisement : Traffic Engineering
              LINK INFO TLV (0002) Len 140 :
                Sub-TLV: 1      Len: 1      LINK_TYPE   : 1
                Sub-TLV: 2      Len: 4      LINK_ID     : 192.0.2.2
                Sub-TLV: 3      Len: 4      LOC_IP_ADDR : 192.168.12.1
                Sub-TLV: 4      Len: 4      REM_IP_ADDR : 192.168.12.2
                Sub-TLV: 5      Len: 4      TE_METRIC   : 10
                Sub-TLV: 6      Len: 4      MAX_BDWTH  : 10000000 Kbps
                Sub-TLV: 7      Len: 4      RSRVBL_BDWTH : 10000000 Kbps
                Sub-TLV: 8      Len: 32     UNRSRVD_CLS0 :
                P0: 10000000 Kbps P1: 10000000 Kbps P2: 5000000 Kbps P3: 5000000 Kbps
                P4: 1000000 Kbps P5: 1000000 Kbps P6: 0 Kbps P7: 0 Kbps
                Sub-TLV: 9      Len: 4      ADMIN_GROUP : 0 None
                Sub-TLV: 17     Len: 36     TELK_BW_CONST:
                BW Model : RDM
                BC0: 10000000 Kbps BC1: 5000000 Kbps BC2: 1000000 Kbps BC3: 1000000 Kbps
                BC4: 1000000 Kbps BC5: 1000000 Kbps BC6: 1000000 Kbps BC7: 1000000 Kbps
---snip---
=====

```

Six TE classes are defined:

- TE0 and TE1 are defined for CT0. They can reserve all the available bandwidth, if it is not required by the other TE classes (100% = 50% for CT0 + 40% for CT1 + 10% for CT7)
- TE2 and TE3 are defined for CT1. They can reserve 50% of the bandwidth (50% = 40% for CT1 + 10% for CT7)
- TE4 and TE5 are defined for CT2. They can reserve 10% of the bandwidth, even though the configured bandwidth percentage for CT2 is 0. The 10% allocated to higher class CT7 can be used.

Bandwidth is more efficiently used in RDM than in MAM.

The BCs and bandwidth per TE class type show that bandwidth can be shared with the outer dolls, as follows:

```

[/]
A:admin@PE-1# show router rsvp interface "int-PE-1-PE-2" detail

=====
RSVP Interface (Detailed) : int-PE-1-PE-2
=====
-----
Interface : int-PE-1-PE-2
-----
Interface       : int-PE-1-PE-2
Port ID        : 1/1/c1/1
Admin State    : Up                Oper State     : Up
Active Sessions : 0                Active Resvs   : 0

```

```

Total Sessions      : 0
Subscription       : 100 %
Total BW           : 10000 Mbps
Port Speed         : 10000 Mbps
Aggregate          : Dsabl
---snip---
Percent Link Bandwidth for Class Types*
Link Bw CT0       : 50
Link Bw CT1       : 40
Link Bw CT2       : 0
Link Bw CT3       : 0
Link Bw CT4       : 0
Link Bw CT5       : 0
Link Bw CT6       : 0
Link Bw CT7       : 10

Bandwidth Constraints for Class Types (Kbps)
BC0               : 10000000
BC1               : 5000000
BC2               : 1000000
BC3               : 1000000
BC4               : 1000000
BC5               : 1000000
BC6               : 1000000
BC7               : 1000000

Bandwidth for TE Class Types (Kbps)
TE0-> Resv. Bw    : 0
TE1-> Resv. Bw    : 0
TE2-> Resv. Bw    : 0
TE3-> Resv. Bw    : 0
TE4-> Resv. Bw    : 0
TE5-> Resv. Bw    : 0
TE6-> Resv. Bw    : 0
TE7-> Resv. Bw    : 0
Unresv. Bw       : 10000000
Unresv. Bw       : 10000000
Unresv. Bw       : 5000000
Unresv. Bw       : 5000000
Unresv. Bw       : 1000000
Unresv. Bw       : 1000000
Unresv. Bw       : 0
Unresv. Bw       : 0
---snip---
=====

```

## Establishing LSPs

LSP-PE-1-PE-3-TE5 could not be established in the MAM model, because there was no bandwidth assigned to TE5 (CT2). However, in the RDM model, TE5 can use the bandwidth of the inner doll CT7 and the LSP is operationally up, as follows:

```

# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-3-TE5" {
        admin-state enable
      }
    }
  }
}

```

```

[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-3-TE5" path detail

```

```

=====
MPLS LSP LSP-PE-1-PE-3-TE5 Path (Detail)
=====

```

Legend :

```

@ - Detour Available          # - Detour In Use
b - Bandwidth Protected      n - Node Protected
s - Soft Preemption
S - Strict                   L - Loose
A - ABR                      + - Inherited

```

```

-----
LSP LSP-PE-1-PE-3-TE5
Path dyn
-----
LSP Name      : LSP-PE-1-PE-3-TE5

```



```

From       : 192.0.2.1
To         : 192.0.2.3
Admin State : Up                               Oper State      : Up
Path Name  : dyn
Path LSP ID : 6144                             Path Type       : Primary
Path Admin  : Up                               Path Oper       : Up
Out Interface : 1/1/c1/1                       Out Label      : 524287
---snip---
Setup Priority : 2                             Oper SetupPriority: 2
Hold Priority   : 2                             Oper HoldPriority : 2
Class Type     : 2                             Oper CT         : 2
Backup CT      : None
---snip---
Actual Hops    :
  192.168.12.1(192.0.2.1) Record Label    : N/A
  -> 192.168.12.2(192.0.2.2) Record Label    : 524287
  -> 192.168.23.2(192.0.2.3) Record Label    : 524287
---snip---
=====

```

The bandwidth reservation on interface int-PE-1-PE-2 is as follows:

```

[/]
A:admin@PE-1# show router rsvp interface "int-PE-1-PE-2" detail

=====
RSVP Interface (Detailed) : int-PE-1-PE-2
=====
-----
Interface : int-PE-1-PE-2
-----
---snip---
Percent Link Bandwidth for Class Types*
Link Bw CT0      : 50                               Link Bw CT4      : 0
Link Bw CT1      : 40                               Link Bw CT5      : 0
Link Bw CT2      : 0                                Link Bw CT6      : 0
Link Bw CT3      : 0                                Link Bw CT7      : 10

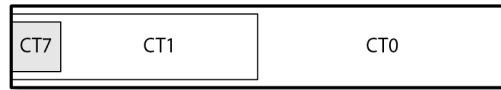
Bandwidth Constraints for Class Types (Kbps)
BC0      : 10000000                                BC4      : 1000000
BC1      : 5000000                                BC5      : 1000000
BC2      : 1000000                                BC6      : 1000000
BC3      : 1000000                                BC7      : 1000000

Bandwidth for TE Class Types (Kbps)
TE0-> Resv. Bw : 0                               Unresv. Bw      : 9000000
TE1-> Resv. Bw : 0                               Unresv. Bw      : 9000000
TE2-> Resv. Bw : 0                               Unresv. Bw      : 4000000
TE3-> Resv. Bw : 0                               Unresv. Bw      : 4000000
TE4-> Resv. Bw : 0                               Unresv. Bw      : 0
TE5-> Resv. Bw : 1000000                       Unresv. Bw      : 0
TE6-> Resv. Bw : 0                               Unresv. Bw      : 0
TE7-> Resv. Bw : 0                               Unresv. Bw      : 0
---snip---
=====

```

This LSP uses all the available bandwidth for CT7. Because TE5 is defined with the best priority (2) of all TE classes, this LSP is not preempted when a new LSP is enabled. Therefore, this bandwidth is subtracted from the amount of unreserved bandwidth. The remaining unreserved bandwidth is for CT0 and CT1 only. LSPs with other CTs cannot be established on this interface. [Figure 23: Reserved bandwidth for LSP with CT2 \(one session\)](#) shows the reserved bandwidth on interface int-PE-1-PE-2 for this LSP.

Figure 23: Reserved bandwidth for LSP with CT2 (one session)



25857

Another LSP is established: LSP-PE-1-PE-3-TE3, with CT1 and priority 4, as follows:

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-3-TE3" {
        admin-state enable
      }
    }
  }
}
```

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-3-TE3" path detail

=====
MPLS LSP LSP-PE-1-PE-3-TE3 Path (Detail)
=====
Legend :
@ - Detour Available          # - Detour In Use
b - Bandwidth Protected      n - Node Protected
s - Soft Preemption
S - Strict                    L - Loose
A - ABR                       + - Inherited

-----
LSP LSP-PE-1-PE-3-TE3
Path dyn
-----
LSP Name      : LSP-PE-1-PE-3-TE3
From          : 192.0.2.1
To           : 192.0.2.3
Admin State   : Up
Path Name    : dyn
Path LSP ID  : 55298
Path Admin   : Up
Out Interface : 1/1/c1/1
---snip---
Setup Priority : 4
Hold Priority  : 4
Class Type    : 1
Backup CT     : None
---snip---
Actual Hops   :
  192.168.12.1(192.0.2.1)
-> 192.168.12.2(192.0.2.2)
-> 192.168.23.2(192.0.2.3)
---snip---
Oper State    : Up
Path Type    : Primary
Path Oper    : Up
Out Label    : 524286
Oper SetupPriority: 4
Oper HoldPriority: 4
Oper CT      : 1
Record Label : N/A
Record Label : 524286
Record Label : 524286
=====
```

The bandwidth reservation on RSVP interface int-PE-1-PE-2 is as follows:

```
[/]
A:admin@PE-1# show router rsvp interface "int-PE-1-PE-2" detail
```

```

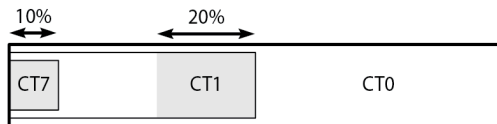
=====
RSVP Interface (Detailed) : int-PE-1-PE-2
=====
-----
Interface : int-PE-1-PE-2
-----
Interface      : int-PE-1-PE-2
Port ID       : 1/1/cl/1
Admin State   : Up                Oper State    : Up
Active Sessions : 2                Active Resvs  : 2
Total Sessions : 2
---snip---
Bandwidth Constraints for Class Types (Kbps)
BC0           : 10000000          BC4           : 1000000
BC1           : 5000000          BC5           : 1000000
BC2           : 1000000          BC6           : 1000000
BC3           : 1000000          BC7           : 1000000

Bandwidth for TE Class Types (Kbps)
TE0-> Resv. Bw : 0                Unresv. Bw   : 7000000
TE1-> Resv. Bw : 0                Unresv. Bw   : 7000000
TE2-> Resv. Bw : 0                Unresv. Bw   : 2000000
TE3-> Resv. Bw : 2000000          Unresv. Bw   : 2000000
TE4-> Resv. Bw : 0                Unresv. Bw   : 0
TE5-> Resv. Bw : 1000000          Unresv. Bw   : 0
TE6-> Resv. Bw : 0                Unresv. Bw   : 0
TE7-> Resv. Bw : 0                Unresv. Bw   : 0
---snip---
=====

```

Figure 24: Bandwidth reservation for LSP with CT2 and LSP with CT1 (two sessions) shows the bandwidth reservation for the two active sessions.

Figure 24: Bandwidth reservation for LSP with CT2 and LSP with CT1 (two sessions)



25858

Another LSP is established for CT1, requesting more bandwidth than the short path via PE-2 has available. Therefore, the longer path via PE-5 and PE-4 is set up for this LSP, as follows:

```

# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-3-TE3-backupTE1" {
        admin-state enable
      }
    }
  }
}

[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-3-TE3-backupTE1" path detail

=====
MPLS LSP LSP-PE-1-PE-3-TE3-backupTE1 Path (Detail)
=====
Legend :

```

```

@ - Detour Available          # - Detour In Use
b - Bandwidth Protected      n - Node Protected
s - Soft Preemption
S - Strict                   L - Loose
A - ABR                      + - Inherited
=====
-----
LSP LSP-PE-1-PE-3-TE3-backupTE1
Path dyn
-----
LSP Name      : LSP-PE-1-PE-3-TE3-backupTE1
From          : 192.0.2.1
To            : 192.0.2.3
Admin State   : Up                      Oper State    : Up
Path Name     : dyn
Path LSP ID   : 32260                    Path Type     : Primary
Path Admin    : Up                      Path Oper     : Up
Out Interface : 1/1/c1/2                  Out Label    : 524287
---snip---
Setup Priority : 4                      Oper SetupPriority: 4
Hold Priority  : 4                      Oper HoldPriority : 4
Class Type    : 1                      Oper CT       : 1
Backup CT     : 0
---snip---
Actual Hops   :
  192.168.15.1(192.0.2.1)              Record Label  : N/A
  -> 192.168.15.2(192.0.2.5)            Record Label  : 524287
  -> 192.168.45.1(192.0.2.4)           Record Label  : 524287
  -> 192.168.34.1(192.0.2.3)           Record Label  : 524285
---snip---
=====

```

The bandwidth for this LSP is reserved on interface int-PE-1-PE-5, because the amount of unreserved bandwidth for TE3 is insufficient and inner dolls cannot use bandwidth assigned to outer dolls. Inner dolls are of higher priority than outer dolls, as follows:

```

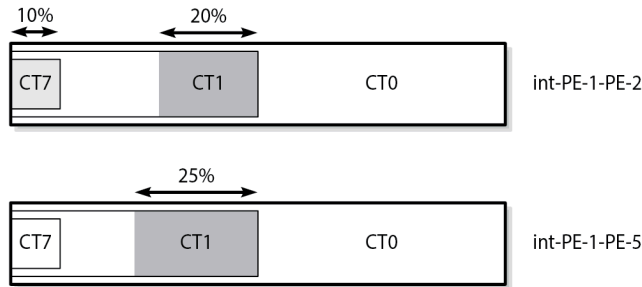
[/]
A:admin@PE-1# show router rsvp interface "int-PE-1-PE-5" detail
=====
RSVP Interface (Detailed) : int-PE-1-PE-5
=====
-----
Interface : int-PE-1-PE-5
-----
---snip---
Bandwidth Constraints for Class Types (Kbps)
BC0          : 10000000                BC4          : 10000000
BC1          : 5000000                 BC5          : 10000000
BC2          : 1000000                 BC6          : 10000000
BC3          : 1000000                 BC7          : 10000000

Bandwidth for TE Class Types (Kbps)
TE0-> Resv. Bw : 0                    Unresv. Bw  : 7500000
TE1-> Resv. Bw : 0                    Unresv. Bw  : 7500000
TE2-> Resv. Bw : 0                    Unresv. Bw  : 2500000
TE3-> Resv. Bw : 2500000              Unresv. Bw  : 2500000
TE4-> Resv. Bw : 0                    Unresv. Bw  : 1000000
TE5-> Resv. Bw : 0                    Unresv. Bw  : 1000000
TE6-> Resv. Bw : 0                    Unresv. Bw  : 0
TE7-> Resv. Bw : 0                    Unresv. Bw  : 0
---snip---

```

Figure 25: Reserved bandwidth on both interfaces of PE-1 (three sessions) shows the reserved bandwidth on both interfaces of PE-1.

Figure 25: Reserved bandwidth on both interfaces of PE-1 (three sessions)



25859

The following LSP is configured on PE-1:

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-3-TE1" {
        admin-state enable
        type p2p-rsvp
        to 192.0.2.3
        path-computation-method local-cspf
        primary "dyn" {
          bandwidth 3000
          priority {
            setup-priority 4
            hold-priority 4
          }
        }
      }
    }
  }
}
```

The class type is by default 0. CT0 and priority 4 corresponds to TE1. There is sufficient bandwidth available on the short path via PE-2. The bandwidth reservation on RSVP interface int-PE-1-PE-2 is as follows:

```
[/]
A:admin@PE-1# show router rsvp interface "int-PE-1-PE-2" detail

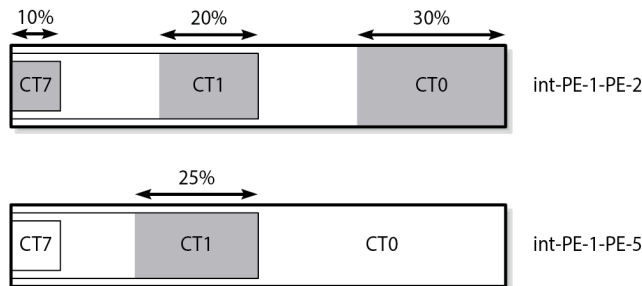
=====
RSVP Interface (Detailed) : int-PE-1-PE-2
=====
-----
Interface : int-PE-1-PE-2
-----
---snip---
Bandwidth for TE Class Types (Kbps)
TE0-> Resv. Bw   : 0                Unresv. Bw   : 4000000
TE1-> Resv. Bw   : 3000000          Unresv. Bw   : 4000000
TE2-> Resv. Bw   : 0                Unresv. Bw   : 2000000
TE3-> Resv. Bw   : 2000000          Unresv. Bw   : 2000000
TE4-> Resv. Bw   : 0                Unresv. Bw   : 0
TE5-> Resv. Bw   : 1000000          Unresv. Bw   : 0
TE6-> Resv. Bw   : 0                Unresv. Bw   : 0
```

```
TE7-> Resv. Bw : 0 Unresv. Bw : 0
---snip---
```

None of the established LSPs can be preempted. Therefore, the sum of the reserved and unreserved bandwidth does not exceed the total bandwidth.

Figure 26: Reserved bandwidth on both interfaces on PE-1 (four sessions) shows the bandwidth reservation on both interfaces.

Figure 26: Reserved bandwidth on both interfaces on PE-1 (four sessions)



25860

The following LSP with CT0 and priority 7 is configured on PE-1:

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-3-TE0" {
        admin-state enable
        type p2p-rsvp
        to 192.0.2.3
        path-computation-method local-cspf
        primary "dyn" {
          bandwidth 4000
          priority {
            setup-priority 7
            hold-priority 7
          }
        }
      }
    }
  }
}
```

The bandwidth is reserved in TE class 0. There is sufficient bandwidth on the short path to PE-3. The bandwidth is now reserved for 100%, as follows:

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

=====
RSVP Interfaces
=====
```

Interface	Total Sessions	Active Sessions	Total BW (Mbps)	Resv BW (Mbps)	Adm	Opr
system	-	-	-	-	Up	Up
<b>int-PE-1-PE-2</b>	<b>4</b>	<b>4</b>	<b>10000</b>	<b>10000</b>	<b>Up</b>	<b>Up</b>
int-PE-1-PE-5	1	1	10000	2500	Up	Up

```
-----
Interfaces : 3
```

The bandwidth reservation on int-PE-1-PE-2 is as follows:

```
[/]
A:admin@PE-1# show router rsvp interface "int-PE-1-PE-2" detail

=====
RSVP Interface (Detailed) : int-PE-1-PE-2
=====
-----
Interface : int-PE-1-PE-2
-----
---snip---
Percent Link Bandwidth for Class Types*
Link Bw CT0      : 50                Link Bw CT4      : 0
Link Bw CT1      : 40                Link Bw CT5      : 0
Link Bw CT2      : 0                 Link Bw CT6      : 0
Link Bw CT3      : 0                 Link Bw CT7      : 10

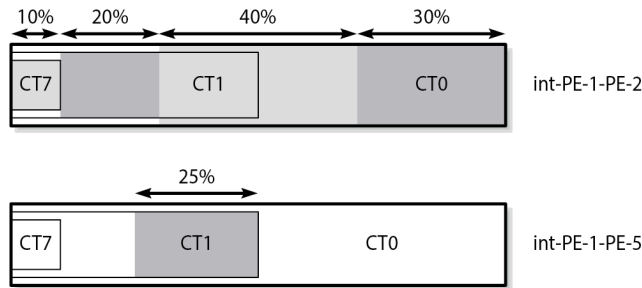
Bandwidth Constraints for Class Types (Kbps)
BC0      : 10000000                BC4      : 1000000
BC1      : 5000000                 BC5      : 1000000
BC2      : 1000000                 BC6      : 1000000
BC3      : 1000000                 BC7      : 1000000

Bandwidth for TE Class Types (Kbps)
TE0-> Resv. Bw : 4000000           Unresv. Bw      : 0
TE1-> Resv. Bw : 3000000           Unresv. Bw      : 4000000
TE2-> Resv. Bw : 0                 Unresv. Bw      : 0
TE3-> Resv. Bw : 2000000           Unresv. Bw      : 2000000
TE4-> Resv. Bw : 0                 Unresv. Bw      : 0
TE5-> Resv. Bw : 1000000           Unresv. Bw      : 0
TE6-> Resv. Bw : 0                 Unresv. Bw      : 0
TE7-> Resv. Bw : 0                 Unresv. Bw      : 0
---snip---
```

Even though the sum of the reserved bandwidth equals the maximum reservable bandwidth on the link, there is still unreserved bandwidth for specific TE classes. When an additional LSP is established requiring bandwidth in TE3 or TE1 (which have setup priority 4), it can preempt another LSP with a lower hold priority. LSPs requiring bandwidth in TE class TE2 have a setup priority 7 and cannot preempt any other LSP. The setup priority in TE1 and TE3 is 4, which is higher than the hold priority in TE2 and TE0 (7 is the lowest priority). There are no LSPs in TE2, so the only LSPs to preempt have bandwidth reserved in TE0.

[Figure 27: Reserved bandwidth on both interfaces of PE-1 \(five sessions\)](#) shows the bandwidth reservation on the interfaces of PE-1.

Figure 27: Reserved bandwidth on both interfaces of PE-1 (five sessions)



25861

## Preemption

The following LSP is configured with CT1, setup priority 4, and hold priority 4, which corresponds to TE class 3:

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-3-TE3-2nd" {
        admin-state enable
        type p2p-rsvp
        to 192.0.2.3
        class-type 1
        path-computation-method local-cspf
        primary "dyn" {
          bandwidth 750
          priority {
            setup-priority 4
            hold-priority 4
          }
        }
      }
    }
  }
}
```

Because the setup priority 4 exceeds the hold priority 7 of LSP-PE-1-PE-3-TE0, this LSP preempts the existing one. The following output shows that the next hop for LSP-PE-1-PE-3-TE0 is 192.168.15.2 (PE-5), while the next hop for LSP-PE-1-PE-3-TE3-2nd is 192.168.12.2 (PE-2):

```
[/]
A:admin@PE-1# show router mpls lsp path

=====
MPLS LSP Path
=====
-----
LSP Name   : LSP-PE-1-PE-3-TE5
From       : 192.0.2.1
To         : 192.0.2.3
Adm State  : Up
Oper State : Up
-----
Path Name  Next Hop      Type      Out I/F  Adm  Opr
-----
dyn        192.168.12.2 Primary    1/1/c1/1 Up    Up
```



```

-----
LSP Name   : LSP-PE-1-PE-3-TE3
From       : 192.0.2.1
To         : 192.0.2.3
Adm State  : Up                               Oper State   : Up
-----
Path Name           Next Hop           Type           Out I/F      Adm  Opr
-----
dyn                 192.168.12.2   Primary        1/1/c1/1     Up   Up
-----

LSP Name   : LSP-PE-1-PE-3-TE3-backupTE1
From       : 192.0.2.1
To         : 192.0.2.3
Adm State  : Up                               Oper State   : Up
-----
Path Name           Next Hop           Type           Out I/F      Adm  Opr
-----
dyn                 192.168.15.2   Primary        1/1/c1/2     Up   Up
-----

LSP Name   : LSP-PE-1-PE-3-TE1
From       : 192.0.2.1
To         : 192.0.2.3
Adm State  : Up                               Oper State   : Up
-----
Path Name           Next Hop           Type           Out I/F      Adm  Opr
-----
dyn                 192.168.12.2   Primary        1/1/c1/1     Up   Up
-----

LSP Name   : LSP-PE-1-PE-3-TE0
From       : 192.0.2.1
To         : 192.0.2.3
Adm State  : Up                               Oper State   : Up
-----
Path Name           Next Hop           Type           Out I/F      Adm  Opr
-----
dyn                 192.168.15.2   Primary        1/1/c1/2     Up   Up
-----

LSP Name   : LSP-PE-1-PE-3-TE3-2nd
From       : 192.0.2.1
To         : 192.0.2.3
Adm State  : Up                               Oper State   : Up
-----
Path Name           Next Hop           Type           Out I/F      Adm  Opr
-----
dyn                 192.168.12.2   Primary        1/1/c1/1     Up   Up
=====

```

The bandwidth reservation on RSVP interface int-PE-1-PE-2 is as follows:

```
[/]
A:admin@PE-1# show router rsvp interface "int-PE-1-PE-2" detail

=====
RSVP Interface (Detailed) : int-PE-1-PE-2
=====
-----
Interface : int-PE-1-PE-2
-----
---snip---
Bandwidth for TE Class Types (Kbps)
TE0-> Resv. Bw   : 0                Unresv. Bw   : 3250000
TE1-> Resv. Bw   : 3000000          Unresv. Bw   : 3250000
TE2-> Resv. Bw   : 0                Unresv. Bw   : 1250000
TE3-> Resv. Bw   : 2750000          Unresv. Bw   : 1250000
TE4-> Resv. Bw   : 0                Unresv. Bw   : 0
TE5-> Resv. Bw   : 1000000          Unresv. Bw   : 0
TE6-> Resv. Bw   : 0                Unresv. Bw   : 0
TE7-> Resv. Bw   : 0                Unresv. Bw   : 0
---snip---
```

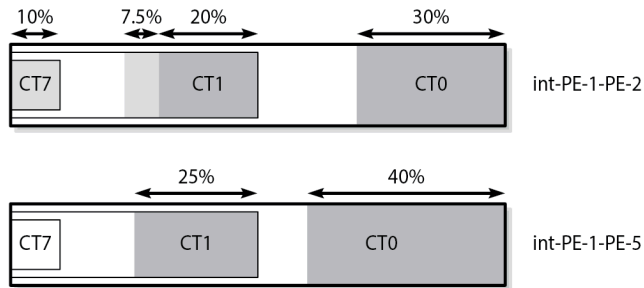
The bandwidth reservation on RSVP interface int-PE-1-PE-5 is as follows

```
[/]
A:admin@PE-1# show router rsvp interface "int-PE-1-PE-5" detail

=====
RSVP Interface (Detailed) : int-PE-1-PE-5
=====
-----
Interface : int-PE-1-PE-5
-----
---snip---
Bandwidth for TE Class Types (Kbps)
TE0-> Resv. Bw   : 4000000          Unresv. Bw   : 3500000
TE1-> Resv. Bw   : 0                Unresv. Bw   : 7500000
TE2-> Resv. Bw   : 0                Unresv. Bw   : 2500000
TE3-> Resv. Bw   : 2500000          Unresv. Bw   : 2500000
TE4-> Resv. Bw   : 0                Unresv. Bw   : 1000000
TE5-> Resv. Bw   : 0                Unresv. Bw   : 1000000
TE6-> Resv. Bw   : 0                Unresv. Bw   : 0
TE7-> Resv. Bw   : 0                Unresv. Bw   : 0
---snip---
```

Figure 28: Reserved bandwidth on both interfaces on PE-1 (six sessions) shows the bandwidth reservation on both interfaces on PE-1 for the six sessions.

Figure 28: Reserved bandwidth on both interfaces on PE-1 (six sessions)



25862

Preemption can also be within the same CT. An LSP with CT0 and priority 4 (TE1) could have preempted the LSP with CT0 and priority 7 (TE0) equally well.

## Bandwidth availability check

A **tools** command can be launched to verify the available bandwidth toward a node for a specific class type (by default CT0) and priority (by default setup priority 7 and hold priority 0). The options for this command are as follows:

```
[/]
A:admin@PE-1# tools perform router mpls cspf to 192.0.2.3 ?

cspf to <IP address> [from <IP address>] [bandwidth <number>] [include-bitmap <string>]
[exclude-bitmap <string>] [hop-limit <number>] [exclude-address <ipv4 address>] [metric-type-te]
[strict-srlg] [srlg-group <number>] [exclude-node <ipv4 address>] [skip-interface <string>] [ds-
class-type <number>] [cspf-reqtype <keyword>] [least-fill-min-thd <number>] [setup-priority
<number>]
[hold-priority <number>]

bandwidth          - rate-in-mbps
cspf-reqtype       - all|random|least-fill : keywords
ds-class-type      - <number> - <0..7>
exclude-address    - a.b.c.d (outbound interface)
exclude-bitmap     - string '<1..40 characters>'
exclude-node       - [a.b.c.d] (outbound interface)
from               - ipv4 address '<d.d.d.d>' or ipv6 address
                   '(<x:x:x:x:x:x:x:x>|<x:x:x:x:x:x:d.d.d.d>)'
hold-priority      - <number> - <0..7>
hop-limit          - <number> - <2..255>
include-bitmap     - [0..4294967295] - accepted in decimal, hex(0x) or binary(0b)
least-fill-min-thd - <number> - <1..100>
metric-type-te     - This element has no values
setup-priority     - <number> - <0..7>
skip-interface     - [max 32 chars]
srlg-group         - <number> - <0..4294967295>
strict-srlg       - This element has no values
```

The following verifies whether an LSP can be set up from PE-1 to PE-3 requesting 100 Mb/s with CT0 (default) and both priorities equal to 4:

```
[/]
A:admin@PE-1# tools perform router mpls cspf to 192.0.2.3 bandwidth 100 setup-priority 4 hold-
priority 4
Req CSPF for all ECMP paths
  from: this node to: 192.0.2.3 w/(DiffServ = RDM) class: 0 , setup Priority 4, Hold Priority
  4 TE Class: 1

CSPF Path
To      : 192.0.2.3
Path 1  : (cost 20)
  Src:   192.0.2.1 (= Rtr)
  Egr:   192.168.12.1      -> Ingr:   192.168.12.2      Rtr:
192.0.2.2 (met 10)
  Egr:   192.168.23.1     -> Ingr:   192.168.23.2      Rtr:
192.0.2.3 (met 10)
  Dst:   192.0.2.3 (= Rtr)
```

The short path via PE-2 has sufficient bandwidth for an LSP with these TE requirements (TE class 1 with CT0 and both priorities 4). This is different for TE class 5 (CT2 and priorities 2), where the bandwidth is completely reserved. The following shows that the longer path via PE-5 and PE-4 must be taken:

```
[/]
A:admin@PE-1# tools perform router mpls cspf to 192.0.2.3 bandwidth 100 ds-class-type 2 setup-
priority 2 hold-priority 2
Req CSPF for all ECMP paths
  from: this node to: 192.0.2.3 w/(DiffServ = RDM) class: 2 , setup Priority 2, Hold Priority
  2 TE Class: 5

CSPF Path
To      : 192.0.2.3
Path 1  : (cost 30)
  Src:   192.0.2.1 (= Rtr)
  Egr:   192.168.15.1     -> Ingr:   192.168.15.2      Rtr:
192.0.2.5 (met 10)
  Egr:   192.168.45.2    -> Ingr:   192.168.45.1      Rtr:
192.0.2.4 (met 10)
  Egr:   192.168.34.2    -> Ingr:   192.168.34.1      Rtr:
192.0.2.3 (met 10)
  Dst:   192.0.2.3 (= Rtr)
```

This **tools** command can only be launched when a TE class is defined with the requested CT and priority. An error is raised when the request cannot be fulfilled, as follows:

```
[/]
A:admin@PE-1# tools perform router mpls cspf to 192.0.2.3 setup-priority 5
MINOR: CLI #2005: Error while processing command - No Te class mapped to Class Type 0 , Setup
Priority 5.

[/]
A:admin@PE-1# tools perform router mpls cspf to 192.0.2.3 setup-priority 4
MINOR: CLI #2005: Error while processing command - No Te class mapped to Class Type 0 , Hold
Priority 0.
```

## Conclusion

DiffServ TE enforces different BCs for different classes of traffic. DiffServ TE controls overbooking and supports preemption. Two BC models are described in this chapter: MAM and RDM.

# Entropy Label

This chapter provides information about the Entropy Label (EL).

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

## Applicability

This chapter was initially written for SR OS Release 14.0.R4. The CLI in the current edition corresponds to SR OS Release 24.3.R1.

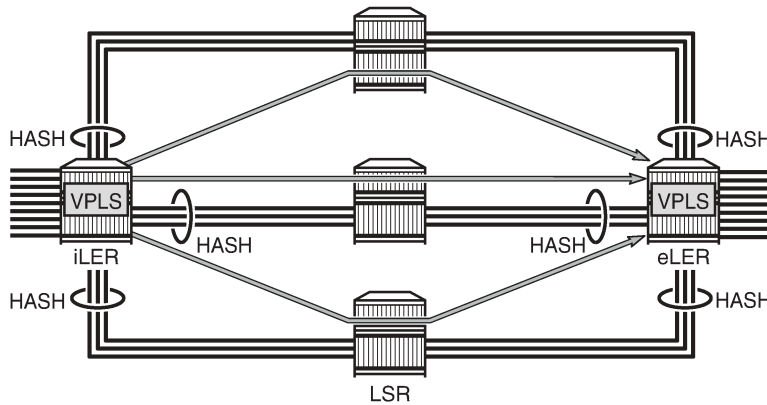
RFC 6391 hash label or flow-aware transport label is supported in SR OS Release 8.0.R1, and later. RFC 6790 Entropy Labels (ELs) are supported on RSVP and BGP tunnels in Release 14.0.R1, and later, and supported on LDP tunnels in Release 14.0.R4, and later.

## Overview

Entropy is the degree of disorder or uncertainty in a system. SR OS supports both the MPLS EL and the hash label, but they are mutually exclusive. These labels allow Label Switched Routers (LSRs) to load-balance labeled packets in a granular way without the need to inspect the IPv4 or IPv6 header. The hash label is applicable to services such as Epipe VLL, VPLS, IES (spoke-SDP), and VPRN services. The main advantage of the EL compared to the hash label is that the EL can be applied to a wider range of services, such as EVPN, BGP VPWS, Fpipe VLL, lpipe VLL, H-VPLS, and BGP RFC 3107 tunnels.

[Figure 29: Load-balancing of flows based on hash label or entropy label](#) shows that different flows from an ingress Label Edge Router (iLER) are load-balanced across different paths in the MPLS network toward the egress Label Edge Router (eLER).

Figure 29: Load-balancing of flows based on hash label or entropy label



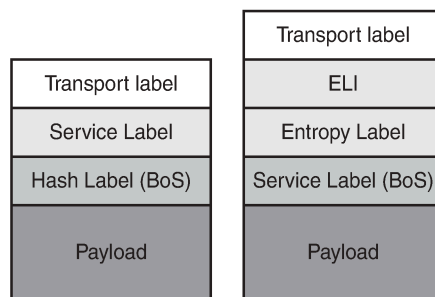
26076

The EL is inserted below the innermost Tunnel Label (TL) in the label stack, along with the Entropy Label Indicator (ELI), which is a special purpose MPLS label with a value of 7 to indicate that the next label in the stack is an EL. As with the hash label, the value of the EL is calculated based on a hash of the packet payload header (IP and Layer 4). The EL is inserted as deep as possible in the stack to ensure preservation for as far as possible through the network. When the EL and ELI are present, load-balancing on the transit LSRs can be configured to only take into account the EL label for Link Aggregation Group (LAG) and Equal Cost Multi-Path (ECMP). Load-balancing on the LSR can be configured to take into account the IP header also, but that is not required when the EL and ELI are present. The eLER removes the EL and ELI before forwarding the packet to its final destination.

The EL requires two additional labels in the label stack, which may result in an unsupported label stack depth in an intermediate (possibly third-party) LSR. SR OS allows control of the EL and ELI insertion and accounting of extra labels in the tunnel table. The supported label stack depth is 12 in SR OS Release 14.0, including transport, service, hash, and OAM labels.

Figure 30: Label stack with hash label versus label stack with EL and ELI shows a comparison between a label stack with a hash label and a label stack with an EL and ELI.

Figure 30: Label stack with hash label versus label stack with EL and ELI



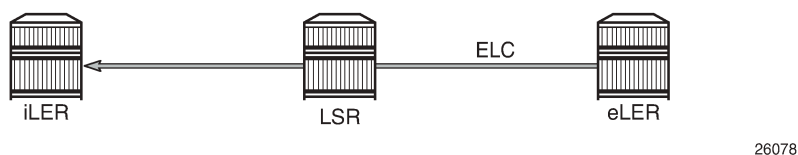
26077

Hash labels or ELs are inserted by the iLER, based on the incoming packet header. The iLER uses a single label to represent one or more flows, based on the hash of the incoming packet header IP and Layer 4. The hash label is a single label that is inserted at the bottom of the stack, below the service label. The EL is inserted together with the ELI below the innermost tunnel label, but above the service label. The entropy and hash label value is outside of the range for MPLS label values, because the most significant bit equals 1 for entropy and hash labels and 0 for MPLS labels.

The EL is used as part of the LSR hashing algorithm for spraying packets over multi-port LAGs, ECMP, or BGP tunnels with multiple downstream interfaces. The packet ordering is preserved because one label is used per conversation. Hashing based on a label stack containing an EL per service will become more granular than when based on a standard label stack alone.

**Figure 31: Downstream LERs signal EL capability to iLER** shows how the eLER signals its EL capability to the iLER.

*Figure 31: Downstream LERs signal EL capability to iLER*



The Entropy Label Capability (ELC) is signaled by the eLER and indicates the ability to receive and process the ELs. This can be advertised for LDP and RSVP. However, ELC signaling is not supported for BGP tunnels, because no agreed standard exists in the IETF. RFC 6790 introduced the ELC BGP attribute that can be signaled by the eLER to indicate that it supports EL. According to RFC 6790, LSRs incapable of processing ELs must remove the ELC BGP attribute, but this requirement could not be guaranteed; therefore, the ELC BGP attribute has been deprecated in RFC 7447.

As a workaround, at the iLER, an override of the ELC for a BGP tunnel can be manually configured. After this ELC has been overridden, the BGP sender assumes that the receiver is capable of receiving and processing the ELs, regardless of the signaled ELC. The iLER inserts an EL on a tunnel for which the ELC is confirmed by the downstream peer or when the ELC is overridden by configuration.

ELC signaling can be enabled for LDP on the LERs, as follows:

```
# on PE-1, PE-2:
configure {
  router "Base" {
    ldp {
      entropy-label-capability true
    }
  }
}
```

ELC signaling for RSVP can be enabled as follows:

```
# on PE-1, PE-2:
configure {
  router "Base" {
    rsvp {
      entropy-label-capability true
    }
  }
}
```

As previously described, the lack of an IETF standard for BGP tunnels means that ELC is ignored by the receiving LER, and no EL is inserted. Therefore, an override is required that assumes that the far-end LER has ELC, and allows insertion of the EL. The override is enabled as follows:

```
# on PE-1, PE-2:
```



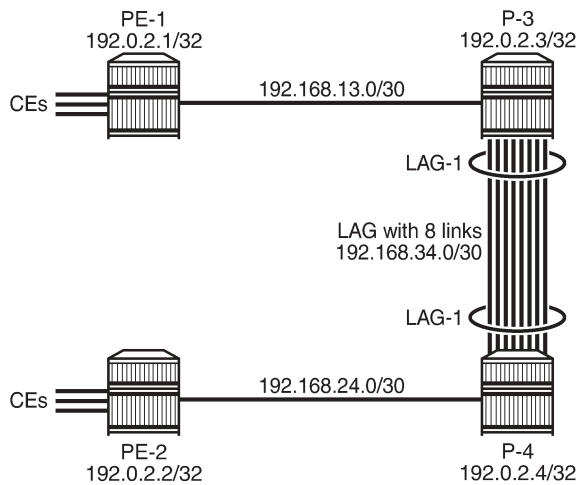
```
configure {  
  router "Base" {  
    bgp {  
      override-tunnel-elc true  
    }  
  }  
}
```

When the ELC is overridden for BGP, the iLER assumes that the receiver can handle the EL.

## Configuration

In this section, an EVPN-VPLS is configured on PE-1 and PE-2 to illustrate LAG hashing based on EL. [Figure 32: Example topology](#) shows the topology used for this example. Load-balancing of the traffic is performed in the P-routers that are connected by a LAG with eight links.

Figure 32: Example topology



26079

The initial configuration of the PE/P-routers includes the following:

- Cards, MDAs, ports
- LAG with eight network links between P-3 and P-4
- Router interfaces
- IGP (IS-IS or OSPF)
- LDP enabled on all interfaces
- MPLS enabled on all interfaces. RSVP enabled.
- iBGP configured with P-3 as route reflector (RR)

For the configuration of EVPN-MPLS, the BGP configuration needs to include the address family EVPN, as follows. See chapter *EVPN for MPLS Tunnels* for more information.

```
# on PE-1, PE-2:  
configure {  
  router "Base" {  
    autonomous-system 64500  
    bgp {  
      address-family evpn {  
        ...  
      }  
    }  
  }  
}
```

```

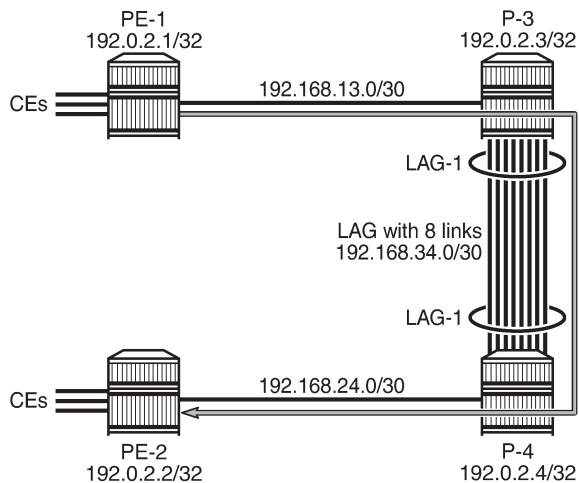
ebgp-default-reject-policy {
  import false
  export false
}
rapid-update {
  evpn true
}
group "iBGP" {
  peer-as 64500
  family {
    evpn true
  }
}
neighbor "192.0.2.3" {
  group "iBGP"
}
}

```

## EVPN service using RSVP tunnel with EL

Figure 33: RSVP LSP "LSP-PE-1-PE-2" from PE-1 to PE-2 via P-3 and P-4 shows LSP "LSP-PE-1-PE-2" from PE-1 to PE-2 via core routers P-3 and P-4.

Figure 33: RSVP LSP "LSP-PE-1-PE-2" from PE-1 to PE-2 via P-3 and P-4



26080

The LSP "LSP-PE-1-PE-2" is configured on PE-1, as follows:

```

# on PE-1:
configure {
  router "Base" {
    mpls {
      path "path-PE-1-PE-2" {
        admin-state enable
        hop 10 {
          ip-address 192.168.13.2
          type strict
        }
        hop 20 {
          ip-address 192.168.34.2
        }
      }
    }
  }
}

```

```

        type strict
    }
    hop 30 {
        ip-address 192.168.24.1
        type strict
    }
}
lsp "LSP-PE-1-PE-2" {
    admin-state enable
    type p2p-rsvp
    to 192.0.2.2
    entropy-label true
    primary "path-PE-1-PE-2" { }
}
}

```

The configuration for LSP "LSP-PE-2-PE-1" on PE-2 is similar.

ELC is disabled by default, and needs to be enabled for RSVP on the eLERs, as follows:

```

# on PE-1, PE-2:
configure {
    router "Base" {
        rsvp {
            entropy-label-capability true
        }
    }
}

```

The EL can be disabled or enabled in the **mpls** context on the iLER, as follows:

```

# on PE-1, PE-2:
configure {
    router "Base" {
        mpls {
            entropy-label ?
        }
    }
}

entropy-label

rsvp-te          - Apply entropy labels to RSVP-TE LSPs
sr-te           - Apply entropy labels to SR-TE LSPs

```

The EL can also be disabled or enabled per LSP, as follows:

```

# on PE-1:
configure {
    router "Base" {
        mpls {
            lsp "LSP-PE-1-PE-2" {
                entropy-label ?
            }
        }
    }
}

entropy-label <keyword>
<keyword> - (false|true)

```

By default, the EL on the LSP inherits the EL settings in the **mpls** context, as follows:

```

# On PE-1:
configure {
    router "Base" {
        mpls {
            lsp "LSP-PE-1-PE-2" {
                info detail | match "entropy-label"
                entropy-label true
            }
        }
    }
}

```

When LSP templates are used, EL can be configured within the **lsp-template** context for single-hop and mesh point-to-point LSPs, as follows:

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp-template "LSPtemplate1" {
        type p2p-rsvp-one-hop
        entropy-label ?

entropy-label <keyword>
<keyword> - (false|true)

Entropy label for an LSP template

Warning: Modifying this element toggles
'configure router "Base" mpls lsp-template "LSPtemplate1" admin-state' automatically for
the new
value to take effect.
```

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp-template "LSPtemplate2" {
        type p2p-rsvp-mesh
        entropy-label ?

entropy-label <keyword>
<keyword> - (false|true)

Entropy label for an LSP template

Warning: Modifying this element toggles
'configure router "Base" mpls lsp-template "LSPtemplate2" admin-state' automatically for
the new
value to take effect.
```

When the EL settings are modified, the changes take effect as a result of an automatic clearing of the LSP or an automatic bouncing of the MPLS.

The following command shows that EL is enabled in MPLS:

```
[/]
A:admin@PE-1# show router mpls status | match "Label"
Entropy Label RSVP-TE : Enabled Entropy Label SR-TE : Enabled
```

The following command shows that EL is enabled in RSVP:

```
[/]
A:admin@PE-1# show router rsvp status | match "Label"
Implicit Null Label: Disabled Node-id in RR0 : Exclude
DiffServTE AdmModel: Basic Entropy Label : Enabled
```

The following command shows that EL is enabled and operational in LSP "LSP-PE-1-PE-2":

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-2" detail | match "Label"
```

**Entropy Label : Enabled**                      **Oper Entropy Label : Enabled**

The following command shows that the LSP "LSP-PE-1-PE-2" has the flag "entropy-label-capable" in the RSVP tunnel table:

```
[/]
A:admin@PE-1# show router tunnel-table protocol rsvp detail

=====
Tunnel Table (Router: Base)
=====
Destination      : 192.0.2.2/32
NextHop          : 192.168.13.2
Tunnel Flags     : exclude-for-lfa entropy-label-capable
Age              : 00h04m22s
CBF Classes      : (Not Specified)
Owner            : rsvp
Tunnel ID        : 1
Encap            : MPLS
Tunnel Label     : 524283
Preference       : 7
Tunnel MTU       : 8918
Tunnel Metric    : 16777215
LSP ID           : 5632
Max Label Stack  : 1
LSP Bandwidth    : 0
Bypass Label     : 0
LSP Weight       : 0
-----
Number of tunnel-table entries      : 1
Number of tunnel-table entries with LFA : 0
=====
```

EL is supported on many Layer 2 and Layer 3 services. In this example, VPLS 1 is configured on PE-1 with BGP EVPN-MPLS, as follows:

```
# on PE-1:
configure {
  service {
    vpls "VPLS 1" {
      service-id 1
      customer "1"
      bgp 1 { }
      bgp-evpn {
        evi 1
        mpls 1 {
          admin-state enable
          entropy-label true
          auto-bind-tunnel {
            resolution filter
            resolution-filter {
              rsvp true
            }
          }
        }
      }
    }
    sap 1/1/c2/1:11 { }
    sap 1/1/c9/1:11 { }
  }
}
```

Auto-bind-tunnel is resolved to the RSVP LSP "LSP-PE-1-PE-2". A similar configuration is applied on PE-2.

The following command shows that the EL is enabled for BGP-EVPN, as follows:

```
[/]
A:admin@PE-1# show service id 1 bgp-evpn

=====
```

```

BGP EVPN
=====
---snip---
=====
BGP EVPN MPLS Information
=====
Admin Status      : Enabled          Bgp Instance      : 1
Force Vlan Fwding : Disabled
Force QinQ Fwding : none
Route NextHop Type : system-ipv4
Control Word      : Disabled
Max Ecmp Routes   : 1
Entropy Label    : Enabled
Default Route Tag : none
Split Horizon Group: (Not Specified)
Ingress Rep BUM Lbl: Disabled
Ingress Ucast Lbl : 524280          Ingress Mcast Lbl : 524280
RestProtSrcMacAct : none
Evpn Mpls Encap   : Enabled          Evpn MplsUdp      : Disabled
Oper Group        : (none)
MH Mode           : network
Evi 3-byte Auto-RT : Disabled
Dyn Egr Lbl Limit : Disabled
Hash Label        : Disabled
Local AC Ingr Lbl : <not-allocated>

BGP EVPN MPLS Auto Bind Tunnel Information
-----
Allow-Flex-Algo-FB : Disabled
Resolution          : filter          Strict Tnl Tag    : Disabled
Max Ecmp Routes     : 1
Filter Tunnel Types : rsvp
Weighted Ecmp       : Disabled
=====

```

The iLER PE-1 adds the EL and ELI to the label stack. Traffic load-balancing is performed in P-3 where the traffic is sprayed over all eight links of the LAG. The options for LSR load-balancing are the following:

```

# on LSRs P-3 and P-4:
configure {
  system {
    load-balancing
    lsr-load-balancing ?

    lsr-load-balancing <keyword>
    <keyword> - (lbl-only|lbl-ip|ip-only|eth-encap-ip|lbl-ip-l4-teid)
    Dynamic Default - lbl-only

    Algorithm for system-wide LSR load balancing

```

By default, the load-balancing settings on P-3 are as follows:

```

# On P-3:
configure {
  system {
    load-balancing
    info detail
  }
}
-----
l2tp-load-balancing false
l4-load-balancing false
lsr-load-balancing lbl-only

```

```
mc-enh-load-balancing false
service-id-lag-hashing false
system-ip-load-balancing false
-----
```

When the EL and ELI are included in the stack, only the EL label is used for the load-balancing (**lbi-only**). It is not required to inspect the IP header or any other header.

For the reverse path, PE-2 is the iLER, and P-4 does the spraying of the packets over all links of the LAG. Therefore, LSR load-balancing with **lbi-only** is also configured on P-4.

## EVPN service using LDP tunnel with EL

ELC is disabled by default for LDP, as follows:

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

=====
LDP Status for IPv4 LSR ID 192.0.2.1
                IPv6 LSR ID ::
=====
---snip---
Admin State      : Up
IPv4 Oper State  : Up
IPv6 Oper State  : Down
---snip---

Entropy Label Capa*: False
---snip---
=====
* indicates that the corresponding row element may have been truncated.
```

The command to enable ELC is as follows:

```
# on PE-1:
configure {
  router "Base" {
    ldp {
      entropy-label-capability true
    }
  }
}
```

In the list of LDP active bindings, the egress labels that are pushed by iLER PE-1 on traffic toward an eLER capable of handling an EL get the indication "e", as follows:

```
[/]
A:admin@PE-1# show router ldp bindings active prefixes ipv4

=====
LDP Bindings (IPv4 LSR ID 192.0.2.1)
                (IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
  (S) - Static (M) - Multi-homed Secondary Support
  (B) - BGP Next Hop (BU) - Alternate Next-hop for Fast Re-Route
```

```

(I) - SR-ISIS Next Hop (0) - 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
--                                   524284e
192.168.13.2                         1/1/c1/1

192.0.2.2/32                         Swap
524282                               524284
192.168.13.2                         1/1/c1/1

192.0.2.3/32                         Push
--                                   524287
192.168.13.2                         1/1/c1/1

192.0.2.4/32                         Push
--                                   524285
192.168.13.2                         1/1/c1/1

192.0.2.4/32                         Swap
524284                               524285
192.168.13.2                         1/1/c1/1

-----
No. of IPv4 Prefix Active Bindings: 6
=====

```

Because the iLER adds the EL, only labels which are pushed get the "e"-indication. Labels which are swapped or popped do not get this label.

The detail of the tunnel table for LDP shows that the LDP tunnel toward PE-2 has tunnel flag entropy-label-capable, as follows:

```

[/]
A:admin@PE-1# show router tunnel-table protocol ldp detail

=====
Tunnel Table (Router: Base)
=====
Destination      : 192.0.2.2/32
NextHop          : 192.168.13.2
Tunnel Flags   : entropy-label-capable
Age              : 00h00m38s
CBF Classes      : (Not Specified)
Owner            : ldp
Encap            : MPLS
Tunnel ID        : 65539
Preference       : 9
Tunnel Label     : 524284
Tunnel Metric    : 30
Tunnel MTU       : 8918
Max Label Stack  : 1
-----
---snip---
=====

```



The following EVPN-VPLS uses an LDP transport tunnel and has EL enabled:

```
# on PE-1:
configure {
  service {
    vpls "VPLS 2" {
      admin-state enable
      service-id 2
      customer "1"
      bgp 1 { }
      bgp-evpn {
        evi 2
        mpls 1 {
          admin-state enable
          entropy-label true
          auto-bind-tunnel {
            resolution filter
            resolution-filter {
              ldp true
            }
          }
        }
      }
    }
    sap 1/1/c2/1:16 { }
    sap 1/1/c9/1:16 { }
  }
}
```

The configuration on PE-1 and PE-2 is similar.

EL is enabled in this service, as follows:

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

=====
BGP EVPN
=====
---snip---
=====
BGP EVPN MPLS Information
=====
Admin Status      : Enabled          Bgp Instance      : 1
Force Vlan Fwding : Disabled
Force Qinq Fwding : none
Route NextHop Type : system-ipv4
Control Word      : Disabled
Max Ecmp Routes   : 1
Entropy Label    : Enabled
Default Route Tag : none
Split Horizon Group: (Not Specified)
Ingress Rep BUM Lbl: Disabled
Ingress Ucast Lbl : 524279          Ingress Mcast Lbl : 524279
RestProtSrcMacAct : none
Evpn Mpls Encap   : Enabled          Evpn MplsUdp      : Disabled
Oper Group        : (none)
MH Mode           : network
Evi 3-byte Auto-RT : Disabled
Dyn Egr Lbl Limit : Disabled
Hash Label        : Disabled
Local AC Ingr Lbl : <not-allocated>

BGP EVPN MPLS Auto Bind Tunnel Information
-----
```

```

Allow-Flex-Algo-FB : Disabled
Resolution         : filter           Strict Tnl Tag    : Disabled
Max Ecmp Routes    : 1
Filter Tunnel Types: ldp
Weighted Ecmp      : Disabled
=====

```

## LSR load-sharing based on EL

The LSR load-sharing based on the EL results in a granular load-sharing for great numbers of flows, for example, thousands of flows. Each flow has four unique variables: source IP address, destination IP address, source UDP port, and destination UDP port. Each flow gets a unique EL and the P-nodes distribute the load over all eight links in the LAG, based on the hashing algorithm "lbl-only".

In a virtual test setup, such great numbers of flows are difficult to simulate. Nevertheless, the spread may be illustrated by comparing the traffic on each link of the LAG, once in the absence and once in the presence of externally launched traffic, as follows:

# On P-3:

1. **clear lag 1 statistics**
2. wait for X seconds (example: X=125)
3. **show lag "lag-1" statistics**

or

# On P-3:

1. **clear port <port> statistics** (where <port> is a port in the LAG)
2. wait for X seconds (example: X=125)
3. **show port statistics**

In the presence of great numbers of flows, the spread may be illustrated with:

- **monitor lag 1 rate packets**
- **monitor port <port> rate monitor port all-ethernet-rates**

As an example:, on P-3, in the absence of externally launched traffic:

```

[/]
A:admin@P-3# show lag "lag-1" statistics

=====
LAG Statistics
=====
Description      : N/A
Port-id          :          Input(Packets)  Output(Packets)
                  :          Input(Bytes)    Output(Bytes)
                  :          Input Errors   Output Errors
-----
1/1/c3/1         :          228                237
                  :          24186               25052
                  :          0                   0
1/1/c4/1         :          124                124
                  :          15872               15872
                  :          0                   0
1/1/c5/1         :          157                156
                  :          19997               19872

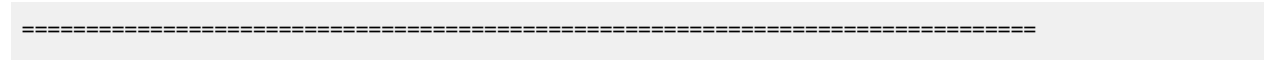
```

	0	0
1/1/c6/1	132	<b>153</b>
	16508	19352
	0	0
1/1/c7/1	153	<b>124</b>
	19352	15872
	0	0
1/1/c8/1	124	<b>124</b>
	15872	15872
	0	0
1/1/c9/1	124	<b>124</b>
	15872	15872
	0	0
1/1/c10/1	124	<b>124</b>
	15872	15872
	0	0
-----		
Totals	1166	1166
	143531	143636
	0	0
=====		

while in the presence of externally launched traffic:

```
[/]
A:admin@P-3# show lag "lag-1" statistics
```

=====			
LAG Statistics			
=====			
Description	: N/A		
Port-id		Input(Packets)	Output(Packets)
		Input(Bytes)	Output(Bytes)
		Input Errors	Output Errors
-----			
1/1/c3/1		278	<b>287</b>
		30260	30668
		0	0
1/1/c4/1		150	<b>150</b>
		19200	19200
		0	0
1/1/c5/1		189	<b>189</b>
		24075	24075
		0	0
1/1/c6/1		160	<b>186</b>
		19995	23408
		0	0
1/1/c7/1		186	<b>150</b>
		23464	19200
		0	0
1/1/c8/1		150	<b>150</b>
		19200	19200
		0	0
1/1/c9/1		150	<b>150</b>
		19200	19200
		0	0
1/1/c10/1		150	<b>150</b>
		19200	19200
		0	0
-----			
Totals		1413	1412
		174594	174151
		0	0



When thousands of flows are sent via P-3 and P-4, the flows are evenly spread in each direction.

## Conclusion

The EL is a standard-compliant way to maintain packet ordering within a conversation while load-balancing across multiple ECMP paths or links in a LAG. The EL is supported for both Layer 2 and Layer 3 services and is, therefore, a more general solution than the MPLS hash label.

# IGP Shortcuts

This chapter provides information about IGP shortcuts.

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

## Applicability

This chapter is applicable to SR OS when the feature is not related to BGP. There are no other prerequisites for this configuration. This chapter was initially written for SR OS Release 12.0.R3, but the MD-CLI in the current edition corresponds to SR OS Release 21.2.R1.

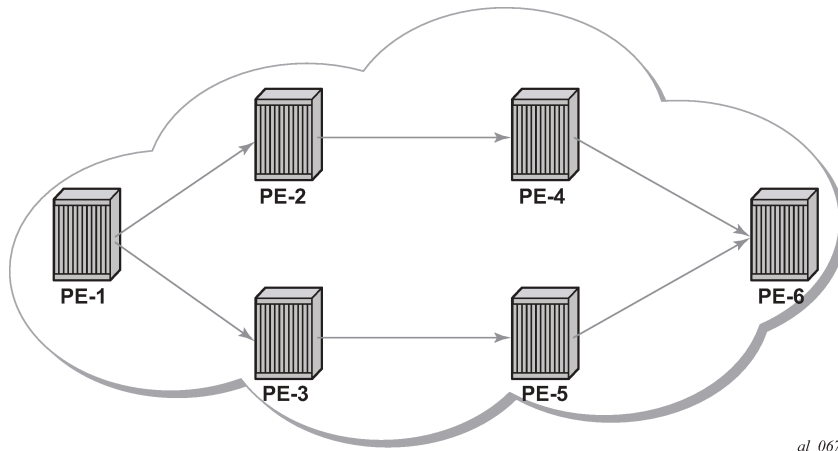
## Overview

Interior Gateway Protocols (IGPs) are routing protocols that operate inside an Autonomous System (AS). An AS is a network domain that is managed under a single administration. Because the scope of operation of an IGP is usually within an AS, IGPs are also called intra-AS protocols. The purpose of an IGP is to provide reachability information to destination nodes that are inside the domain. IGPs can be one or more of a variety of protocols, including routing protocols such as Routing Information Protocol (RIP) version 1 or 2, Open Shortest Path First (OSPF), and Intermediate System to Intermediate System (IS-IS).

IGPs such as OSPF and IS-IS are link-state protocols that use a Shortest Path First (SPF) algorithm to compute the shortest path tree to all nodes in a network. The results of such computations indicate the destination node, next hop address, and output interface, where the output interface is a physical interface. Optionally, Multi-Protocol Label Switching (MPLS) Label Switched Paths (LSPs) can be included in the SPF algorithm on the node performing the calculations, as LSPs behave as logical interfaces directly connected to remote nodes in the network. Because the SPF algorithm treats the LSPs in the same way as a physical interface (being a potential output interface), the computation results could be to select a destination node together with an output LSP, using the LSP as a shortcut through the network to the destination node.

[Figure 34: Normal SPF Tree Sourced by PE-1](#) shows a normal SPF tree sourced by PE-1 (Provider Edge-1).

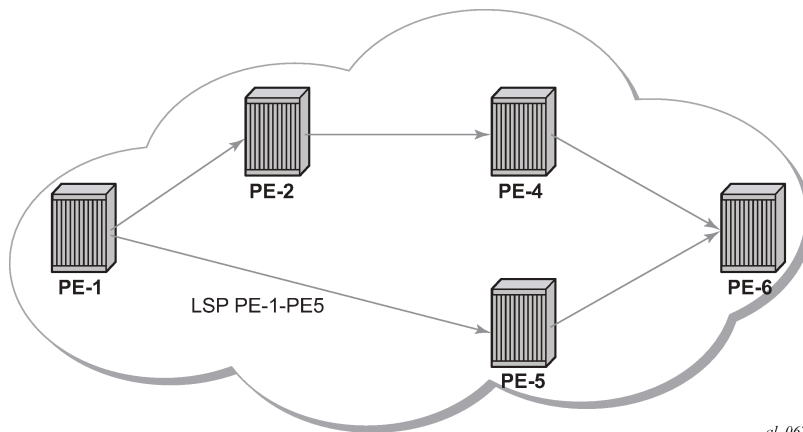
Figure 34: Normal SPF Tree Sourced by PE-1



al\_0674

If there is an LSP that connects PE-1 to PE-5, and IGP shortcuts are configured on PE-1, the SPF tree will be as shown in [Figure 35: SPF Tree Sourced by PE-1 Using LSP Shortcuts](#).

Figure 35: SPF Tree Sourced by PE-1 Using LSP Shortcuts



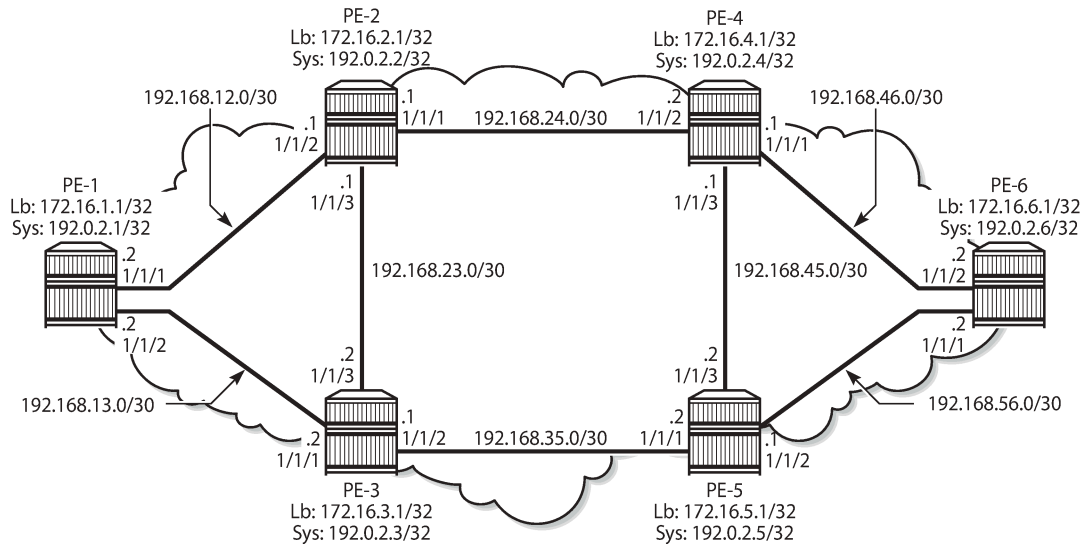
al\_0675

IGP shortcuts are enabled on a per router basis; SPF computations are independent and irrelevant to other routers, so there is no need to enable shortcuts on every single router.

The example topology used in this example is shown in [Figure 36: Example Topology](#). The setup consists of six 7750 service routers. There is a single AS and a single IGP area. The following configuration tasks should be completed first:

- IS-IS or OSPF on all interfaces within the AS (configuration has been done using IS-IS but using OSPF shows exactly the same behavior).
- Label Distribution Protocol (LDP) and Resource Reservation Protocol (RSVP) on all interfaces within the AS.

Figure 36: Example Topology



26168

In all figures, **Lb** stands for Loopback and **Sys** stands for the system IP addresses.

## Configuration

The first step is to configure the IGP (IS-IS) on all nodes, where IS-IS redistributes route reachability to all routers. To facilitate IS-IS configuration, all routers are L2-L1 capable within the same IS-IS area-id so there is only a single topology area in the network (all routers share the same topology). Traffic engineering (TE) is enabled on the IGP as it is a requirement for RSVP. The metric is using the default values: because no reference bandwidth command is used, the default metric of 10 is applicable on all interfaces. The configuration for PE-2 is as follows.

```
[/]
A:admin@PE-2# configure {
  router "Base" {
    interface "int-PE-2-PE-1" {
      port 1/1/2
      ipv4 {
        primary {
          address 192.168.12.2
          prefix-length 30
        }
      }
    }
    interface "int-PE-2-PE-3" {
      port 1/1/3
      ipv4 {
        primary {
          address 192.168.23.1
          prefix-length 30
        }
      }
    }
  }
}
```

```

interface "int-PE-2-PE-4" {
    port 1/1/1
    ipv4 {
        primary {
            address 192.168.24.1
            prefix-length 30
        }
    }
}
interface "system" {
    ipv4 {
        primary {
            address 192.0.2.2
            prefix-length 32
        }
    }
}
}
isis 0 {
    admin-state enable
    traffic-engineering true
    area-address [49.0001]
    interface "int-PE-2-PE-1" {
        interface-type point-to-point
    }
    interface "int-PE-2-PE-3" {
        interface-type point-to-point
    }
    interface "int-PE-2-PE-4" {
        interface-type point-to-point
    }
    interface "system" {
        passive true
    }
}
}
}
}

```

The configuration for the other nodes is similar. The IP addresses can be derived from [Figure 36: Example Topology](#).

The global route table (GRT) for PE-2 is as follows:

```

[/]
A:admin@PE-2# show router route-table
=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]                               Type  Proto  Age           Pref
  Next Hop[Interface Name]                       Metric
-----
192.0.2.1/32                                     Remote  ISIS   00h02m31s    15
    192.168.12.1                                10
192.0.2.2/32                                     Local   Local  00h02m32s    0
    system                                       0
192.0.2.3/32                                     Remote  ISIS   00h02m05s    15
    192.168.23.2                                10
192.0.2.4/32                                     Remote  ISIS   00h01m51s    15
    192.168.24.2                                10
192.0.2.5/32                                     Remote  ISIS   00h01m31s    15
    192.168.23.2                                20
192.0.2.6/32                                     Remote  ISIS   00h01m09s    15
    192.168.24.2                                20

```



```

192.168.12.0/30          Local  Local  00h02m32s  0
   int-PE-2-PE-1
192.168.13.0/30          Remote ISIS  00h02m31s  15
   192.168.12.1
192.168.23.0/30          Local  Local  00h02m32s  0
   int-PE-2-PE-3
192.168.24.0/30          Local  Local  00h02m32s  0
   int-PE-2-PE-4
192.168.35.0/30          Remote ISIS  00h02m05s  15
   192.168.23.2
192.168.45.0/30          Remote ISIS  00h01m51s  15
   192.168.24.2
192.168.46.0/30          Remote ISIS  00h01m51s  15
   192.168.24.2
192.168.56.0/30          Remote ISIS  00h01m31s  15
   192.168.23.2
-----
No. of Routes: 14
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
=====

```

## LDP and RSVP Shortcuts

Interface Label Distribution Protocol (iLDP) is enabled on all interfaces (except system interfaces, which is not allowed) in all routers. The configuration on all nodes is similar and the IP addresses are derived from [Figure 36: Example Topology](#). The configuration of PE-4 is as follows:

```

[/]
A:admin@PE-4# configure {
  router "Base" {
    ldp {
      interface-parameters {
        interface "int-PE-4-PE-2" {
          ipv4 {
          }
        }
        interface "int-PE-4-PE-5" {
          ipv4 {
          }
        }
        interface "int-PE-4-PE-6" {
          ipv4 {
          }
        }
      }
    }
  }
}

```

With iLDP enabled, PE-4 establishes iLDP sessions with its directly connected neighbors, as follows:

```

[/]
A:admin@PE-4# show router ldp session ipv4

```

```

=====
LDP IPv4 Sessions
=====

```

Peer LDP Id	Adj Type	State	Msg Sent	Msg Recv	Up Time
192.0.2.2:0	Link	<b>Established</b>	54	54	0d 00:01:54
192.0.2.5:0	Link	<b>Established</b>	49	50	0d 00:01:36
192.0.2.6:0	Link	<b>Established</b>	42	43	0d 00:01:19

No. of IPv4 Sessions: 3

The following tunnel table shows that there is a Label Switched Path (LSP) to every other router. The reason is that the LDP label distribution mode is downstream unsolicited (DU) by default, originating label bindings for system addresses only (which are used by iLDP as transport address by default). The command also shows the preference of the LSPs (where the preference is 9 for LDP) and the metric of the LSPs (the metric is inherited from the IGP, each hop counts as a metric of 10), as follows. The metric to destinations PE-1 and PE-3 is 20 because there are two hops in between (PE-4 is two hops away from PE-1 and PE-3).

```
[/]
A:admin@PE-4# show router tunnel-table

=====
IPv4 Tunnel Table (Router: Base)
=====
Destination      Owner      Encap TunnelId  Pref  Nexthop      Metric
  Color
-----
192.0.2.1/32     ldp       MPLS  65538      9    192.168.24.1  20
192.0.2.2/32     ldp       MPLS  65537      9    192.168.24.1  10
192.0.2.3/32     ldp       MPLS  65539      9    192.168.24.1  20
192.0.2.5/32     ldp       MPLS  65540      9    192.168.45.2  10
192.0.2.6/32     ldp       MPLS  65541      9    192.168.46.2  10
-----
Flags: B = BGP or MPLS backup hop available
      L = Loop-Free Alternate (LFA) hop available
      E = Inactive best-external BGP route
      k = RIB-API or Forwarding Policy backup hop
=====
```

In order to configure RSVP shortcuts, RSVP must be enabled on all interfaces where traffic engineering is required, but in this example, MPLS and RSVP are enabled on all interfaces of the network. By default, MPLS is enabled on the system interface, therefore, it need not be configured explicitly. When RSVP is in **admin-state enable**, it is automatically configured on the interfaces where MPLS is configured. The configuration for PE-6 is as follows.

```
[/]
A:admin@PE-6# configure {
  router "Base" {
    mpls {
      admin-state enable
      interface "int-PE-6-PE-4" {
      }
      interface "int-PE-6-PE-5" {
      }
    }
    rsvp {
      admin-state enable
      interface "int-PE-6-PE-4" {
      }
      interface "int-PE-6-PE-5" {
      }
    }
  }
}
```

```

    }
  }
}

```

The configuration of the other nodes is similar. The IP addresses can be derived from [Figure 36: Example Topology](#). Because there are no RSVP LSPs configured yet, the tunnel table has no RSVP LSPs and only contains LDP LSPs.

## LDP Static Route (IP Tunneled in LDP Tunnel)

Using LDP LSP shortcuts for static route resolution enables forwarding of IPv4 packets over LDP LSPs instead of using a regular IP next hop. In other words, the traffic to the resolved static routes is forwarded using MPLS LDP LSP rather than plain IP.

The configuration defines a static route pointing to the destination PE (remote loopback, which is an indirect next hop in the example), and explicitly indicates that it should use LDP rather than IGP. Taking PE-1 and PE-6 as an example, two loopback interfaces are configured (172.16.X.1/32), where X = PE number, and a static route is defined according to the preceding explanation. The following shows the configuration on PE-1.

```

[/]
A:admin@PE-1# configure {
  router "Base" {
    interface "Loopback" {
      ipv4 {
        primary {
          address 172.16.1.1
          prefix-length 32
        }
      }
      loopback
    }
    static-routes {
      route 172.16.6.1/32 route-type unicast {
        indirect 192.0.2.6 {
          admin-state enable
          tunnel-next-hop {
            disallow-igp true
            resolution filter
            resolution-filter {
              ldp true
            }
          }
        }
      }
    }
  }
}

```

Looking at the GRT or forwarding information base (FIB), there are two new entries corresponding to the two configured loopback interfaces. One entry has the protocol set to LOCAL (the local loopback on the PE), and the other entry has the protocol set to STATIC, where the next hop is reached using an LDP LSP.

```

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

```

```

=====
FIB Display

```

```

=====
Prefix [Flags]                                Protocol
NextHop
-----
172.16.1.1/32                                LOCAL
  172.16.1.1 (loopback)
172.16.6.1/32                                STATIC
  192.0.2.6 (Transport:LDP)
192.0.2.1/32                                LOCAL
  192.0.2.1 (system)
192.0.2.2/32                                ISIS
  192.168.12.2 (int-PE-1-PE-2)
192.0.2.3/32                                ISIS
  192.168.13.2 (int-PE-1-PE-3)
192.0.2.4/32                                ISIS
  192.168.12.2 (int-PE-1-PE-2)
192.0.2.5/32                                ISIS
  192.168.13.2 (int-PE-1-PE-3)
192.0.2.6/32                                ISIS
  192.168.12.2 (int-PE-1-PE-2)
192.168.12.0/30                             LOCAL
  192.168.12.0 (int-PE-1-PE-2)
192.168.13.0/30                             LOCAL
  192.168.13.0 (int-PE-1-PE-3)
192.168.23.0/30                             ISIS
  192.168.12.2 (int-PE-1-PE-2)
192.168.24.0/30                             ISIS
  192.168.12.2 (int-PE-1-PE-2)
192.168.35.0/30                             ISIS
  192.168.13.2 (int-PE-1-PE-3)
192.168.45.0/30                             ISIS
  192.168.12.2 (int-PE-1-PE-2)
192.168.46.0/30                             ISIS
  192.168.12.2 (int-PE-1-PE-2)
192.168.56.0/30                             ISIS
  192.168.13.2 (int-PE-1-PE-3)
-----
Total Entries : 16
=====

```

The following output shows that a **ping** sourced by the loopback interface on PE-1 is able to reach the loopback interface on PE-6, and **traceroute** demonstrates that the traffic is following the LDP LSP. The **ping** and **traceroute** traffic cannot follow the IGP path because the static route command states that the IGP is disallowed when no LDP LSP toward PE-6 is available (also, the loopback interfaces are not enabled on IS-IS).

```

[/]
A:admin@PE-1# ping 172.16.6.1 source ping 172.16.6.1 source-address 172.16.1.1
PING 172.16.6.1 56 data bytes
64 bytes from 172.16.6.1: icmp_seq=1 ttl=64 time=3.18ms.
64 bytes from 172.16.6.1: icmp_seq=2 ttl=64 time=3.30ms.
. 64 bytes from 172.16.6.1: icmp_seq=3 ttl=64 time=3.03ms.
. 64 bytes from 172.16.6.1: icmp_seq=4 ttl=64 time=3.25ms.
. 64 bytes from 172.16.6.1: icmp_seq=5 ttl=64 time=2.72ms.

---- 172.16.6.1 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 2.72ms, avg = 3.10ms, max = 3.30ms, stddev = 0.208ms

[/]
A:admin@PE-1# traceroute 172.16.6.1 source traceroute 172.16.6.1 source-address 172.16.1.1

```

```
traceroute to 172.16.6.1 from 172.16.1.1, 30 hops max, 40 byte packets
 1  0.0.0.0  * * *
 2  0.0.0.0  * * *
 3  172.16.6.1 (172.16.6.1)    2.79 ms  2.35 ms  2.61 ms
```

With the **traceroute** command, there are three hops from PE-1 to PE-6. There is no information regarding IP for the first two hops because the traffic is encapsulated in an MPLS LSP. The reason why the hops are displayed even when there is an MPLS LSP tunnel is because by default, the SR router propagates (copies) the Time To Live (TTL) from the IP header in the MPLS header. This is known as uniform mode.

However, a service provider might not want to show how many MPLS hops (nodes) there are in their network if a **traceroute** command is executed from outside their network. To prevent internal hops being shown, **no propagate** commands are needed in the LDP configuration, as follows. This is known as pipe mode.

```
[/]
A:admin@PE-1# configure {
  router "Base" {
    ldp {
      shortcut-local-ttl-propagate false
      shortcut-transit-ttl-propagate false
    }
  }
}
```

When TTL propagation is disabled, the hops are not displayed any longer when running the **traceroute** command.

```
[/]
A:admin@PE-1# traceroute 172.16.6.1 source-address 172.16.1.1
traceroute to 172.16.6.1 from 172.16.1.1, 30 hops max, 40 byte packets
 1  172.16.6.1 (172.16.6.1)    1.86 ms  2.04 ms  1.96 ms
```

For more information about uniform mode and pipe mode, see the "Tunneling of ICMP Reply Packets over MPLS LSPs" chapter in the *7450 ESS, 7750 SR, 7950 XRS, and VSR MPLS Advanced Configuration Guide for Classic CLI*.

## RSVP Static Route (IP Tunneled in RSVP Tunnel)

Using RSVP LSP shortcuts for static route resolution enables forwarding of IPv4 packets over RSVP LSPs instead of using a regular IP next hop. In other words, the traffic to the resolved static routes is forwarded using an MPLS RSVP LSP rather than plain IP.

The configuration defines a static route pointing to a destination PE (remote loopback, which is an indirect next hop in the example), and explicitly indicates that it should use RSVP rather than IGP. Taking PE-6 and PE-1 as an example, two loopback interfaces are configured (172.16.X.1/32), where X = PE number, and a static route is defined according to the preceding explanation. The following shows the configuration on PE-6.

```
[/]
A:admin@PE-6# configure {
  router "Base" {
    static-routes {
      route 172.16.1.1/32 route-type unicast {
        indirect 192.0.2.1 {
          admin-state enable
        }
      }
    }
  }
}
```



The RSVP LSP is used to resolve the indirect next hop (PE-1 system address) in the static route (the LSP used is identified with the tunnel ID, in this case 1), therefore, the route for prefix 172.16.1.1 in the GRT looks as follows:

```
[/]
A:admin@PE-6# show router route-table 172.16.1.1

=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]                Type   Proto   Age      Pref
  Next Hop[Interface Name]                Metric
-----
172.16.1.1/32                    Remote Static  00h00m00s  5
  192.0.2.1 (tunneled:RSVP: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
=====
```

As in the LDP shortcut with static route example, between PE-6 and PE-1, TTL propagation is disabled, as follows.

```
[/]
A:admin@PE-1# configure {
  router "Base" {
    ldp {
      shortcut-local-ttl-propagate false
      shortcut-transit-ttl-propagate false
    }
  }
}
```

The output is the following when running a traceroute:

```
[/]
A:admin@PE-6# traceroute 172.16.1.1 source traceroute 172.16.1.1 source-address 172.16.6.1
traceroute to 172.16.1.1 from 172.16.6.1, 30 hops max, 40 byte packets
 1 172.16.1.1 (172.16.1.1)  2.81 ms  3.24 ms  3.02 ms
```

The two static routes that have been defined to use the LDP and RSVP shortcuts follow the static routes default values and have a preference of 5 and a metric of 1.

## LDP Shortcut for IGP Route Resolution

Using LDP shortcuts for IGP route resolution enables forwarding of packets to IGP learned routes over an LDP LSP. The default is to disable the LDP shortcut across all interfaces in the node.

When LDP shortcuts are enabled, LDP populates the Route Table Manager (RTM) with next hop entries corresponding to all prefixes for which it activated an LDP Forwarding Equivalence Class (FEC). For a prefix, two route entries are populated in the RTM. One corresponds to the LDP shortcut next hop and has an owner of LDP. The other one is the regular IP next hop. The LDP shortcut next hop always takes preference over the regular IP next hop for forwarding user packets and specific control packets over an outgoing interface to the route next hop.

When LDP has activated a FEC for a prefix and programmed the RTM, it also programs the ingress tunnel table in the line card with the LDP tunnel information.

When an IPv4 packet is received on an ingress network interface and the preferred RTM entry corresponds to an LDP shortcut, a subscriber Internet Enhanced Service (IES) interface, or a regular IES interface, the lookup of the packet by the ingress line card results in the packet being sent labeled with the label stack corresponding to the Next Hop Label Forwarding Entry (NHLFE) of the LDP LSP. If the preferred RTM entry corresponds to an IP next hop, the IPv4 packet is forwarded unlabeled. The activation of the FEC by LDP is done by performing an exact match with an IGP route prefix in the RTM, but it can also be done by performing a longest prefix match with an IGP route in the RTM if the aggregate-prefix-match option is enabled globally in LDP.

## Handling of Control Packets

All control plane packets will not see the LDP shortcut route entry in the RTM with the exception of the following control packets which will be forwarded over an LDP shortcut when enabled:

- A locally generated or in transit ICMP ping and UDP traceroute of an IGP route. The transit message appears as a user packet to the ingress LER node.
- A locally generated response to a received ICMP ping or UDP traceroute message.

All other control plane packets that require an RTM lookup and have knowledge of which destination is reachable over the LDP shortcut will continue to be forwarded over the IP next hop route in the RTM.

## Handling of Multicast Packets

LDP shortcuts apply to unicast FEC types and are used for forwarding IP unicast packets in the data path. IP multicast packets forwarded over an multicast Label Distribution Protocol (mLDP) Point-to-Multi-Point (P2MP) LSP make use of a multicast FEC and thus cannot make use of the LDP unicast shortcut.

## ECMP Considerations

When Equal Cost Multi-Path (ECMP) is enabled and multiple equal cost next hops exist for the IGP route, the ingress line card will spray the packets for this route based on the hashing routine supported for IPv4 packets. When the preferred RTM entry corresponds to an LDP shortcut route, spraying is performed across the multiple next hops for the LDP FEC. The FEC next hops can either be direct link LDP neighbors, or T-LDP (targeted LDP) neighbors reachable over RSVP LSPs in the case of LDP-over-RSVP, but not both. This is as per ECMP for LDP in the existing implementation. When the preferred RTM entry corresponds to a regular IP route, spraying will be performed across regular IP next hops for the prefix. Spraying across regular IP next hops and LDP shortcut next hops concurrently is not supported.

Configuring IGP LDP shortcuts is straightforward, and only applies to the node where there is interest to provision the LDP shortcut. In this example, only PE-1 is provisioned with LDP shortcuts, as follows:

```
[/]
A:admin@PE-1# configure {
  router "Base" {
    ldp {
      ldp-shortcut {
        ipv4 true
      }
    }
  }
}
```



```
}  
}
```

Now, all tunnel LSPs that resolve an IGP next hop will replace the IP next hops, as shown in the following output:

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

Route Table (Router: Base)				
Dest Prefix[Flags] Next Hop[Interface Name]	Type	Proto	Age Metric	Pref
192.0.2.1/32 system	Local	Local	00h50m05s 0	0
<b>192.0.2.2/32</b> <b>192.168.12.2 (tunneled)</b>	<b>Remote</b>	<b>LDP</b>	<b>00h00m00s</b> <b>10</b>	<b>9</b>
<b>192.0.2.3/32</b> <b>192.168.13.2 (tunneled)</b>	<b>Remote</b>	<b>LDP</b>	<b>00h00m00s</b> <b>10</b>	<b>9</b>
<b>192.0.2.4/32</b> <b>192.168.12.2 (tunneled)</b>	<b>Remote</b>	<b>LDP</b>	<b>00h00m00s</b> <b>20</b>	<b>9</b>
<b>192.0.2.5/32</b> <b>192.168.13.2 (tunneled)</b>	<b>Remote</b>	<b>LDP</b>	<b>00h00m00s</b> <b>20</b>	<b>9</b>
<b>192.0.2.6/32</b> <b>192.168.12.2 (tunneled)</b>	<b>Remote</b>	<b>LDP</b>	<b>00h00m00s</b> <b>30</b>	<b>9</b>
192.168.12.0/30 int-PE-1-PE-2	Local	Local	00h50m05s 0	0
192.168.13.0/30 int-PE-1-PE-3	Local	Local	00h50m05s 0	0
192.168.23.0/30 192.168.12.2	Remote	ISIS	00h49m41s 20	15
192.168.24.0/30 192.168.12.2	Remote	ISIS	00h49m41s 20	15
192.168.35.0/30 192.168.13.2	Remote	ISIS	00h49m20s 20	15
192.168.45.0/30 192.168.12.2	Remote	ISIS	00h49m06s 30	15
192.168.46.0/30 192.168.12.2	Remote	ISIS	00h49m06s 30	15
192.168.56.0/30 192.168.13.2	Remote	ISIS	00h48m46s 30	15

```
-----  
No. of Routes: 14  
Flags: n = Number of times nexthop is repeated  
       B = BGP backup route available  
       L = LFA nexthop available  
       S = Sticky ECMP requested  
=====
```

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

FIB Display	
Prefix [Flags] NextHop	Protocol
192.0.2.1/32 192.0.2.1 (system)	LOCAL

```
-----
```

```

192.0.2.2/32                                LDP
  192.0.2.2 (Transport:LDP)
192.0.2.3/32                                LDP
  192.0.2.3 (Transport:LDP)
192.0.2.4/32                                LDP
  192.0.2.4 (Transport:LDP)
192.0.2.5/32                                LDP
  192.0.2.5 (Transport:LDP)
192.0.2.6/32                                LDP
  192.0.2.6 (Transport:LDP)
192.168.12.0/30                              LOCAL
  192.168.12.0 (int-PE-1-PE-2)
192.168.13.0/30                              LOCAL
  192.168.13.0 (int-PE-1-PE-3)
192.168.23.0/30                              ISIS
  192.168.12.2 (int-PE-1-PE-2)
192.168.24.0/30                              ISIS
  192.168.12.2 (int-PE-1-PE-2)
192.168.35.0/30                              ISIS
  192.168.13.2 (int-PE-1-PE-3)
192.168.45.0/30                              ISIS
  192.168.12.2 (int-PE-1-PE-2)
192.168.46.0/30                              ISIS
  192.168.12.2 (int-PE-1-PE-2)
192.168.56.0/30                              ISIS
  192.168.13.2 (int-PE-1-PE-3)
-----
Total Entries : 14
-----
=====

```

Applying LDP IGP shortcuts only on PE-1 implies that IP traffic from PE-1 to any of the system addresses of the rest of the nodes will use the LDP shortcut, however, the traffic replied from any PE back to PE-1 will be native IP because IGP shortcuts have not been provisioned in the other nodes.

## RSVP Shortcut for IGP Route Resolution

Using RSVP LSP shortcuts when resolving IGP routes enables forwarding of packets to IGP learned routes over an RSVP LSP. The use of RSVP shortcuts for resolving IGP routes is enabled at the IS-IS (or OSPF) routing protocol level or at the LSP level, and instructs IS-IS and OSPF to include RSVP LSPs originating on this node and terminating on the system address (router ID) of a remote node and considers them as direct links. RSVP LSPs with a destination address corresponding to an interface address or any other loopback interface address of a remote node are automatically not considered by IS-IS or OSPF.

By default, RSVP shortcuts are disabled in all IGP instances.

RSVP LSPs are included in the IGP SPF computation with the following characteristics:

- RSVP LSP is modeled as a point-to-point link IP interface and its metric is used in the computation of the shortest path of IGP routes
- Next hop and interface include the NHLFE of the shortcut LSP when the IGP path cost using the RSVP LSP is the best.
- Shortcuts are not used when the destination RSVP LSP is in a different IGP area. In addition, IGP adjacencies across an RSVP LSP are not supported.

RSVP shortcuts are enabled at IGP instance level as follows:

```
[/]
```

```
A:admin@<all PEs># configure {
  router "Base" {
    isis 0 {
      igp-shortcut {
        admin-state enable
        tunnel-next-hop {
          family ipv4 {
            resolution filter
            resolution-filter {
              rsvp true
            }
          }
        }
      }
    }
  }
}
```

The configuration can be done at the IGP level or per LSP level. When RSVP shortcuts are enabled at the IGP instance level, all RSVP LSPs originating on this node are eligible by default. The user can, however, exclude a specific RSVP LSP from being used as a shortcut for resolving IGP routes by entering the command (e.g. on PE-6):

```
[/]
A:admin@PE-6# configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-6-PE-1" {
        delete igp-shortcut
      }
    }
  }
}
```

As RSVP shortcuts can coexist with LDP shortcuts or IP next hops, SPF computation and path selection follows the procedures in RFC 3906:

- SPF picks the RSVP shortcut next hop if there is an RSVP LSP directly to that address regardless of the path cost compared to the IGP next hop.
- SPF picks the RSVP shortcut next hop or the IGP next hop based on path lowest cost if there is an IGP path to the prefix that does not go via the tail-end of the LSP.
- If the IGP next hop is picked, then it can be an LDP shortcut next hop or a regular IP next hop. The LDP shortcut next hop always has preference over the regular IP next hop.

## Handling of Control Packets

All control plane packets requiring an RTM lookup and whose destination is reachable over the RSVP shortcut are forwarded over the shortcut. This is because the RTM keeps a single route entry for each prefix, except if there is ECMP over different outgoing interfaces. Interface bound control packets are not impacted by the RSVP shortcut because RSVP LSPs with a destination address different than the router ID are not included by IGP in its SPF calculation.

RSVP shortcuts for IGP shortcut resolution should only be used with CSPF LSPs or with fully explicit path non-CSPF LSPs. RSVP hop-by-hop Path messages will try to use the shortcut and consequently LSPs without CSPF enabled, or that use a loose/empty hop path, will not come up. However, LSPs with CSPF enabled or using a strict hop path will come up. This is because in the former case, the RTM lookup to get

the next hop results in using the shortcut and so the path messages are sent directly to the destination of the LSP, where they are dropped. With CSPF enabled, the next hop (and the entire path) is provided by CSPF and the path messages are sent unlabeled to the directly connected neighbor which corresponds to the next hop of the destination of the LSP. Similar processing occurs if a strict hop path is used, as is the case in the following example.

## Handling of Multicast Packets

IP multicast packets cannot be forwarded over an RSVP shortcut, they can only be forwarded over an RSVP P2MP LSP. However, RSVP shortcut routes appear in the RTM and are seen by all applications when they are the best route. When the Reverse Path Forwarding (RPF) check for the source of the multicast packet matches an RSVP shortcut route, the check will pass if both the RSVP shortcut and the multicast-import options are enabled in the IGP, as follows, because the RTM is populated with next hops only and not with tunnels (RPFs will fail for source prefixes resolved to a tunnel NH).

```
[/]
A:admin@PE-6# configure {
  router "Base" {
    isis multicast-import ?

  multicast-import

  ipv4          - Submit IPv4 routes into the multicast RPF of the RTM
  ipv6          - Submit IPv6 routes into the multicast RPF of the RTM
```

The unicast RTM can still use the tunnel next hop for the same prefix. SPF keeps track of both the direct first hop and the tunneled first hop of a node which is added to the Dijkstra tree.

## ECMP Considerations

When ECMP is enabled and multiple equal cost paths exist for the route over a set of tunnel next hops based on the hashing routine supported for IPv4 packets, there are two possibilities:

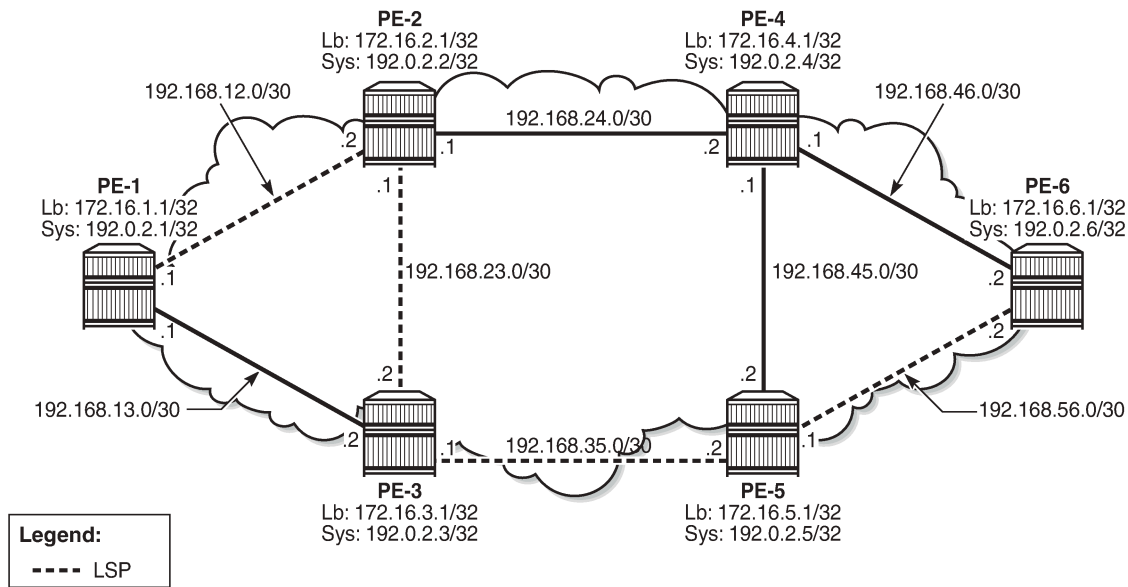
- Destination is tunnel endpoint: the system selects the tunnel with lowest tunnel ID (IP next hop is never used).
- Destination is different from the tunnel endpoint: it selects tunnel endpoints when the LSP metric is not greater than the IGP cost and it prefers tunnel endpoint over IP next hop.

ECMP is not performed across the IP and tunnel next hops simultaneously.

## RSVP Shortcuts Configuration

Configuring RSVP LSP shortcuts is straightforward, and only applies to the node where there is interest to provision the RSVP shortcut. Two LSPs, from PE-6 to PE-1 and from PE-1 to PE-6, with strict hops, are provisioned according to [Figure 37: LSPs Between PE-1 and PE-6](#).

Figure 37: LSPs Between PE-1 and PE-6



25826

The configuration on PE-1 and PE-6 is similar (replacing the IP addresses), so only the configuration for PE-6 is shown:

```
[/]
A:admin@PE-6# configure {
  router "Base" {
    isis 0 {
      igp-shortcut {
        admin-state enable
        tunnel-next-hop {
          family ipv4 {
            resolution filter
            resolution-filter {
              rsvp true
            }
          }
        }
      }
    }
  }
}
```

```
[/]
A:admin@PE-6# configure {
  router "Base" {
    mpls {
      path "path-to-PE-1" {
        admin-state enable
        hop 10 {
          ip-address 192.0.2.5
          type strict
        }
        hop 20 {
          ip-address 192.0.2.3
        }
      }
    }
  }
}
```



```
B = BGP backup route available
L = LFA nexthop available
S = Sticky ECMP requested
```

The RSVP LSP in the output has a metric of 16777215, the LSP administrative metric matches the maximum value allowed for an IS-IS link using the wide-metric (24-bit value with a range of [0 — 16777215]). The following metric rules apply:

- A dynamic strict path non-CSPF LSP has the maximum metric (16777215).
- A dynamic CSPF LSP has a metric equal to the cumulative IGP cost.
  - If the user enabled the use of the TE metric on this LSP (`[/] A:admin@<a PE># configure { router "Base" { mpls { lsp <LSP Name> { path-computation-method local-cspf metric-type te } } }`), then the metric for the LSP is the maximum (16777215).
  - If the user enabled the use of the TE metric on this LSP and provisioned a specific metric on the lsp (`[/] A:admin@<a PE># configure { router "Base" { mpls { lsp <LSP Name> { path-computation-method local-cspf metric-type te metric <value> } } }`), then the metric for the LSP is the one provisioned. When configuring the metric of an LSP, the parameter "metric-type" is not required.
- A static LSP has a maximum metric (16777215).
- Manual and dynamic bypass LSPs have the maximum metric (16777215).

The RSVP shortcuts section detailed the importance of the LSP metric when using CSPF LSPs or when importing RSVP tunnel links into the IGP. The LSP metric can be inherited from the IGP, or can be manually modified by configuring a specific LSP metric or relative metric offset. Because IP and LDP FECs resolve to RSVP LSPs when the metric is equal or lower compared to the regular routing metric, configuring a specific static LSP metric (lower than the IGP metric) or relative metric offset is strongly recommended when using RSVP shortcuts, so that the GRT and LDP FEC resolution will always prefer the RSVP LSP shortcuts when the CSPF path computation is not using the shortest path.

For the preceding example, the first rule applies.

## Advertising RSVP LSP Tunnel Links in the IGP: Forwarding Adjacency Feature

If configured, an RSVP LSP can also be advertised into the IGP similar to regular links so that other routers in the network can include that RSVP LSP into their SPF computations. The forwarding adjacency feature can be enabled independently from the RSVP shortcut feature in CLI. If both are configured for an IGP instance, the forwarding adjacency takes precedence. An RSVP LSP must exist in the reverse direction in order for the advertised link to pass the bi-directional link check and be usable by other routers in the network. However, this is not required for the node which originates the LSP. The LSP is advertised as an unnumbered point-to-point link and the Link State Protocol data unit (LSP) and Link State Advertisement (LSA) have no traffic engineering opaque sub-TLVs as per RFC 3906.

Reusing the RSVP IGP shortcuts set up previously (PE-1 and PE-6 RSVP IGP shortcut example according to [Figure 37: LSPs Between PE-1 and PE-6](#)), the outcome is a route linked with an RSVP LSP as next hop, as follows:

```
[/]
A:admin@PE-6# show router route-table 192.0.2.1/32

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

```

Next Hop[Interface Name]                               Metric
-----
192.0.2.1/32                                           Remote  ISIS    00h02m25s 15
    192.0.2.1 (tunneled:RSVP:2)                       16777215
-----
No. of Routes: 1
Flags: n = Number of times nexthop is repeated
      B = BGP backup route available
      L = LFA nexthop available
      S = Sticky ECMP requested
=====

```

The route tunneled through RSVP has a metric of 16777215, so it is not used by PE-6 GRT to reach any other routes because the metric is very high. After enabling the forwarding adjacency feature (tunnel links) to use shortcuts in the configuration, PE-1 and PE-6 have a direct connection through the RSVP LSP (as a virtual link). This configuration command must be executed in both routers.

```

[/]
A:admin@PE-1(&6)# configure {
  router "Base" {
    isis 0 {
      advertise-tunnel-link
    }
  }
}

```

When the shortcut is advertised by IS-IS, the route will disappear from the RTM because the metric of the shortcut is greater than the IGP cost.

```

[/]
A:admin@PE-6# show router route-table 192.0.2.1/32

=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]                               Type  Proto   Age           Pref
  Next Hop[Interface Name]                       Metric
-----
192.0.2.1/32                                     Remote ISIS    00h01m22s 15
    192.168.46.1                                30
-----
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
=====

```

If the LSP is reconfigured to use a metric equal to or smaller than the IGP cost, the router PE-6 will use the RSVP shortcut again. In the example, the LSP is reconfigured with a metric of 30 (e.g. on PE-6):

```

[/]
A:admin@PE-6# configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-6-PE-1-strict" {
        metric 30
      }
    }
  }
}

```



```
}

```

Now the shortcut shows up as the preferred next hop to reach PE-1 from PE-6.

```
[/]
A:admin@PE-6# show router route-table 192.0.2.1/32

=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]                Type  Proto  Age           Pref
  Next Hop[Interface Name]                Metric
-----
192.0.2.1/32                      Remote ISIS   00h00m53s  15
  192.0.2.1 (tunneled:RSVP:2)          30

-----
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
=====
```

As explained earlier, this could be combined with ECMP, so if ECMP is configured to 2, the system shows the two equal cost paths.

```
[/]
A:admin@PE-6# configure {
  router "Base" {
    ecmp 2
  }
}

[/]
A:admin@PE-6# show router route-table 192.0.2.1/32

=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]                Type  Proto  Age           Pref
  Next Hop[Interface Name]                Metric
-----
192.0.2.1/32                      Remote ISIS   00h00m21s  15
  192.0.2.1 (tunneled:RSVP:2)          30
192.0.2.1/32                      Remote ISIS   00h00m21s  15
  192.168.46.1                          30

-----
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 GRT on PE-4 displays the route to reach PE-1 (192.0.2.1/32) with a metric of 20 via PE-2 as next hop. Although PE-6 is announcing the RSVP LSP-PE-6-PE-1 to the other routers, the LSP shortcut is not used by PE-4, because the metric to reach PE-6 (10) plus the metric of the LSP shortcut from PE-6 to PE-1 (metric 30) is greater than 20.

```
[/]
A:admin@PE-4# show router route-table
```

```

=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]                               Type  Proto  Age           Pref
  Next Hop[Interface Name]                       Metric
-----
192.0.2.1/32                                     Remote ISIS   01h06m15s  15
      192.168.24.1                               20
192.0.2.2/32                                     Remote ISIS   01h06m15s  15
      192.168.24.1                               10
192.0.2.3/32                                     Remote ISIS   01h06m15s  15
      192.168.24.1                               20
192.0.2.4/32                                     Local  Local   01h06m16s  0
      system                                     0
192.0.2.5/32                                     Remote ISIS   01h05m50s  15
      192.168.45.2                               10
192.0.2.6/32                                     Remote ISIS   01h05m33s  15
      192.168.46.2                               10
192.168.12.0/30                                  Remote ISIS   01h06m15s  15
      192.168.24.1                               20
192.168.13.0/30                                  Remote ISIS   01h06m15s  15
      192.168.24.1                               30
192.168.23.0/30                                  Remote ISIS   01h06m15s  15
      192.168.24.1                               20
192.168.24.0/30                                  Local  Local   01h06m16s  0
      int-PE-4-PE-2                              0
192.168.35.0/30                                  Remote ISIS   01h05m50s  15
      192.168.45.2                               20
192.168.45.0/30                                  Local  Local   01h06m16s  0
      int-PE-4-PE-5                              0
192.168.46.0/30                                  Local  Local   01h06m16s  0
      int-PE-4-PE-6                              0
192.168.56.0/30                                  Remote ISIS   01h05m33s  15
      192.168.45.2                               20
-----
No. of Routes: 14
Flags: n = Number of times nexthop is repeated
      B = BGP backup route available
      L = LFA nexthop available
      S = Sticky ECMP requested
=====

```

If the metric of the LSP LSP-PE-6-PE-1 is modified to a value between 1 and 9, there is a better metric (less than 20) so that PE-4 will change the next hop via PE-6. First the metric of the LSP is modified to 9:

```

[/]
A:admin@PE-6# configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-6-PE-1-strict" {
        metric 9
      }
    }
  }
}

```

The GRT on PE-4 shows that the next hop to reach PE-1 has changed, from next hop PE-2 to next hop PE-6 (therefore, using the LSP shortcut), and the metric is 19 (10 to reach PE-6 plus metric 9 of the LSP PE-6-PE-1 shortcut):

```

[/]

```

```
A:admin@PE-4# show router route-table 192.0.2.1/32

=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]                Type  Proto  Age      Pref
  Next Hop[Interface Name]                Metric
-----
192.0.2.1/32                      Remote ISIS  00h00m29s 15
  192.168.46.2                      19
-----
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
=====
```

Because the metric of the LSP shortcut was modified to a value of 9, the GRT of PE-6 shows that the next hops of several routes have changed and are also using the shortcut LSP PE-6-PE-1 because the metric is better than the regular IS-IS metric. IGP shortcuts will not be used to resolve prefixes downstream of the LSP endpoint when the LSP metric is higher than the underlying IGP cumulative metric.

```
[/]
A:admin@PE-6# show router route-table

=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]                Type  Proto  Age      Pref
  Next Hop[Interface Name]                Metric
-----
192.0.2.1/32                      Remote ISIS  00h00m54s 15
  192.0.2.1 (tunneled:RSVP:2)          9
192.0.2.2/32                      Remote ISIS  00h00m54s 15
  192.0.2.1 (tunneled:RSVP:2)          19
192.0.2.3/32                      Remote ISIS  00h00m54s 15
  192.0.2.1 (tunneled:RSVP:2)          19
192.0.2.4/32                      Remote ISIS  00h02m06s 15
  192.168.46.1                      10
192.0.2.5/32                      Remote ISIS  00h02m06s 15
  192.168.56.1                      10
192.0.2.6/32                      Local  Local  01h08m12s 0
  system                              0
192.168.12.0/30                   Remote ISIS  00h00m54s 15
  192.0.2.1 (tunneled:RSVP:2)          19
192.168.13.0/30                   Remote ISIS  00h00m54s 15
  192.0.2.1 (tunneled:RSVP:2)          19
192.168.23.0/30                   Remote ISIS  00h00m54s 15
  192.0.2.1 (tunneled:RSVP:2)          29
192.168.24.0/30                   Remote ISIS  00h02m06s 15
  192.168.46.1                      20
192.168.35.0/30                   Remote ISIS  00h02m06s 15
  192.168.56.1                      20
192.168.45.0/30                   Remote ISIS  00h02m06s 15
  192.168.46.1                      20
192.168.46.0/30                   Local  Local  01h08m12s 0
  int-PE-6-PE-4                      0
192.168.56.0/30                   Local  Local  01h08m12s 0
  int-PE-6-PE-5                      0
-----
No. of Routes: 14
Flags: n = Number of times nexthop is repeated
```

B = BGP backup route available  
L = LFA nexthop available  
S = Sticky ECMP requested

=====

There are also cases where an LDP FEC can resolve to an RSVP LSP, if the user enables the LDP-over-RSVP feature or IGP shortcut feature when **prefer-tunnel-in-tunnel** is set to true in LDP and the endpoint of the RSVP LSP matches the FEC prefix. For those cases, the metric to the prefix is the sum of the RSVP LSP metric and the remaining IGP path cost.

[Table 2: RSVP LSP Role As Outcome of LSP Level and IGP Level Configuration Options](#) provides a summary of the outcome when configuring the forwarding adjacency, LDPoRSVP and RSVP shortcut options at both the IGP instance level and at the LSP level.

Table 2: RSVP LSP Role As Outcome of LSP Level and IGP Level Configuration Options

	IGP Instance Level Configurations					
LSP Level Configuration	advertise-tunnel-link enabled/ rsvp-shortcut enabled/ldp-over-rsvp enabled	advertise-tunnel-link enabled/ rsvp-shortcut enabled/ldp-over-rsvp disabled	advertise-tunnel-link enabled/ rsvp-shortcut disabled/ldp-over-rsvp disabled	advertise-tunnel-link disabled/ rsvp-shortcut disabled/ldp-over-rsvp disabled	advertise-tunnel-link disabled/ rsvp-shortcut enabled/ldp-over-rsvp enabled	advertise-tunnel-link disabled/ rsvp-shortcut disabled/ldp-over-rsvp enabled
igp-shortcut enabled/ldp-over-rsvp enabled	Forwarding Adjacency	Forwarding Adjacency	Forwarding Adjacency	None	IGP Shortcut	LDP-over-RSVP
igp-shortcut enabled/ldp-over-rsvp disabled	Forwarding Adjacency	Forwarding Adjacency	Forwarding Adjacency	None	IGP Shortcut	None
igp-shortcut disabled/ldp-over-rsvp enabled	None	None	None	None	None	LDP-over-RSVP
igp-shortcut disabled/ldp-over-rsvp disabled	None	None	None	None	None	None

### LSP Relative Metric

It is possible to use relative metrics for IGP shortcuts as per RFC 3906, *Calculating Interior Gateway Protocol (IGP) Routes Over Traffic Engineering Tunnels*, with the following command (e.g. on PE-6):

```
[/]
A:admin@PE-6# configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-6-PE-1-strict" {
        igp-shortcut ?
```

```

admin-state          - Administrative state of the LSP to be used as a shortcut while
resolving IGP routes

Choice: igp-choice
lfa-type             :- Specific RSVP LSP by IS-IS and OSPF routing protocols to assign as
shortcut or as forwarding adjacency
relative-metric      :- Shortest IGP cost between the endpoints of the LSP plus the configured
offset

[/]
A:admin@PE-6# configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-6-PE-1-strict" {
        igp-shortcut {
          lfa-type ?

lfa-type <keyword>
<keyword> - (lfa-protect|lfa-only)

        Specific RSVP LSP by IS-IS and OSPF routing protocols to assign as shortcut or as
        forwarding adjacency

[/]
A:admin@PE-6# configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-6-PE-1-strict" {
        igp-shortcut {
          relative-metric ?

relative-metric <number>
<number> - <-10..10>

        Shortest IGP cost between the endpoints of the LSP plus the configured offset

```

When this feature is enabled, IGP applies the shortest IGP cost between the endpoints of the LSP, plus the value of a configured offset when computing the cost of the prefix that is resolved to the LSP.

The offset value is optional and can have a value between -10 and 10, and defaults to zero (0). An offset value of zero (0) is used when the relative metric option is enabled without specifying the offset parameter value. The minimum net cost for the prefix is capped to the value of one (1) after applying the offset:

$$\text{Prefix cost} = \max(1, \text{IGP Cost} + \text{relative metric offset})$$

The **relative-metric** option is ignored when **advertise-tunnel-link** is enabled in IS-IS or OSPF. In that case, the IGP advertises the LSP as a P2P unnumbered link using the LSP operational metric.

The **relative-metric** option is mutually exclusive with the **lfa-protect** (Loop-Free Alternate (LFA)) or the **lfa-only** options. An LSP with **relative-metric** option enabled cannot be included in the LFA SPF and vice versa when RSVP shortcuts are enabled in the IGP (see chapter [LDP/IP FRR LFA for IGP Shortcut Using IS-IS/OSPF](#) for more information).

The offset can be used to enforce the preference of the shortcut path over the other paths for the prefix. Using an example, a new CSPF LSP with empty path and relative metric of -10 is created between PE-6 and PE-1. Whereas the operational or absolute metric is 30 (IGP cost and populated in the Tunnel Table Manager, TTM), the metric that the RTM shows is 20 after applying the offset:

```

[/]
A:admin@PE-6# configure {
  router "Base" {

```

```

mpls {
  admin-state enable
  lsp "LSP-PE-6-PE-1-loose" {
    to 192.0.2.1
    type p2p-rsvp
    primary "loose_path" { }
    path-computation-method local-cspf
    igp-shortcut {
      relative-metric -10
    }
  }
}

```

```

[/]
A:admin@PE-6# show router tunnel-table 192.0.2.1

```

=====  
IPv4 Tunnel Table (Router: Base)  
=====

Destination Color	Owner	Encap	TunnelId	Pref	Nexthop	Metric
192.0.2.1/32	rsvp	MPLS	1	7	192.168.46.1	30
192.0.2.1/32	rsvp	MPLS	2	7	192.168.56.1	16777215
192.0.2.1/32	ldp	MPLS	65546	9	192.168.46.1	30

-----  
Flags: B = BGP or MPLS backup hop available  
L = Loop-Free Alternate (LFA) hop available  
E = Inactive best-external BGP route  
k = RIB-API or Forwarding Policy backup hop  
=====

```

[/]
A:admin@PE-6# show router route-table 192.0.2.1

```

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

Dest Prefix[Flags] Next Hop[Interface Name]	Type	Proto	Age	Pref Metric
192.0.2.1/32 192.168.46.1	Remote	ISIS	00h03m11s	15 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/IP FRR LFA for IGP Shortcut Using IS-IS/OSPF

MPLS LDP/IP FRR LFA for IGP shortcuts allows for the use of RSVP LSP-based IGP shortcuts as Loop-Free Alternate (LFA) backups, this way expanding the coverage of the IP Fast-Reroute (FRR) and the LDP FRR capabilities for IS-IS and OSPF prefixes. For a detailed description about IP and LDP FRR, see chapter [MPLS LDP FRR using ISIS as IGP](#).

When an RSVP LSP is used as a shortcut by IS-IS or OSPF, it is included by the SPF as a P2P link and it can also be optionally advertised into the rest of the network by the IGP.

Two LSP-level configuration options are provided:

- The **lfa-protect** option includes the RSVP LSP in both the main SPF and the LFA SPFs. If the prefix primary next hop (NH) is tunneled, no LFA NH is computed. The protection in this case is provided by RSVP FRR. If the prefix primary NH is direct, then an LFA NH is computed. A direct LFA NH is preferred over a tunneled LFA NH. Within each LFA NH type, node protection is preferred over link protection. The configuration command is:

```
[/]
A:admin@a PE># configure {
  router "Base" {
    mpls {
      lsp "<LSP Name>" {
        igp-shortcut {
          lfa-type lfa-protect
        }
      }
    }
  }
}
```

- The **lfa-only** option includes the LSP in the LFA SPFs only so that the introduction of IGP shortcuts does not impact the main SPF decision. The prefix primary NH is always direct and the prefix LFA NH is computed. A direct LFA NH is preferred over a tunneled LFA NH. Within each LFA NH type, node protection is preferred over link protection. The configuration command is:

```
[/]
A:admin@a PE># configure {
  router "Base" {
    mpls {
      lsp "<LSP Name>" {
        igp-shortcut {
          lfa-type lfa-only
        }
      }
    }
  }
}
```

LDP/IP FRR is a local decision, so it can be enabled per node and there are no interoperability issues with other nodes. In the topology, PE-2 is provisioned with IS-IS LFA (OSPF configuration for the rest of this section is similar):

```
[/]
A:admin@PE-2# configure {
  router "Base" {
    isis 0 {
      loopfree-alternate { }
    }
  }
}
```

The second item to configure is whether LDP or IP FRR is provisioned. To configure IP FRR, the command is:

```
[/]
```

```
A:admin@PE-2# configure {
  routing-options {
    ip-fast-reroute true
  }
}
```

To configure LDP FRR, the following command is used:

```
[/]
A:admin@PE-2# configure {
  router "Base" {
    ldp {
      fast-reroute
    }
  }
}
```

Although not shown, it is recommended to enable IGP-LDP synchronization per interface to avoid possible traffic black-holes.

LFA is enabled in all routers of the topology. The following command shows the LFA coverage on PE-2 where four nodes out of five are protected (80%) and seven of the ten prefixes are protected (70%). IPv4 prefixes are protected (IPv6 is not configured). The following output shows L1 and L2 because this node is provisioned as an L1-L2 IS-IS router.

```
[/]
A:admin@PE-2# show router isis lfa-coverage

=====
Rtr Base ISIS Instance 0 LFA Coverage
=====
Topology          Level  Node      IPv4          IPv6
-----
IPV4 Unicast      L1     4/5(80%)   7/10(70%)    0/0(0%)
IPV6 Unicast      L1     0/0(0%)    0/0(0%)      0/0(0%)
IPV4 Multicast    L1     0/0(0%)    0/0(0%)      0/0(0%)
IPV6 Multicast    L1     0/0(0%)    0/0(0%)      0/0(0%)
IPV4 Unicast      L2     4/5(80%)   7/10(70%)    0/0(0%)
IPV6 Unicast      L2     0/0(0%)    0/0(0%)      0/0(0%)
IPV4 Multicast    L2     0/0(0%)    0/0(0%)      0/0(0%)
IPV6 Multicast    L2     0/0(0%)    0/0(0%)      0/0(0%)
=====
```

PE-2, PE-3, PE-4, and PE-5 share the same results, whereas only PE-1 and PE-6 have a coverage of 100% as shown in the following output.

```
[/]
A:admin@PE-1# show router isis lfa-coverage

=====
Rtr Base ISIS Instance 0 LFA Coverage
=====
Topology          Level  Node      IPv4          IPv6
-----
IPV4 Unicast      L1     5/5(100%)  11/11(100%)  0/0(0%)
IPV6 Unicast      L1     0/0(0%)    0/0(0%)      0/0(0%)
IPV4 Multicast    L1     0/0(0%)    0/0(0%)      0/0(0%)
IPV6 Multicast    L1     0/0(0%)    0/0(0%)      0/0(0%)
IPV4 Unicast      L2     5/5(100%)  11/11(100%)  0/0(0%)
IPV6 Unicast      L2     0/0(0%)    0/0(0%)      0/0(0%)
IPV4 Multicast    L2     0/0(0%)    0/0(0%)      0/0(0%)
=====
```



```
IPV6 Multicast L2 0/0(0%) 0/0(0%) 0/0(0%)
=====
```

Taking a deeper look into the IS-IS LFA on PE-2, it can be seen that the node which is not protected is PE-4 (system address 192.0.2.4, because it is the one missing):

```
[/]
A:admin@PE-2# show router route-table alternative | match LFA pre-lines 2
192.0.2.1/32 Remote ISIS 01h21m18s 15
    192.168.12.1 10
    192.168.23.2 (LFA) 20
192.0.2.3/32 Remote ISIS 01h20m52s 15
    192.168.23.2 10
    192.168.12.1 (LFA) 20
192.0.2.5/32 Remote ISIS 01h20m18s 15
    192.168.23.2 20
    192.168.24.2 (LFA) 20
192.0.2.6/32 Remote ISIS 01h19m56s 15
    192.168.24.2 20
    192.168.23.2 (LFA) 30
192.168.13.0/30 Remote ISIS 01h21m18s 15
    192.168.12.1 20
    192.168.23.2 (LFA) 30
192.168.35.0/30 Remote ISIS 01h20m52s 15
    192.168.23.2 20
    192.168.12.1 (LFA) 30
192.168.56.0/30 Remote ISIS 01h20m18s 15
    192.168.23.2 30
    192.168.24.2 (LFA) 30
Flags: n = Number of times nexthop is repeated
Backup = BGP backup route
LFA = Loop-Free Alternate nexthop
```

LFA is improved by taking advantage of RSVP shortcuts when it is properly provisioned. The reason why PE-4 cannot be protected with an LFA path is because the direct NH is using the direct link between PE-2 and PE-4 (the shortest IGP) and the intended LFA path through PE-3 is not valid (when LFA tries to find an alternate path via PE-3, the IGP cost from PE-3 to PE-4 is the same going via PE-5 as the path back via PE-2, invalidating that LFA calculation because there is a loop). This is normal because PE-2, PE-3, PE-4 and PE-5 are forming a ring. LFA coverage is increased by adding a link between PE-2 and PE-5, which can be done using a physical link or a virtual link with an RSVP shortcut. From the two possible options (**lfa-only** and **lfa-protect**), a new LSP "LSP-PE-2-PE-5" is configured with **igp-shortcut lfa-only**.

```
[/]
A:admin@PE-2# configure {
  router "Base" {
    mpls {
      admin-state enable
      path "path-to-PE-5" {
        admin-state enable
        hop 10 {
          ip-address 192.0.2.3
          type strict
        }
        hop 20 {
          ip-address 192.0.2.5
          type strict
        }
      }
    }
    lsp "LSP-PE-2-PE-5" {
      admin-state enable
      to 192.0.2.5
    }
  }
}
```



```

192.168.23.2 10
192.168.12.1 (LFA) 20
192.0.2.4/32 Remote ISIS 01h22m47s 15
192.168.24.2 10
192.0.2.5 (LFA) (tunneled:RSVP:1) 16777225
192.0.2.5/32 Remote ISIS 01h22m27s 15
192.168.23.2 20
192.168.24.2 (LFA) 20
192.0.2.6/32 Remote ISIS 01h22m05s 15
192.168.24.2 20
192.168.23.2 (LFA) 30
192.168.13.0/30 Remote ISIS 01h23m27s 15
192.168.12.1 20
192.168.23.2 (LFA) 30
192.168.35.0/30 Remote ISIS 01h23m01s 15
192.168.23.2 20
192.168.12.1 (LFA) 30
192.168.45.0/30 Remote ISIS 01h22m47s 15
192.168.24.2 20
192.0.2.5 (LFA) (tunneled:RSVP:1) 16777235
192.168.46.0/30 Remote ISIS 01h22m47s 15
192.168.24.2 20
192.0.2.5 (LFA) (tunneled:RSVP:1) 16777235
192.168.56.0/30 Remote ISIS 01h22m27s 15
192.168.23.2 30
192.168.24.2 (LFA) 30
Flags: n = Number of times nexthop is repeated
Backup = BGP backup route
LFA = Loop-Free Alternate nexthop

```

The tunnel table shows the RSVP LSP used as a shortcut and its operational metric.

```

[/]
A:admin@PE-2# show router tunnel-table 192.0.2.5

=====
IPv4 Tunnel Table (Router: Base)
=====
Destination      Owner      Encap TunnelId  Pref  Nexthop      Metric
  Color
-----
192.0.2.5/32      rsvp       MPLS 1          7     192.168.23.2 16777215
192.0.2.5/32 [L]  ldp        MPLS 65540         9     192.168.23.2 20
-----
Flags: B = BGP or MPLS backup hop available
      L = Loop-Free Alternate (LFA) hop available
      E = Inactive best-external BGP route
      k = RIB-API or Forwarding Policy backup hop
=====

```

If the LSP "LSP-PE-2-PE-5" is provisioned with **lfa-protect** instead of **lfa-only**, the result is that the LSP "LSP-PE-2-PE-5" is used by normal SPF to define the primary NH and it is not used by LFA SPF anymore.

```

[/]
A:admin@PE-2# configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-2-PE-5" {
        igp-shortcut {
          lfa-type lfa-protect
        }
      }
    }
  }
}

```

```
)
)
}
```

The coverage when lfa-protect is used also shows a 100% for nodes and 100% for prefixes, as follows.

```
[/]
A:admin@PE-2# show router isis lfa-coverage

=====
Rtr Base ISIS Instance 0 LFA Coverage
=====
Topology          Level   Node           IPv4           IPv6
-----
IPV4 Unicast      L1      5/5(100%)      9/9(100%)      0/0(0%)
IPV6 Unicast      L1      0/0(0%)        0/0(0%)        0/0(0%)
IPV4 Multicast    L1      0/0(0%)        0/0(0%)        0/0(0%)
IPV6 Multicast    L1      0/0(0%)        0/0(0%)        0/0(0%)
IPV4 Unicast      L2      5/5(100%)      9/9(100%)      0/0(0%)
IPV6 Unicast      L2      0/0(0%)        0/0(0%)        0/0(0%)
IPV4 Multicast    L2      0/0(0%)        0/0(0%)        0/0(0%)
IPV6 Multicast    L2      0/0(0%)        0/0(0%)        0/0(0%)
=====
```

In this case, the GRT looks as follows, the main difference being that now PE-5 (192.0.2.5) has a direct shortcut from PE-2:

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

=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]          Type   Proto   Age           Pref
  Next Hop[Interface Name]  Metric
  Alt-NextHop               Alt-
                             Metric
-----
192.0.2.1/32                Remote ISIS   01h24m50s  15
  192.168.12.1              10
  192.168.23.2 (LFA)        20
192.0.2.2/32                Local  Local   01h24m51s   0
  system                    0
192.0.2.3/32                Remote ISIS   01h24m24s  15
  192.168.23.2              10
  192.168.12.1 (LFA)        20
192.0.2.4/32                Remote ISIS   01h24m10s  15
  192.168.24.2              10
  192.0.2.5 (LFA) (tunneled:RSVP:1) 16777225
192.0.2.5/32                Remote ISIS 00h00m45s 15
  192.0.2.5 (tunneled:RSVP:1) 16777215
192.0.2.6/32                Remote ISIS   01h23m28s  15
  192.168.24.2              20
  192.168.23.2 (LFA)        30
192.168.12.0/30             Local  Local   01h24m51s   0
  int-PE-2-PE-1            0
192.168.13.0/30             Remote ISIS   01h24m50s  15
  192.168.12.1              20
  192.168.23.2 (LFA)        30
192.168.23.0/30             Local  Local   01h24m51s   0
  int-PE-2-PE-3            0
192.168.24.0/30             Local  Local   01h24m51s   0
```

```

int-PE-2-PE-4
192.168.35.0/30 Remote ISIS 01h24m24s 15
192.168.23.2 20
192.168.12.1 (LFA) 30
192.168.45.0/30 Remote ISIS 01h24m10s 15
192.168.24.2 20
192.0.2.5 (LFA) (tunneled:RSVP:1) 16777235
192.168.46.0/30 Remote ISIS 01h24m10s 15
192.168.24.2 20
192.0.2.5 (LFA) (tunneled:RSVP:1) 16777235
192.168.56.0/30 Remote ISIS 00h00m45s 15
192.168.24.2 30
192.168.23.2 (LFA) 40
-----
No. of Routes: 14
Flags: n = Number of times nexthop is repeated
Backup = BGP backup route
LFA = Loop-Free Alternate nexthop
S = Sticky ECMP requested
=====

```

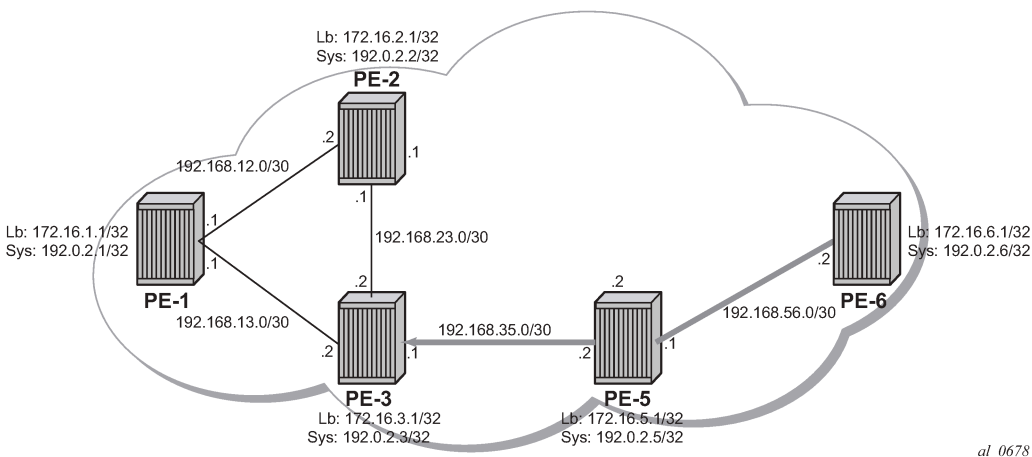
### Rules Determining the Installation of Shortcuts into the RTM

Although it was already mentioned in the RSVP-TE LSP shortcut for IGP route resolution section, the rules determining how shortcuts are installed into the RTM are (sorted by higher priority):

- RSVP shortcut.
- LDP shortcut.
- IGP route with regular IP next hop.
- The implementation is compliant with RFC 3906.

To check the rules, the network configuration is iLDP in all interfaces with LDP shortcuts enabled, there is also an RSVP LSP from PE-6 to PE-3 available but RSVP shortcuts are disabled. The topology is shown in [Figure 39: Network Topology to Verify Installation of Shortcuts into the RTM](#).

Figure 39: Network Topology to Verify Installation of Shortcuts into the RTM



The following RSVP LSP is needed between PE-6 and PE-3.

```
[/]
A:admin@PE-6# configure {
  router "Base" {
    ldp {
      ldp-shortcut {
        ipv4 true
      }
    }
  }
}
```

```
[/]
A:admin@PE-6# configure {
  router "Base" {
    isis 0 {
      delete loopfree-alternate
      igp-shortcut {
        tunnel-next-hop {
          family ipv4 {
            resolution-filter {
              rsvp false
            }
            resolution none
          }
        }
        admin-state disable
      }
    }
  }
}
```

```
[/]
A:admin@PE-6# configure {
  router "Base" {
    mpls {
      path "loose_path" {
        admin-state enable
      }
      lsp "LSP-PE-6-PE-3" {
        admin-state enable
        to 192.0.2.3
        type p2p-rsvp
        primary "loose_path" { }
        path-computation-method local-cspf
      }
    }
  }
}
```

The routes in the routing table on PE-6 are the following:

```
[/]
A:admin@PE-6# show router route-table

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

```

-----
192.0.2.1/32                               Remote LDP      00h00m00s 9
      192.168.56.1 (tunneled)              30
192.0.2.2/32                               Remote LDP      00h00m00s 9
      192.168.56.1 (tunneled)              30
192.0.2.3/32                               Remote LDP      00h00m00s 9
      192.168.56.1 (tunneled)              20
192.0.2.5/32                               Remote LDP      00h00m00s 9
      192.168.56.1 (tunneled)              10
192.0.2.6/32                               Local  Local      01h28m21s 0
      system                                0
192.168.12.0/30                            Remote ISIS     00h02m08s 15
      192.168.56.1                          40
192.168.13.0/30                            Remote ISIS     00h19m24s 15
      192.168.56.1                          30
192.168.23.0/30                            Remote ISIS     00h02m08s 15
      192.168.56.1                          30
192.168.35.0/30                            Remote ISIS     00h22m14s 15
      192.168.56.1                          20
192.168.56.0/30                            Local  Local      01h28m21s 0
      int-PE-6-PE-5                          0
-----
No. of Routes: 10
Flags: n = Number of times nexthop is repeated
      B = BGP backup route available
      L = LFA nexthop available
      S = Sticky ECMP requested
=====

```

The tunnel table shows the LSPs available for the shortcuts, and therefore, these are used in the GRT for LDP (but not for RSVP):

```

[/]
A:admin@PE-6# show router tunnel-table

=====
IPv4 Tunnel Table (Router: Base)
=====
Destination      Owner      Encap TunnelId  Pref  Nexthop      Metric
  Color
-----
192.0.2.1/32     ldp        MPLS  65547    9     192.168.56.1  30
192.0.2.2/32     ldp        MPLS  65544    9     192.168.56.1  30
192.0.2.3/32     rsvp       MPLS   4        7     192.168.56.1  20
192.0.2.3/32     ldp        MPLS  65545    9     192.168.56.1  20
192.0.2.5/32     ldp        MPLS  65540    9     192.168.56.1  10
-----
Flags: B = BGP or MPLS backup hop available
      L = Loop-Free Alternate (LFA) hop available
      E = Inactive best-external BGP route
      k = RIB-API or Forwarding Policy backup hop
=====

```

So far, LDP shortcuts are preferred over the IGP next hops for the system addresses (router ID). After enabling RSVP shortcuts in the IS-IS context, the changes in the GRT are:

```

[/]
A:admin@PE-6# configure {
  router "Base" {
    isis 0 {
      igp-shortcut {
        admin-state enable
      }
    }
  }
}

```





```

=====
Dest Prefix[Flags]                                Type  Proto  Age      Pref
  Next Hop[Interface Name]                        Metric
-----
192.0.2.3/32                                       Remote ISIS  00h00m53s 15
  192.0.2.3 (tunneled:RSVP:4)                       100
-----
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
=====

```

- SPF also picks the RSVP LSP shortcut if both the LSP path and the IGP path to the prefix are via the tail-end of the LSP. This is regardless of the path cost compared to the IGP next hop. When paths over multiple RSVP shortcuts have the same lowest cost, if ECMP is enabled on the system, the LSP with the lowest tunnel ID is chosen. In this example, 192.168.13.0 and 192.168.23.0 are using the shortcut but 192.168.12.0 is not.

```

[/]
A:admin@PE-6# show router route-table

=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]                                Type  Proto  Age      Pref
  Next Hop[Interface Name]                        Metric
-----
192.0.2.1/32                                       Remote LDP   00h00m53s 9
  192.168.56.1 (tunneled)                          30
192.0.2.2/32                                       Remote LDP   00h00m53s 9
  192.168.56.1 (tunneled)                          30
192.0.2.3/32                                       Remote ISIS  00h00m53s 15
  192.0.2.3 (tunneled:RSVP:4)                       100
192.0.2.5/32                                       Remote LDP   00h01m25s 9
  192.168.56.1 (tunneled)                          10
192.0.2.6/32                                       Local  Local    01h29m45s 0
  system
192.168.12.0/30                                     Remote ISIS  00h00m53s 15
  192.168.56.1                                     40
192.168.13.0/30                                     Remote ISIS  00h00m53s 15
  192.0.2.3 (tunneled:RSVP:4)                       110
192.168.23.0/30                                     Remote ISIS  00h00m53s 15
  192.0.2.3 (tunneled:RSVP:4)                       110
192.168.35.0/30                                     Remote ISIS  00h23m39s 15
  192.168.56.1                                     20
192.168.56.0/30                                     Local  Local    01h29m45s 0
  int-PE-6-PE-5                                     0
-----
No. of Routes: 10
Flags: n = Number of times nexthop is repeated
      B = BGP backup route available
      L = LFA nexthop available
      S = Sticky ECMP requested
=====

```

## LDP/RSVP LSP Shortcut for BGP NH Resolution

Using LDP/RSVP LSP shortcuts for resolving BGP next hops allows IPv4 packet forwarding to routes resolved via a BGP next hop using an LDP/RSVP LSP instead of using a regular IP next hop. In the network topology of [Figure 36: Example Topology](#), both PE-3 and PE-6 have a single peer configured, initially without any shortcuts enabled under the **bgp** context. Also, one static route is configured in PE-3 and PE-6 and that is redistributed into BGP. The relevant configuration on PE-3 is the following:

```
[/]
A:admin@PE-3# configure {
  router "Base" {
    interface "int-static-route" {
      ipv4 {
        primary {
          address 172.16.33.1
          prefix-length 30
        }
      }
      port 1/1/4:33
    }
    autonomous-system 65536
    static-routes {
      route 10.10.10.0/24 route-type unicast {
        next-hop "172.16.33.2" {
          admin-state enable
        }
      }
    }
  }
}

[/]
A:admin@PE-3# configure {
  policy-options {
    policy-statement "static-routes-policy" {
      description "export static-routes for I-BGP"
      entry 10 {
        from {
          protocol {
            name [static]
          }
        }
        to {
          protocol {
            name [bgp]
          }
        }
        action {
          action-type accept
          next-hop self
        }
      }
      default-action {
        action-type reject
      }
    }
  }
}

[/]
A:admin@PE-3# configure {
```

```

router "Base" {
  bgp {
    export {
      policy ["static-routes-policy"]
    }
    group "ibgp" {
      peer-as 65536
    }
    neighbor 192.0.2.6 {
      group "ibgp"
    }
  }
}

```

Checking the static route received on PE-6 via BGP, the next hop is the PE-3 system address:

```

[/]
A:admin@PE-6# show router bgp routes 10.10.10.0/24 detail
=====
BGP Router ID:192.0.2.6      AS:65536      Local AS:65536
=====
Legend -
Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                l - leaked, x - stale, > - best, b - backup, p - purge
Origin codes  : i - IGP, e - EGP, ? - incomplete
=====
BGP IPv4 Routes
=====
Original Attributes

Network       : 10.10.10.0/24
NextHop       : 192.0.2.3
Path Id       : None
From          : 192.0.2.3
Res. Protocol : ISIS                Res. Metric   : 20
Res. NextHop  : 192.168.56.1
Local Pref.   : 100                  Interface Name : int-PE-6-PE-5
Aggregator AS : None                 Aggregator    : None
Atomic Aggr.  : Not Atomic           MED           : None
AIGP Metric   : None                 IGP Cost      : 20
Connector     : None
Community     : No Community Members
Cluster       : No Cluster Members
Originator Id : None                 Peer Router Id : 192.0.2.3
Fwd Class     : None                 Priority       : None
Flags         : Used Valid Best Incomplete In-RTM
Route Source  : Internal
AS-Path       : No As-Path
Route Tag     : 0
Neighbor-AS   : n/a
Orig Validation: NotFound
Source Class  : 0                     Dest Class    : 0
Add Paths Send : Default
RIB Priority   : Normal
Last Modified  : 00h01m32s

Modified Attributes

---snip---
-----

```

```
Routes : 1
=====
```

Three of the BGP peering configuration possibilities are LDP, RSVP, or BGP. The other resolution filter options are related to segment routing and are beyond the scope of this chapter. In case both LDP and RSVP are included in the filter, RSVP is preferred. Disabling the IGP is also allowed (meaning that unless there is a shortcut, the BGP peering will not fall back to IGP):

```
[/]
A:admin@PE-6# configure {
  router "Base" {
    bgp {
      next-hop-resolution {
        shortcut-tunnel {
          family ipv4 {
            resolution ?

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

Resolution mode for binding BGP routes to tunnel types

[/]
A:admin@PE-6# configure {
  router "Base" {
    bgp {
      next-hop-resolution {
        shortcut-tunnel {
          family ipv4 {
            resolution-filter ?

resolution-filter

bgp           - Use BGP tunneling for next-hop resolution
ldp           - Use LDP tunneling for next-hop resolution
mpls-fwd-policy - Use MPLS forwarding policy for next-hop resolution
rib-api       - Use RIB API gRPC service for next-hop resolution
rsvp          - Use RSVP tunneling for next-hop resolution
sr-isis       - Use IS-IS SR tunneling for next-hop resolution
sr-ospf       - Use OSPF SR tunneling for next-hop resolution
sr-ospf3      - Use OSPFv3 SR tunneling for next-hop resolution
sr-policy     - Use SR policies for next-hop resolution
sr-te        - Use SR-TE tunneling for next-hop resolution
```

When enabling LDP shortcuts on PE-6, the output changes showing the detail of the received BGP route indicating that the next hop is resolved using LDP:

```
configure {
  router "Base" {
    bgp {
      next-hop-resolution {
        shortcut-tunnel {
          family ipv4 {
            resolution-filter {
              ldp true
            }
            resolution filter
          }
        }
      }
    }
  }
}
```

```

}
}
}

[/]
A:admin@PE-6# show router bgp routes 10.10.10.0/24 detail
=====
BGP Router ID:192.0.2.6      AS:65536      Local AS:65536
=====
Legend -
Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                l - leaked, x - stale, > - best, b - backup, p - purge
Origin codes  : i - IGP, e - EGP, ? - incomplete
=====
BGP IPv4 Routes
=====
Original Attributes

Network       : 10.10.10.0/24
Nexthop       : 192.0.2.3
Path Id       : None
From          : 192.0.2.3
Res. Protocol : LDP                      Res. Metric   : 20
Res. Nexthop  : 192.0.2.3 (LDP)
Local Pref.   : 100                      Interface Name : NotAvailable
Aggregator AS : None                      Aggregator    : None
Atomic Aggr.  : Not Atomic                MED           : None
AIGP Metric   : None                      IGP Cost      : 20
Connector     : None
Community     : No Community Members
Cluster       : No Cluster Members
Originator Id : None                      Peer Router Id : 192.0.2.3
Fwd Class     : None                      Priority       : None
Flags         : Used Valid Best Incomplete In-RTM
Route Source  : Internal
AS-Path       : No As-Path
Route Tag     : 0
Neighbor-AS   : n/a
Orig Validation: NotFound
Source Class  : 0                          Dest Class    : 0
Add Paths Send : Default
RIB Priority   : Normal
Last Modified : 00h02m32s

Modified Attributes

---snip---

-----
Routes : 1
=====

```

The GRT output command also shows that the route is reachable using LDP (indicated as tunneled):

```

[/]
A:admin@PE-6# show router route-table next-hop-type tunneled
=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]                                Type  Proto  Age      Pref
Next Hop[Interface Name]                          Metric

```

```
-----
10.10.10.0/24                               Remote BGP           00h00m30s 170
      192.0.2.3 (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
=====
```

The previously created LSP LSP-PE-6-PE-3 is up and running:

```
[/]
A:admin@PE-6# show router mpls lsp "LSP-PE-6-PE-3" path detail

=====
MPLS LSP LSP-PE-6-PE-3 Path (Detail)
=====
Legend :
  @ - Detour Available           # - Detour In Use
  b - Bandwidth Protected       n - Node Protected
  s - Soft Preemption
  S - Strict                     L - Loose
  A - ABR                       + - Inherited
=====

LSP LSP-PE-6-PE-3
Path loose_path
-----
LSP Name      : LSP-PE-6-PE-3
From          : 192.0.2.6
To           : 192.0.2.3
Admin State : Up
Oper State  : Up
Path Name    : loose_path
Path LSP ID  : 9216
Path Admin   : Up
Out Interface : 1/1/1
Path Up Time : 0d 00:13:27
Retry Limit  : 0
Retry Attempt : 0
Path Type    : Primary
Path Oper    : Up
Out Label    : 524283
Path Down Time : 0d 00:00:00
Retry Timer  : 30 sec
Next Retry In : 0 sec

---snip---

Adspec       : Disabled
PathCompMethod : local-cspf
MetricType   : igp
Least Fill   : Disabled
FRR          : Disabled
Propagate Adm Grp: Disabled
Inter-area   : False
Oper Adspec  : Disabled
OperPathCompMethod: local-cspf
Oper MetricType : igp
Oper LeastFill : Disabled
Oper FRR      : Disabled
Oper Prop Adm Grp : Disabled

---snip---

Adaptive     : Enabled
Preference   : n/a
Path Trans   : 1
Failure Code : noError
Failure Node : n/a
Explicit Hops :
  No Hops Specified
Actual Hops  :
  192.168.56.2(192.0.2.6)
Oper Metric  : 100
CSPF Queries : 1
Record Label : N/A
```

```

-> 192.168.56.1(192.0.2.5)          Record Label      : 524283
-> 192.168.35.1(192.0.2.3)       Record Label      : 524284
Computed Hops      :
  192.168.56.2(S)
-> 192.168.56.1(S)
-> 192.168.35.1(S)
Resignal Eligible: False
Last Resignal     : n/a                CSPF Metric       : 20
=====

```

After adding **resolution-filter rsvp** to the shortcut-tunnel configuration in the **bgp** context, the output shows that the BGP peer is reachable using an RSVP LSP (switched from LDP to RSVP because RSVP is preferred):

```

[/]
A:admin@PE-6# configure {
  router "Base" {
    bgp {
      next-hop-resolution {
        shortcut-tunnel {
          family ipv4 {
            resolution-filter {
              ldp true
              rsvp true
            }
            resolution filter
          }
        }
      }
    }
  }
}

[/]
A:admin@PE-6# show router bgp routes 10.10.10.0/24 detail
=====
BGP Router ID:192.0.2.6      AS:65536      Local AS:65536
=====
Legend -
Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                l - leaked, x - stale, > - best, b - backup, p - purge
Origin codes  : i - IGP, e - EGP, ? - incomplete
=====
BGP IPv4 Routes
=====
Original Attributes

Network       : 10.10.10.0/24
Nexthop      : 192.0.2.3
Path Id      : None
From         : 192.0.2.3
Res. Protocol : RSVP                      Res. Metric   : 100
Res. Nexthop : 192.0.2.3 (RSVP LSP: 4)
Local Pref.  : 100
Aggregator AS : None                      Interface Name : NotAvailable
Atomic Aggr. : Not Atomic                 Aggregator    : None
AIGP Metric  : None                       MED           : None
Connector    : None                       IGP Cost      : 100
Community    : No Community Members
Cluster      : No Cluster Members
Originator Id : None                      Peer Router Id : 192.0.2.3
Fwd Class    : None                       Priority       : None

```

```

Flags      : Used Valid Best Incomplete In-RTM
Route Source : Internal
AS-Path    : No As-Path
Route Tag   : 0
Neighbor-AS : n/a
Orig Validation: NotFound
Source Class : 0                      Dest Class : 0
Add Paths Send : Default
RIB Priority : Normal
Last Modified : 00h03m59s

Modified Attributes

---snip---

-----
Routes : 1
=====

```

The GRT output command also shows that the route is reachable using RSVP (indicated as tunneled:RSVP:4):

```

[/]
A:admin@PE-6# show router route-table next-hop-type tunneled

=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]                Type   Proto   Age           Pref
  Next Hop[Interface Name]                Metric
-----
10.10.10.0/24                    Remote BGP     00h00m33s  170
  192.0.2.3 (tunneled:RSVP:4)                100
-----
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
=====

```

If the RSVP LSP is **disabled**, the system reverts back to the LDP LSP:

```

configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-6-PE-3" {
        admin-state disable
      }
    }
  }
}

```

```

[/]
A:admin@PE-6# show router bgp routes 10.10.10.0/24 detail

=====
BGP Router ID:192.0.2.6      AS:65536      Local AS:65536
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid

```



```

Origin codes      l - leaked, x - stale, > - best, b - backup, p - purge
                  i - IGP, e - EGP, ? - incomplete

=====
BGP IPv4 Routes
=====
Original Attributes

Network          : 10.10.10.0/24
Nextthop         : 192.0.2.3
Path Id          : None
From             : 192.0.2.3
Res. Protocol    : LDP                      Res. Metric    : 20
Res. Nextthop    : 192.0.2.3 (LDP)
Local Pref.     : 100                      Interface Name : NotAvailable
Aggregator AS   : None                    Aggregator     : None
Atomic Aggr.    : Not Atomic              MED            : None
AIGP Metric     : None                    IGP Cost       : 20
Connector       : None
Community       : No Community Members
Cluster         : No Cluster Members
Originator Id   : None                    Peer Router Id  : 192.0.2.3
Fwd Class       : None                    Priority        : None
Flags           : Used Valid Best Incomplete In-RTM
Route Source    : Internal
AS-Path         : No As-Path
Route Tag       : 0
Neighbor-AS     : n/a
Orig Validation : NotFound
Source Class    : 0                      Dest Class     : 0
Add Paths Send  : Default
RIB Priority    : Normal
Last Modified   : 00h04m21s

Modified Attributes

---snip---

-----
Routes : 1
=====

```

When the shortcut tunnel with **resolution-filter rsvp** is enabled at the BGP level, all RSVP LSPs originating on this node are eligible to be used by default as long as the destination address of the LSP corresponds to that of the BGP next hop for that prefix. It is also possible to exclude a specific RSVP LSP from BGP next hop resolution, similar to the exclusion of a specific RSVP LSP being used as a shortcut for resolving IGP routes. In this example, if the RSVP LSP LSP-PE-6-PE-3 is excluded to be eligible for BGP next hop resolution, it reverts back to LDP.

```

[/]
A:admin@PE-6# configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-6-PE-3" {
        admin-state enable
        delete bgp-shortcut
      }
    }
  }
}

```

```

}

[/]
A:admin@PE-6# show router route-table 10.10.10.0

=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]                               Type   Proto   Age           Pref
  Next Hop[Interface Name]                       Metric
-----
10.10.10.0/24                                     Remote BGP      00h00m29s    170
  192.0.2.3 (tunneled:RSVP:4)                    100
-----
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
=====

```

If the configuration is using **disallow-igp**, and neither LDP nor RSVP LSPs are available, the remote route received via BGP is removed from the GRT although the BGP peer session remains up. A field in the detailed show BGP route output indicates that the next hop is "Unresolved":

```

[/]
A:admin@PE-6# configure {
  router "Base" {
    bgp {
      next-hop-resolution {
        shortcut-tunnel {
          family ipv4 {
            disallow-igp true
          }
        }
      }
    }
  }
}

```

```

[/]
A:admin@PE-6# configure {
  router "Base" {
    ldp {
      admin-state disable
    }
  }
}

```

```

[/]
A:admin@PE-6# show router bgp routes 10.10.10.0/24 detail

=====
BGP Router ID:192.0.2.6      AS:65536      Local AS:65536
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
              l - leaked, x - stale, > - best, b - backup, p - purge
Origin codes : i - IGP, e - EGP, ? - incomplete
=====

```

```

BGP IPv4 Routes
=====
Original Attributes

Network       : 10.10.10.0/24
Nexthop      : 192.0.2.3
Path Id      : None
From         : 192.0.2.3
Res. Protocol : INVALID           Res. Metric   : 0
Res. Nexthop : Unresolved
Local Pref.  : 100                    Interface Name : NotAvailable
Aggregator AS : None                  Aggregator    : None
Atomic Aggr. : Not Atomic             MED           : None
AIGP Metric  : None                   IGP Cost      : 0
Connector    : None
Community    : No Community Members
Cluster      : No Cluster Members
Originator Id : None                  Peer Router Id : 192.0.2.3
Fwd Class    : None                   Priority       : None
Flags       : Invalid Incomplete Nexthop-Unresolved
Route Source : Internal
AS-Path      : No As-Path
Route Tag    : 0
Neighbor-AS  : n/a
Orig Validation: NotFound
Source Class : 0                       Dest Class    : 0
Add Paths Send : Default
RIB Priority  : Normal
Last Modified : 00h10m33s

Modified Attributes

---snip---

-----
Routes : 1
=====
  
```

Because the route is unresolved, it does not appear in the GRT:

```

[/]
A:admin@PE-6# show router route-table 10.10.10.0

=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]                Type  Proto  Age      Pref
Next Hop[Interface Name]          Metric
-----
No. of Routes: 0
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
=====
  
```

## MPLS/GRE Shortcut for BGP NH Resolution within a VRF

Using RSVP/LDP or GRE shortcuts for resolving BGP next hops within a Virtual Private Routed Network (VPRN), also known as auto-bind-tunnel, allows a VPRN service to automatically resolve the BGP next hop for VPRN routes to an MPLS LSP or a GRE tunnel. Three possible mechanisms exist to provide transport tunnels for forwarding traffic between PE routers within an RFC 4364, *BGP/MPLS IP Virtual Private Networks (VPNs)*, network:

- RSVP-TE protocol to create tunnel LSPs between PE routers.
- LDP protocol to create tunnel LSPs between PE routers.
- GRE tunnels between PE routers.

These transport tunneling mechanisms provide the flexibility to use dynamically created LSPs where the service tunnels are automatically bound (the **auto-bind-tunnel** feature), and the ability to provide certain VPN services with their own transport tunnels by explicitly binding SDPs if desired. All services using the auto-bind-tunnel feature use the same set of LSPs, which does not allow for alternate tunneling mechanisms (like GRE) or the ability to craft sets of LSPs with bandwidth reservations for specific customers, as is available with explicit SDPs for the service.

The auto-bind-tunnel configuration is as follows:

```
[/]
A:admin@PE-2# configure {
  service {
    vprn "NON_EXISTING_VPRN" {
      bgp-ipvpn {
        mpls {
          auto-bind-tunnel
          ?

allow-flex-algo-      - Enable flexible algorithm fallback
  fallback
apply-groups         - Apply a configuration group at this level
apply-groups-exclude - Exclude a configuration group at this level
ecmp                 - Maximum ECMP routes allowed
enforce-strict-     - Allow enforcement of strict tunnel tagging
  tunnel-tagging
resolution           - Resolution method for tunnel selection
resolution-filter    + Enter the resolution-filter context
weighted-ecmp       - Allow weighted load-balancing

[/]
A:admin@PE-2# configure {
  service {
    vprn "NON_EXISTING_VPRN" {
      bgp-ipvpn {
        mpls {
          auto-bind-tunnel {
            resolution ?

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

Resolution method for tunnel selection

[/]
A:admin@PE-2# configure {
```

```

service {
  vprn "NON_EXISTING_VPRN" {
    bgp-ipvpn {
      mpls {
        auto-bind-tunnel {
          resolution-filter ?
        }
      }
    }
  }
}

resolution-filter

bgp          - Use BGP tunneling for next-hop resolution
gre          - Use GRE tunneling for next-hop resolution
ldp          - Use LDP tunneling for next-hop resolution
mpls-fw-d-policy - Use MPLS forwarding policy for next-hop resolution
rib-api      - Use RIB API gRPC service for next-hop resolution
rsvp        - Use RSVP tunneling for next-hop resolution
sr-isis     - Use IS-IS SR tunneling for next-hop resolution
sr-ospf     - Use OSPF SR tunneling for next-hop resolution
sr-ospf3    - Use OSPFv3 SR tunneling for next-hop resolution
sr-policy   - Use SR policies for next-hop resolution
sr-te       - Use SR-TE tunneling for next-hop resolution
udp         - Use MPLS over UDP tunneling for next-hop resolution

```

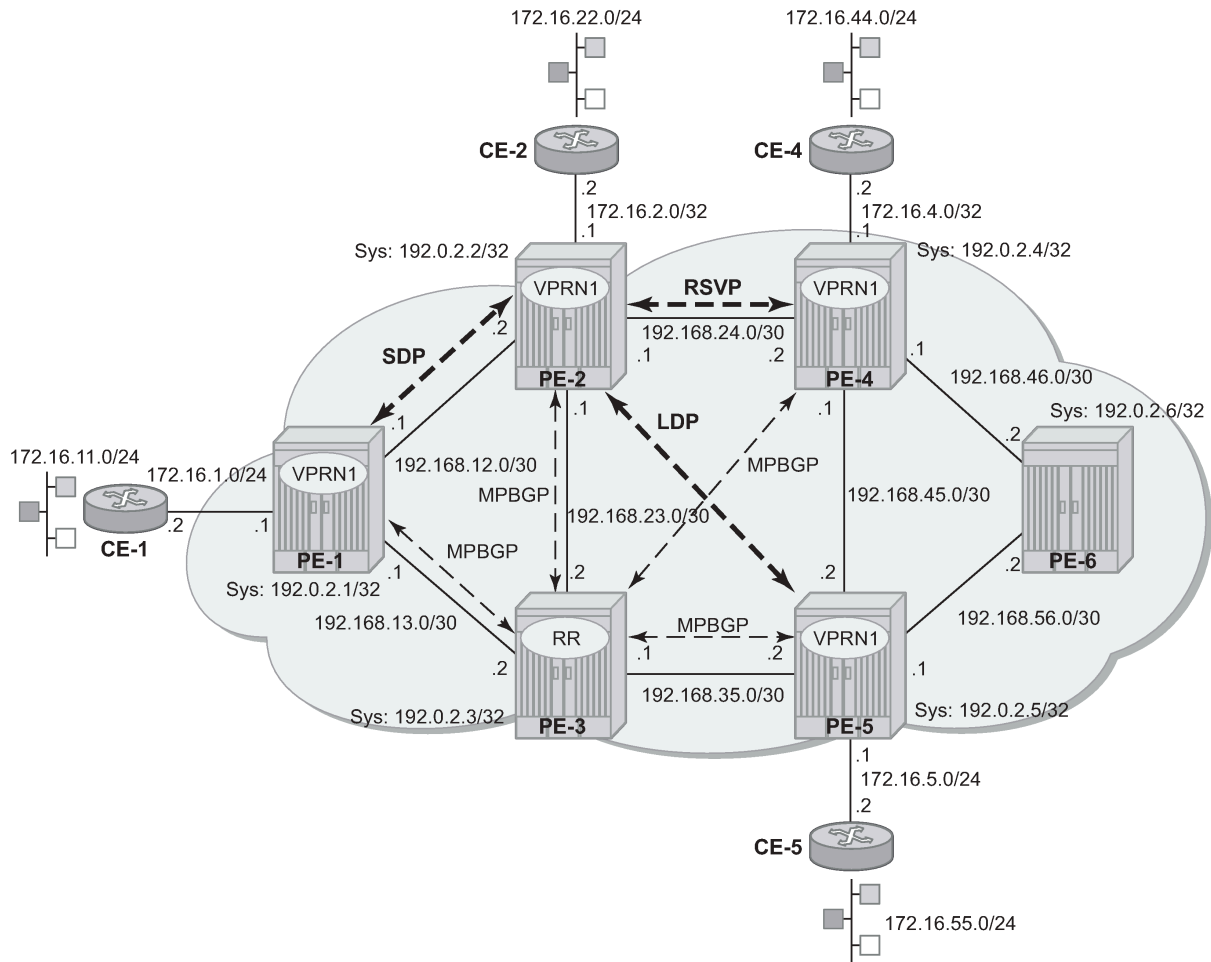
Parameter descriptions:

- **ldp** — Specifies LDP-based LSPs should be used to resolve the BGP next hop for VPRN routes in an associated VPRN instance.
- **gre** — Specifies GRE-based tunnels to be used to resolve the BGP next hop for VPRN routes in an associated VPRN instance. GRE is out of the scope regarding shortcuts, refer to SR OS documentation for further details.
- **rsvp** — Specifies RSVP-TE LSPs should be used to resolve the BGP next hop for VPRN routes in an associated VPRN instance.
- the remaining parameters are beyond the scope of this chapter.

In all cases, if an explicit spoke-SDP is specified in the VPRN, it is always preferred over automatically selected tunnels (even if the SDP is down, the route becomes inactive; there is no fallback to the automatic selection).

The network is configured according to the topology shown in [Figure 40: Shortcuts Within a VRF Topology Network](#). Four PEs (PE-1, PE-2, PE-4, and PE-5) are connected forming a meshed IP-VPN (named VPRN\_1), using a route reflector on PE-3 for MP-BGP peering. All PEs have LDP tunnels enabled so at a minimum all can establish LDP shortcut tunnels to the others. In order to have not only LDP but also RSVP-TE LSPs and static SDPs (using an RSVP LSP) in the network, a mix of tunneling methods is configured. For brevity, the configuration of PE-2 only is given, providing the details about the shortcuts created by auto-bind-tunnel. PE-2 has a static SDP (RSVP-based) with PE-1, an RSVP LSP with PE-4, and an LDP LSP with PE-5. Every PE has a CE connected, so each PE has an interface connected to the CE as well as a static route to a CE LAN (although redistribution routing policies are needed, they are not shown for brevity).

Figure 40: Shortcuts Within a VRF Topology Network



OSSG627

On PE-2, VPRN1 is configured as follows:

```
configure {
  service {
    sdp 1 {
      admin-state enable
      far-end {
        ip-address 192.0.2.1
      }
      delivery-type mpls
      lsp "LSP-PE-2-PE-1"
    }
  }
  vprn "VPRN_1" {
    admin-state enable
    service-id 1
    customer "1"
    bgp-ipvpn {
      mpls {
        admin-state enable
        route-distinguisher "65536:1"
        vrf-import {
          policy ["VPN1-import"]
        }
      }
    }
  }
}
```



```

192.0.2.5 (tunneled) 20
-----
No. of Routes: 8
Flags: n = Number of times nexthop is repeated
      B = BGP backup route available
      L = LFA nexthop available
      S = Sticky ECMP requested
=====

```

As can be seen, there are eight routes because every PE has two routes (one direct PE-CE interface and one static route), so six routes are received from other PEs via MP-BGP. The VPRN\_1 routing table can be understood by looking at the tunnel table (active LSPs for remote system IDs):

```

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

=====
IPv4 Tunnel Table (Router: Base)
=====
Destination      Owner      Encap TunnelId  Pref  Nexthop      Metric
  Color
-----
192.0.2.1/32     sdp        MPLS    1         5     192.0.2.1      0
192.0.2.1/32     rsvp       MPLS    2         7     192.168.12.1  10
192.0.2.1/32     ldp        MPLS   65537      9     192.168.12.1  10
192.0.2.3/32     ldp        MPLS   65538      9     192.168.23.2  10
192.0.2.4/32     rsvp       MPLS    3         7     192.168.24.2  10
192.0.2.4/32     rsvp       MPLS    4         7     192.168.24.2 16777215
192.0.2.4/32     ldp        MPLS   65545      9     192.168.24.2  10
192.0.2.5/32     ldp        MPLS   65542      9     192.168.23.2  20
192.0.2.6/32     ldp        MPLS   65546      9     192.168.24.2  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
=====

```

The tunnel table shows one entry per LSP per remote PE. The following tunnel selection rules apply:

- SDP has the lowest (best) preference, followed by RSVP and then by LDP.
- If the preference is the same, the lowest metric is selected (ECMP is possible with LDP).

PE-2 has three possibilities to reach PE-1 (192.0.2.1): an SDP tunnel ID 1 with preference 5, an RSVP tunnel ID 2 with preference 7, and an LDP LSP with preference 9. Because SDP tunnel ID 1 has the lowest preference, it is the chosen option. PE-2 has three possibilities to reach PE-4 (192.0.2.4): an RSVP tunnel ID 3 with preference 7 and metric 10, an RSVP tunnel ID 4 with preference 7 and metric 16777215, and an LDP LSP with preference 9; so RSVP tunnel ID 3 is selected. PE-2 only has one option to reach PE-5 and PE-6 (192.0.2.5 and 192.0.2.6) using an LDP LSP.

The following FIB for router VPRN\_1 on PE-2 provides more detailed information on the tunneling:

```

[/]
A:admin@PE-2# show router 1 fib 1

=====
FIB Display
=====
Prefix [Flags]      Protocol
  NextHop
-----

```



```

172.16.1.0/24                                BGP_VPN
  192.0.2.1 (VPRN Label:524281 Transport:SDP:1)
172.16.2.0/24                                LOCAL
  172.16.2.0 (to-CE-2)
172.16.4.0/24                                BGP_VPN
  192.0.2.4 (VPRN Label:524280 Transport:RSVP LSP:3)
172.16.5.0/24                                BGP_VPN
  192.0.2.5 (VPRN Label:524282 Transport:LDP)
172.16.11.0/24                               BGP_VPN
  192.0.2.1 (VPRN Label:524281 Transport:SDP:1)
172.16.22.0/24                               STATIC
  172.16.2.2 (to-CE-2)
172.16.44.0/24                               BGP_VPN
  192.0.2.4 (VPRN Label:524280 Transport:RSVP LSP:3)
172.16.55.0/24                               BGP_VPN
  192.0.2.5 (VPRN Label:524282 Transport:LDP)
-----
Total Entries : 8
=====

```

The FIB shows the chosen transport tunnel, specifying SDP ID, RSVP Tunnel ID, and LDP, as well as service label information linked to the routes.

Static SDP tunnels are preferred over dynamic tunnels (RSVP or LDP auto-bind-tunnel). When the static SDP 1 is shut down or the LSP goes down (there is no fallback to dynamic tunneling), the associated routes are removed:

```

[/]
A:admin@PE-2# configure {
  service {
    sdp 1 {
      admin-state disable
    }
  }
}

```

```

[/]
A:admin@PE-2# show router 1 fib 1

=====
FIB Display
=====
Prefix [Flags]                                Protocol
NextHop
-----
172.16.2.0/24                                LOCAL
  172.16.2.0 (to-CE-2)
172.16.4.0/24                                BGP_VPN
  192.0.2.4 (VPRN Label:524280 Transport:RSVP LSP:3)
172.16.5.0/24                                BGP_VPN
  192.0.2.5 (VPRN Label:524282 Transport:LDP)
172.16.22.0/24                               STATIC
  172.16.2.2 (to-CE-2)
172.16.44.0/24                               BGP_VPN
  192.0.2.4 (VPRN Label:524280 Transport:RSVP LSP:3)
172.16.55.0/24                               BGP_VPN
  192.0.2.5 (VPRN Label:524282 Transport:LDP)
-----
Total Entries : 6
-----

```

To avoid this fallback issue, the configuration is modified and the manual spoke-SDPs are removed from the configuration of PE-1 and PE-2; the rest of the configuration remains the same. Now the connectivity between PE-1 and PE-2 is using an RSVP LSP, as shown in the PE-1 following output (RSVP LSP which was used by SDP 1 has disappeared):

```
[/]
A:admin@PE-1# configure {
  service {
    vprn "VPRN_1" {
      delete spoke-sdp 1:1
    }
  }
}
```

```
[/]
A:admin@PE-2# configure {
  service {
    vprn "VPRN_1" {
      delete spoke-sdp 1:1
    }
  }
}
```

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

```
=====
Route Table (Service: 1)
=====
```

Dest Prefix[Flags] Next Hop[Interface Name]	Type	Proto	Age Metric	Pref
172.16.1.0/24 to-CE-1	Local	Local	00h14m16s 0	0
172.16.2.0/24 192.0.2.2 (tunneled:RSVP:2)	Remote	BGP VPN	00h00m23s 10	170
172.16.4.0/24 192.0.2.4 (tunneled)	Remote	BGP VPN	00h00m23s 20	170
172.16.5.0/24 192.0.2.5 (tunneled)	Remote	BGP VPN	00h00m23s 20	170
172.16.11.0/24 172.16.1.2	Remote	Static	00h14m16s 1	5
172.16.22.0/24 192.0.2.2 (tunneled:RSVP:2)	Remote	BGP VPN	00h00m23s 10	170
172.16.44.0/24 192.0.2.4 (tunneled)	Remote	BGP VPN	00h00m23s 20	170
172.16.55.0/24 192.0.2.5 (tunneled)	Remote	BGP VPN	00h00m23s 20	170

```
-----
No. of Routes: 8
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
=====
```

If RSVP is disabled, the connectivity falls back to LDP as the output shows:

```
configure {
```

```

router "Base" {
  mpls {
    admin-state disable
  }
}

```

```

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

```

```

=====
FIB Display
=====
Prefix [Flags]                                Protocol
NextHop
-----
172.16.1.0/24                                LOCAL
  172.16.1.0 (to-CE-1)
172.16.2.0/24                                BGP_VPN
  192.0.2.2 (VPRN Label:524281 Transport:LDP)
172.16.4.0/24                                BGP_VPN
  192.0.2.4 (VPRN Label:524280 Transport:LDP)
172.16.5.0/24                                BGP_VPN
  192.0.2.5 (VPRN Label:524282 Transport:LDP)
172.16.11.0/24                               STATIC
  172.16.1.2 (to-CE-1)
172.16.22.0/24                               BGP_VPN
  192.0.2.2 (VPRN Label:524281 Transport:LDP)
172.16.44.0/24                               BGP_VPN
  192.0.2.4 (VPRN Label:524280 Transport:LDP)
172.16.55.0/24                               BGP_VPN
  192.0.2.5 (VPRN Label:524282 Transport:LDP)
-----
Total Entries : 8
=====

```

If LDP is disabled, the connectivity falls back to GRE as the output shows:

```

[/]
A:admin@PE-1# configure {
  router "Base" {
    ldp {
      admin-state disable
    }
  }
}

```

```

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

```

```

=====
FIB Display
=====
Prefix [Flags]                                Protocol
NextHop
-----
172.16.1.0/24                                LOCAL
  172.16.1.0 (to-CE-1)
172.16.2.0/24                                BGP_VPN
  192.0.2.2 (VPRN Label:524280 Transport:GRE)
172.16.4.0/24                                BGP_VPN
  192.0.2.4 (VPRN Label:524280 Transport:GRE)

```

```
172.16.5.0/24                                BGP_VPN
  192.0.2.5 (VPRN Label:524286 Transport:GRE)
172.16.11.0/24                               STATIC
  172.16.1.2 (to-CE-1)
172.16.22.0/24                               BGP_VPN
  192.0.2.2 (VPRN Label:524280 Transport:GRE)
172.16.44.0/24                               BGP_VPN
  192.0.2.4 (VPRN Label:524280 Transport:GRE)
172.16.55.0/24                               BGP_VPN
  192.0.2.5 (VPRN Label:524286 Transport:GRE)
-----
Total Entries : 8
-----
=====
```

## Conclusion

IGP shortcuts provide a variety of shortcuts in IP, MPLS, and IP-VPN scenarios to customers who want to use new options for building routing topologies. Because IGP shortcuts are enabled on a per router basis, SPF computations are independent and irrelevant to other routers, so there is no need to enable shortcuts globally. This network example shows the configuration of IGP shortcuts together with the associated show outputs which can be used for verification and troubleshooting.

# Inter-Area TE Point-to-Point LSPs

This chapter describes inter-area Traffic Engineering (TE) Point-to-Point (P2P) Label Switched Paths (LSPs) configurations.

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

## Applicability

This chapter was initially written for SR OS Release 11.0.R4, but the MD-CLI in the current edition corresponds to SR OS Release 21.2.R1.

## Overview

Multi-Protocol Label Switching with Traffic Engineering (MPLS TE) is implemented on a wide scale in current Internet Service Provider (ISP) networks to steer traffic across their backbones to facilitate efficient use of available bandwidth between the routers and to guarantee fast convergence in case a link or node fails.

Regular TE LSPs in MPLS networks are confined to only a single Interior Gateway Protocol (IGP) area or level. This is because the head-end has information in the TE database of only the local area for Open Shortest Path First (OSPF) or level for Intermediate System to Intermediate System (IS-IS). As the name implies, inter-area TE LSPs can cross the area or level borders of the IGP.

## Inter-Area TE LSP based on Explicit Route expansion

Inter-area TE LSP using Explicit Route Object (ERO) expansion enables the head-end to calculate the ERO path within its own area or level and keep the remaining Area Border Routers (ABRs) of other areas/levels as loose hops in the ERO path. On receiving a PATH message with a loose hop ERO and based on local configuration, each ABR does a partial Constrained Shortest Path First (CSPF) calculation to the next ABR or a full CSPF calculation to reach the destination.

Automatic selection of ABRs is supported so that the head-end node can work with an empty primary path. When the **to** field of an LSP definition is in an area/level different from the head-end node, CSPF will automatically compute the segment to the exit ABR router which advertised the prefix and which currently is the best path for resolving the prefix in the Route Table Manager (RTM).

## ABR protection

Link and node protection within the respective areas are supported through the TE capabilities of the IGP and Resource Reservation Protocol (RSVP) in each area. To support ABR node protection, a bypass is required from the Point of Local Repair (PLR; node prior to ABR) to the Merge Point (MP; next-hop node to ABR).

Two methods are possible: static ABR protection and dynamic ABR protection. Static ABR protection uses Manual Bypass Tunnels (MBTs), statically configured by the operator between the PLR and the MP. For dynamic ABR protection, node ID propagation and signaling of an eXclude Route Object (XRO) in RSVP PATH messages must both be supported.

Because the Record Route Object (RRO) Node ID sub-object description in RFC 4561 (*Definition of a Record Route Object (RRO) Node-Id Sub-Object*) is not clear about the format of the included node address (S), interface address (I) and label (L), the system supports multiple formats: IL, SL, ISL, SIL, SLI, ILSL and SLIL. The system uses the SLIL (node-address, label, interface-address, label) format to include the node ID itself.

The exclude route object (XRO) inclusion (RFC 4874, *Exclude Routes - Extension to Resource Reservation Protocol-Traffic Engineering*) in bypass RSVP PATH messages is required to exclude the protected ABR from the bypass path. The XRO object contains the ABR system IP address.

## Example topology

The example topology in this chapter contains ten nodes in three areas, as shown in [Figure 41: Inter-area TE LSP setup](#).

Figure 41: Inter-area TE LSP setup

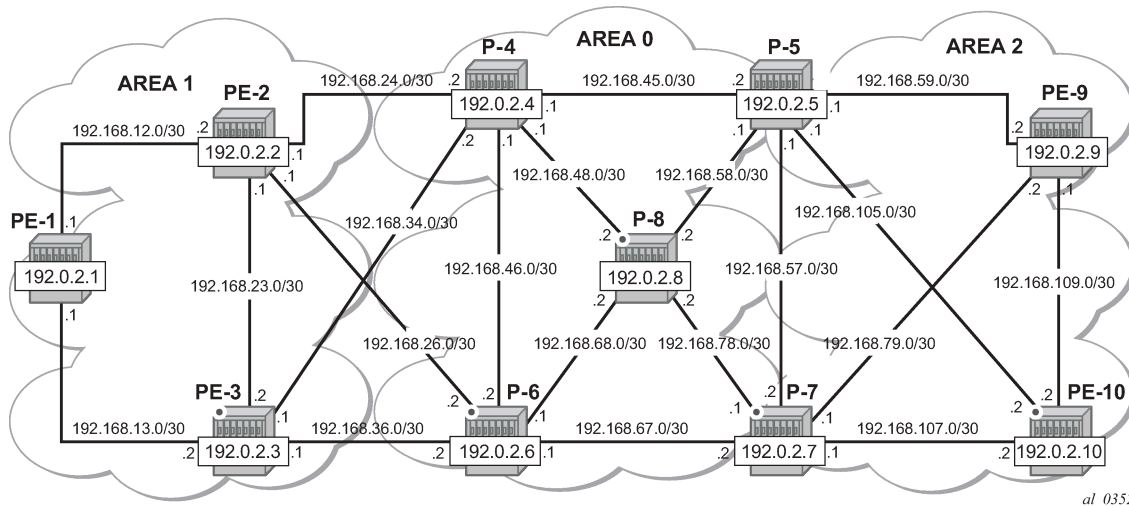
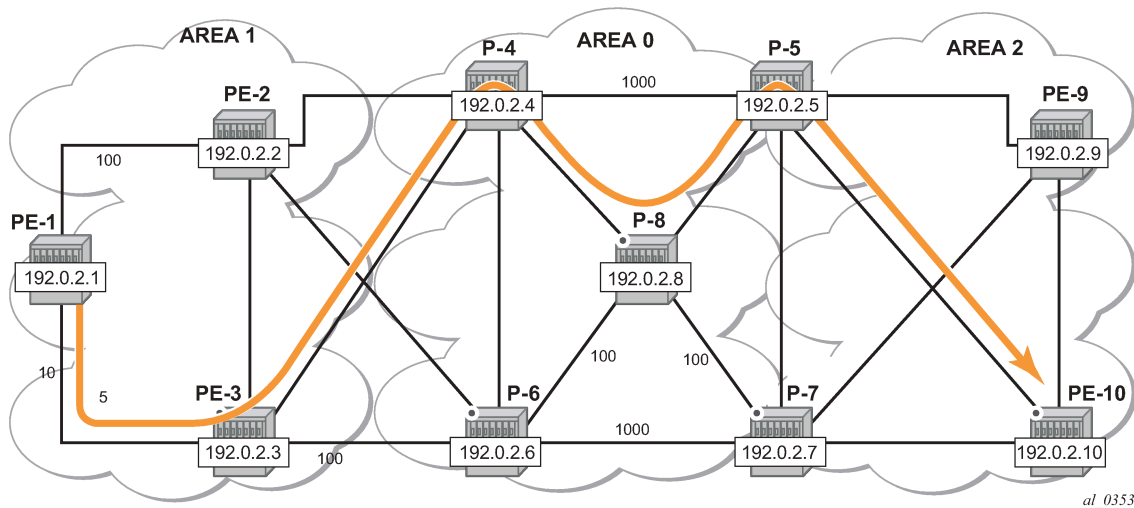


Figure 42: Inter-area TE LSP path shows the LSP path intended to be set up through the network. An empty MPLS path is used. At the head-end node PE-1, the destination address PE-10 is learned via ABR node P-4 and ABR node P-5.

Figure 42: Inter-area TE LSP path



## Configuration

The following base configuration has been implemented on the nodes:

- Cards, MDAs, and ports configured
- Interfaces configured
- IGP areas configured and converged
- Traffic Engineering configured for the IGP
- MPLS and RSVP configured on all links in the network

OSPF or IS-IS can be configured as the IGP; OSPF is used in this chapter.

The following output shows the opaque database of PE-1:

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

=====
Rtr Base OSPFv2 Instance 0 Opaque Link State Database (type: All)
=====
Type  Id                Link State Id      Adv Rtr Id         Age  Sequence          Cksum
-----
Area  0.0.0.1            1.0.0.1            192.0.2.1          269  0x80000002        0x9234
Area  0.0.0.1            1.0.0.2            192.0.2.1          236  0x80000001        0xa53c
Area  0.0.0.1            1.0.0.3            192.0.2.1          215  0x80000001        0xf3e9
Area  0.0.0.1            1.0.0.1            192.0.2.2          230  0x80000002        0x962e
Area  0.0.0.1            1.0.0.2            192.0.2.2          199  0x80000001        0xe6df
Area  0.0.0.1            1.0.0.3            192.0.2.2          165  0x80000001        0x8d32
Area  0.0.0.1            1.0.0.4            192.0.2.2          237  0x80000001        0x6d72
Area  0.0.0.1            1.0.0.5            192.0.2.2          215  0x80000001        0x7056
Area  0.0.0.1            1.0.0.1            192.0.2.3          209  0x80000002        0x9a28
Area  0.0.0.1            1.0.0.2            192.0.2.3          200  0x80000001        0x773a
Area  0.0.0.1            1.0.0.3            192.0.2.3          165  0x80000001        0x1e8c
Area  0.0.0.1            1.0.0.4            192.0.2.3          216  0x80000001        0xa933
```

Area	0.0.0.1	1.0.0.5	192.0.2.3	216	0x80000001	0x4c7a
Area	0.0.0.1	1.0.0.1	192.0.2.4	194	0x80000002	0x9e22
Area	0.0.0.1	1.0.0.6	192.0.2.4	201	0x80000001	0x7e44
Area	0.0.0.1	1.0.0.7	192.0.2.4	201	0x80000001	0x218b
Area	0.0.0.1	1.0.0.1	192.0.2.6	160	0x80000002	0xa616
Area	0.0.0.1	1.0.0.6	192.0.2.6	166	0x80000001	0xf6c5
Area	0.0.0.1	1.0.0.7	192.0.2.6	166	0x80000001	0x990d
-----						
No. of Opaque LSAs: 19						
=====						

The information is only about routers that are part of area 0.0.0.1. PE-1 cannot calculate an end-to-end CSPF path to node PE-10 because this would require TE topology information from area 0.0.0.0 and area 0.0.0.2.

Each node announces its router ID and each attached link that is part of that area, resulting in 19 opaque LSAs in area 0.0.0.1. The system interfaces of P-4 and P-6 are configured in backbone area 0.0.0.0, not in area 0.0.0.1.

In [Figure 42: Inter-area TE LSP path](#), the LSP passes through node PE-3 and node P-8. To prefer a dynamic path from PE-1 to P-4 via PE-3 rather than via PE-2, it is necessary to configure on PE-1 a lower IGP metric on the interface to PE-3 (the default metric is derived from the interface speed; in this case the metric is 10 by default).

```
# on PE-1:
configure {
  router "Base" {
    ospf 0 {
      area 0.0.0.1 {
        interface "int-PE-1-PE-3" {
          metric 5
        }
      }
    }
  }
}
```

Similarly, in the core, the IGP metric between P-4 and P-5, and between P-6 and P-7 is increased to force the LSP to pass through the core P-8 node.

```
# on P-4:
configure {
  router "Base" {
    ospf 0 {
      area 0.0.0.0 {
        interface "int-P-4-P-5" {
          metric 1000
        }
      }
    }
  }
}
```

```
# on P-6:
configure {
  router "Base" {
    ospf 0 {
      area 0.0.0.0 {
        interface "int-P-6-P-7" {
          metric 1000
        }
      }
    }
  }
}
```

Other metrics have also been manipulated as shown on [Figure 42: Inter-area TE LSP path](#).



## MPLS path configuration

An empty MPLS path is sufficient on the head-end node PE-1 because automatic ABR selection is performed. Using an empty MPLS path will ease the provisioning process and brings consistency because this empty MPLS path can be used for both intra and inter-area/level type LSPs.

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      path "empty_path" {
        admin-state enable
      }
    }
  }
}
```

## MPLS LSP configuration

On PE-1, the following LSP to PE-10 is configured with the previously created MPLS path as primary path. CSPF and fast reroute (FRR) facility are enabled on the LSP.

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-10" {
        admin-state enable
        type p2p-rsvp
        to 192.0.2.10
        path-computation-method local-cspf
        fast-reroute {
          frr-method facility
        }
        primary "empty_path" {
        }
      }
    }
  }
}
```

At this stage, the LSP is in an operational Down state with a failure code of badNode at failure node 192.168.34.2 (ABR P-4), as follows.

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-10" path

=====
MPLS LSP LSP-PE-1-PE-10 Path
=====
-----
LSP Name   : LSP-PE-1-PE-10
From       : 0.0.0.0
To         : 192.0.2.10
Adm State  : Up
Oper State : Down
-----
Path Name      Next Hop      Type      Out I/F  Adm  Opr
-----
empty_path    n/a          Primary   n/a      Up   Dwn
```

```

=====
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-10" path detail

=====
MPLS LSP LSP-PE-1-PE-10 Path (Detail)
=====
Legend :
  @ - Detour Available          # - Detour In Use
  b - Bandwidth Protected       n - Node Protected
  s - Soft Preemption
  S - Strict                    L - Loose
  A - ABR                       + - Inherited
=====

-----
LSP LSP-PE-1-PE-10
Path empty_path
-----
LSP Name      : LSP-PE-1-PE-10
From          : 192.0.2.1
To            : 192.0.2.10
Admin State   : Up                Oper State    : Down
Path Name     : empty_path
Path LSP ID   : 24596             Path Type     : Primary
Path Admin    : Up                Path Oper     : Down
Out Interface : n/a              Out Label     : n/a
Path Up Time  : 0d 00:00:00      Path Down Time : 0d 00:05:22
Retry Limit   : 0                 Retry Timer    : 30 sec
Retry Attempt : 11                Next Retry In  : 20 sec

BFD Configuration and State
Template      : None                Ping Interval  : 60
Enable       : False              State          : notApplicable
WaitForUpTimer : 4 sec           OperWaitForUpTimer: N/A
WaitForUpTmLeft : 0 sec

Adspec       : Disabled            Oper Adspec    : N/A
PathCompMethod : local-cspf        OperPathCompMethod: N/A
MetricType    : igp                Oper MetricType : N/A
Least Fill    : Disabled           Oper LeastFill  : N/A
FRR           : Enabled            Oper FRR        : N/A
FRR NodeProtect : Enabled          Oper FRR NP     : N/A
FR Hop Limit  : 16                 Oper FRHopLimit : N/A
FR Prop Admin Gr* : Disabled        Oper FRPropAdmGrp : N/A
Propagate Adm Grp: Disabled        Oper Prop Adm Grp : N/A
Inter-area    : N/A

PCE Report    : Disabled+          Oper PCE Report : Disabled
PCE Control   : Disabled          Oper PCE Control : Disabled
PCE Update ID : 0

Neg MTU       : 0                  Oper MTU        : N/A
Bandwidth     : No Reservation     Oper Bandwidth   : N/A
Hop Limit     : 255                 Oper HopLimit    : N/A
Record Route  : Record             Oper Record Route : N/A
Record Label  : Record             Oper Record Label : N/A
Setup Priority : 7                  Oper SetupPriority: N/A
Hold Priority  : 0                  Oper HoldPriority : N/A
Class Type    : 0                  Oper CT          : N/A
Backup CT     : None
MainCT Retry  : Infinite
Rem           :

```

```

MainCT Retry      : 0
  Limit           :
Include Groups    :                               Oper IncludeGroups:
None                                                      N/A
Exclude Groups    :                               Oper ExcludeGroups:
None                                                      N/A

Adaptive          : Enabled                       Oper Metric       : N/A
Preference        : n/a
Path Trans        : 0                             CSPF Queries      : 11
Failure Code     : badNode
Failure Node    : 192.168.34.2
Explicit Hops     :
  No Hops Specified
Actual Hops       :
  No Hops Specified
Computed Hops     :
  No Hops Specified
Resignal Eligible: False
Last Resignal     : n/a                           CSPF Metric       : N/A
=====
* indicates that the corresponding row element may have been truncated.

```

To get around the intra-area CSPF confinement, the ERO-expansion feature is enabled on all ABR nodes.

```

# on P-4, P-5, P-6, P-7:
configure {
  router "Base" {
    mpls {
      cspf-on-loose-hop true
    }
  }
}

```

**cspf-on-loose-hop** is only required if FRR or TE parameters are configured on the LSP. If any of these parameters is configured on the LSP while one of the ABRs along the path is not configured with **cspf-on-loose-hop true**, the LSP will stay operationally down with failure code: badNode and an indication of the interface address of the failure node.

The LSP path can also contain other strict and/or loose hops. However, **cspf-on-loose-hop** must be enabled in the **mpls** context whenever loose hops are configured in the MPLS path. This command enables ERO expansion and is required for inter-area LSPs on all possible ABR nodes and all nodes not belonging to the area where the iLER is located, which have a loose hop reference in the MPLS path.



**Note:** The LSP may fail to set up if **cspf-on-loose-hop** is enabled on an LSR that is not an ABR and that receives a PATH message without a proper next loose hop in the ERO.

On all nodes, debugging is enabled—in classic CLI—for RSVP PATH messages, as follows:

```

# on all nodes:
debug
  router "Base"
    rsvp
      packet
        path detail
      exit
    exit
  exit
exit

```

The following RSVP PATH message on PE-1 shows the ERO calculation on the head-end to the first ABR.

```

# on PE-1:

```

```

2 2021/05/12 08:13:36.522 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: PATH Msg
Send PATH From:192.0.2.1, To:192.0.2.10
          TTL:255, Checksum:0x5b2d, Flags:0x0
Session   - EndPt:192.0.2.10, TunnId:1, ExtTunnId:192.0.2.1
SessAttr  - Name:LSP-PE-1-PE-10::empty_path
          SetupPri:7, HoldPri:0, Flags:0x17
RSVPHop   - Ctype:1, Addr:192.168.13.1, LIH:3
TimeValue - RefreshPeriod:30
SendTempl - Sender:192.0.2.1, LspId:24648
SendTSpec - Ctype:QOS, CDR:0.000 bps, PBS:0.000 bps, PDR:infinity
          MPU:20, MTU:1564
LabelReq  - IfType:General, L3ProtID:2048
RRO      - IpAddr:192.168.13.1, Flags:0x0
ERO      - IPv4Prefix 192.168.13.2/32, Strict
          IPv4Prefix 192.168.34.2/32, Strict
          IPv4Prefix 192.0.2.10/32, Loose
FRRObj    - SetupPri:7, HoldPri:0, HopLimit:16, BW:0.000 bps, Flags:0x2
          ExcAny:0x0, IncAny:0x0, IncAll:0x0
"

```

On the ABR P-4, the ERO is expanded to include the nodes of area 0.0.0.0 of which P-4 is also part. The RRO contains all the hops the PATH message has passed so far.

```

# on P-4:
2 2021/05/12 08:13:24.842 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: PATH Msg
Send PATH From:192.0.2.1, To:192.0.2.10
          TTL:253, Checksum:0xe9a8, Flags:0x0
Session   - EndPt:192.0.2.10, TunnId:1, ExtTunnId:192.0.2.1
SessAttr  - Name:LSP-PE-1-PE-10::empty_path
          SetupPri:7, HoldPri:0, Flags:0x17
RSVPHop   - Ctype:1, Addr:192.168.48.1, LIH:4
TimeValue - RefreshPeriod:30
SendTempl - Sender:192.0.2.1, LspId:24648
SendTSpec - Ctype:QOS, CDR:0.000 bps, PBS:0.000 bps, PDR:infinity
          MPU:20, MTU:1564
LabelReq  - IfType:General, L3ProtID:2048
RRO      - IpAddr:192.168.48.1, Flags:0x0
          IpAddr:192.168.34.1, Flags:0x0
          IpAddr:192.168.13.1, Flags:0x0
ERO      - IPv4Prefix 192.168.48.2/32, Strict
          IPv4Prefix 192.168.58.1/32, Strict
          IPv4Prefix 192.0.2.10/32, Loose
FRRObj    - SetupPri:7, HoldPri:0, HopLimit:16, BW:0.000 bps, Flags:0x2
          ExcAny:0x0, IncAny:0x0, IncAll:0x0
"

```

Finally, the P-5 ABR will expand the ERO to the final destination PE-10:

```

# on P-5:
7 2021/05/12 08:13:38.230 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: PATH Msg
Send PATH From:192.0.2.1, To:192.0.2.10
          TTL:251, Checksum:0x9207, Flags:0x0
Session   - EndPt:192.0.2.10, TunnId:1, ExtTunnId:192.0.2.1
SessAttr  - Name:LSP-PE-1-PE-10::empty_path
          SetupPri:7, HoldPri:0, Flags:0x17
RSVPHop   - Ctype:1, Addr:192.168.105.1, LIH:5
TimeValue - RefreshPeriod:30
SendTempl - Sender:192.0.2.1, LspId:24648
SendTSpec - Ctype:QOS, CDR:0.000 bps, PBS:0.000 bps, PDR:infinity

```

```

LabelReq      MPU:20, MTU:1564
RR0           - IfType:General, L3ProtID:2048
              - IpAddr:192.168.105.1, Flags:0x0
              - IpAddr:192.168.58.2, Flags:0x0
              - IpAddr:192.168.48.1, Flags:0x0
              - IpAddr:192.168.34.1, Flags:0x0
              - IpAddr:192.168.13.1, Flags:0x0
ER0           - IPv4Prefix 192.168.105.2/32, Strict
FRR0obj      - SetupPri:7, HoldPri:0, HopLimit:16, BW:0.000 bps, Flags:0x2
              - ExcAny:0x0, IncAny:0x0, IncAll:0x0
"
    
```

The MPLS LSP is now operational Up and the LSP path can be shown in detail on the head-end, PE-1:

```

[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-10" path

=====
MPLS LSP LSP-PE-1-PE-10 Path
=====
-----
LSP Name      : LSP-PE-1-PE-10
From          : 0.0.0.0
To            : 192.0.2.10
Adm State     : Up
Oper State    : Up
-----
Path Name      Next Hop      Type      Out I/F  Adm  Opr
-----
empty_path     192.168.13.2  Primary   1/1/2    Up   Up
=====
    
```

```

[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-10" path detail

=====
MPLS LSP LSP-PE-1-PE-10 Path (Detail)
=====
Legend :
@ - Detour Available      # - Detour In Use
b - Bandwidth Protected  n - Node Protected
s - Soft Preemption
S - Strict                L - Loose
A - ABR                   + - Inherited
=====
-----
LSP LSP-PE-1-PE-10
Path empty_path
-----
LSP Name      : LSP-PE-1-PE-10
From          : 192.0.2.1
To            : 192.0.2.10
Admin State   : Up
Oper State    : Up
Path Name     : empty_path
Path LSP ID   : 24648
Path Admin    : Up
Path Type     : Primary
Out Interface : 1/1/2
Path Oper     : Up
Out Label     : 524287
Path Up Time  : 0d 00:02:07
Path Down Time : 0d 00:00:00
Retry Limit   : 0
Retry Timer   : 30 sec
Retry Attempt : 0
Next Retry In : 0 sec
-----
---snip---
    
```

```

Adspec           : Disabled           Oper Adspec      : Disabled
PathCompMethod  : local-cspf         OperPathCompMethod: local-cspf
MetricType      : igp                Oper MetricType  : igp
Least Fill      : Disabled           Oper LeastFill   : Disabled
FRR             : Enabled            Oper FRR         : Enabled
FRR NodeProtect : Enabled            Oper FRR NP      : Enabled
FR Hop Limit    : 16                 Oper FRHopLimit  : 16
FR Prop Admin Gr* : Disabled         Oper FRPropAdmGrp : Disabled
Propagate Adm Grp : Disabled         Oper Prop Adm Grp : Disabled
Inter-area      : True

---snip---

Adaptive        : Enabled            Oper Metric      : 15
Preference      : n/a
Path Trans      : 1                  CSPF Queries     : 37
Failure Code    : noError
Failure Node    : n/a
Explicit Hops   :
  No Hops Specified
Actual Hops     :
  192.168.13.1(192.0.2.1) @ n       Record Label     : N/A
-> 192.168.13.2(192.0.2.3) @      Record Label     : 524287
-> 192.168.34.2(192.0.2.4) @ n    Record Label     : 524287
-> 192.168.48.2 @                  Record Label     : 524287
-> 192.168.58.1 @                  Record Label     : 524287
-> 192.168.105.2 @                 Record Label     : 524287
Computed Hops   :
  192.168.13.1(S)
-> 192.168.13.2(S)
-> 192.168.34.2(SA)
-> 192.0.2.10(L)
Resignal Eligible: False
Last Resignal   : n/a              CSPF Metric      : 15
=====
* indicates that the corresponding row element may have been truncated.

```

## ABR node protection

The LSP is configured with facility FRR protection; link and node protection are established within each area, as shown in the preceding output. Node protection is available for nodes PE-3 in area 1 (bypass originating in PE-1), and P-8 in area 0 (bypass originating in P-4), but not for the ABRs P-4 and P-5. No bypass tunnels for node protection originate in PLRs PE-3 (for ABR P-4) or P-8 (for ABR P-5). The bypass tunnels originating in PE-3 and P-8 only offer link protection. Dynamic ABR node protection requires the setup of a bypass tunnel from the PLR (node just upstream of the ABR) to the MP (node just downstream of the ABR). The following two things are required to establish a bypass tunnel for an ABR:

- The PLR node (part of area x) needs to know the system IP address of the MP node (part of area y) to set up the bypass. For this reason, the node ID of the MP must be included in the RESV message so that the PLR can link the manual bypass tunnel to the primary path to protect the ABR. By default, the node ID is not included in the RESV message, but it can be configured on the MPs with the **include-node-id-in-rro** command.
- The other ABR node receiving the RSVP bypass PATH message for the protected ABR needs to do an ERO expansion toward the MP node. For this reason, the XRO object is included in the RSVP bypass PATH message, containing the node ID of the protected ABR. As an example, the following bypass PATH message is shown on node PE-3.

The XRO object includes the system IP address of the protected ABR node P-4 and the ERO object has MP node P-8 as loose destination:

```
# on PE-3:
46 2021/05/12 08:20:46.609 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: PATH Msg
Send PATH From:192.0.2.3, To:192.0.2.8
      TTL:17, Checksum:0xfddd, Flags:0x0
Session   - EndPt:192.0.2.8, TunnId:61442, ExtTunnId:192.0.2.3
SessAttr  - Name:bypass-node192.0.2.4-61442
          SetupPri:7, HoldPri:0, Flags:0x2
RSVPHop   - Ctype:1, Addr:192.168.36.1, LIH:3
TimeValue - RefreshPeriod:30
SendTempl - Sender:192.0.2.3, LspId:4
SendTSpec - Ctype:QOS, CDR:0.000 bps, PBS:0.000 bps, PDR:infinity
          MPU:20, MTU:1564
LabelReq  - IfType:General, L3ProtID:2048
RR0       - IpAddr:192.168.36.1, Flags:0x0
ER0       - IPv4Prefix 192.168.36.2/32, Strict
IPv4Prefix 192.0.2.8/32, Loose
XRO       - IPv4Prefix: 192.0.2.4/32, Attribute: Node, LBit: Exclude
AdSpec    - General BreakBit:0, NumISHops:0, PathBwEstimate:0
          MinPathLatency:4294967295, CompPathMTU:1564
          Controlled BreakBit:0
"
```

## Node-ID inclusion in the RESV message

P-8 will be the MP for the bypass of ABR P-4 and PE-10 will be the MP for the bypass of ABR P-5. So P-8 and PE-10 need to include their node ID in the RESV message, inside the Record Route Object (RRO).

```
# on P-8 and PE-10:
configure {
  router "Base" {
    rsvp {
      include-node-id-in-rro true
    }
  }
}
```

The default is **include-node-id-in-rro false**.

On PE-3, debugging is enabled—in classic CLI—for RSVP RESV messages, as follows:

```
# on PE-3:
debug
  router "Base"
    rsvp
      packet
        resv detail
      exit
    exit
```

The following RESV message is received on PLR node PE-3. The RRO contains the MP node P-8 information in SLIL format:

```
52 2021/05/12 08:21:19.609 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: RESV Msg
Send RESV From:192.168.13.2, To:192.168.13.1
      TTL:255, Checksum:0x7e4b, Flags:0x0
Session   - EndPt:192.0.2.10, TunnId:1, ExtTunnId:192.0.2.1
RSVPHop   - Ctype:1, Addr:192.168.13.2, LIH:3
```

```

TimeValue - RefreshPeriod:30
Style - SE
FlowSpec - Ctype:QOS, CDR:0.000 bps, PBS:0.000 bps, PDR:infinity
           MPU:20, MTU:1560, RSpecRate:0, RSpecSlack:0
FilterSpec - Sender:192.0.2.1, LspId:24648, Label:524287
RR0 - InterfaceIp:192.168.13.2, Flags:0x9
      Label:524287, Flags:0x1
      InterfaceIp:192.168.34.2, Flags:0x9
      Label:524287, Flags:0x1
      SystemIp:192.0.2.8, Flags:0x29
      Label:524287, Flags:0x1
      InterfaceIp:192.168.48.2, Flags:0x9
      Label:524287, Flags:0x1
      SystemIp:192.0.2.5, Flags:0x21
      Label:524287, Flags:0x1
      InterfaceIp:192.168.58.1, Flags:0x1
      Label:524287, Flags:0x1
      SystemIp:192.0.2.10, Flags:0x20
      Label:524287, Flags:0x1
      InterfaceIp:192.168.105.2, Flags:0x0
      Label:524287, Flags:0x1
"

```

### Bypass configuration for ABR protection

Because dynamic ABR protection is supported and used in this example, no explicit Manual Bypass Tunnels (MBTs) are configured to protect the ABRs. Each PLR first checks if an MBT tunnel exists between the PLR and the MP matching the constraints and protecting the ABR. If no MBT is available, the PLR will signal a bypass tunnel in a dynamic way toward the MP node.

Figure 43: ABR protection shows the two dynamic ABR node protections that are signaled for this LSP.

Figure 43: ABR protection

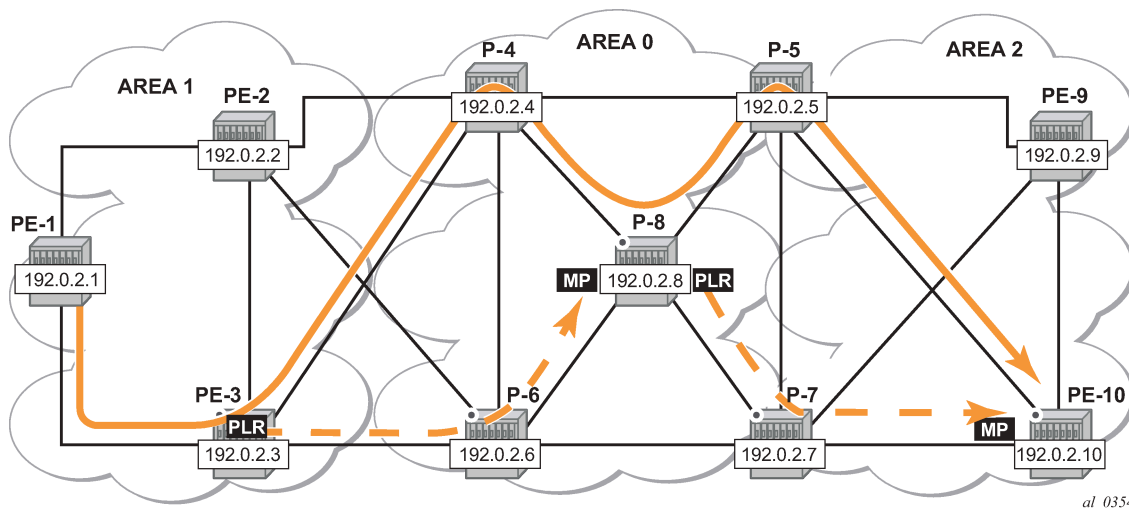
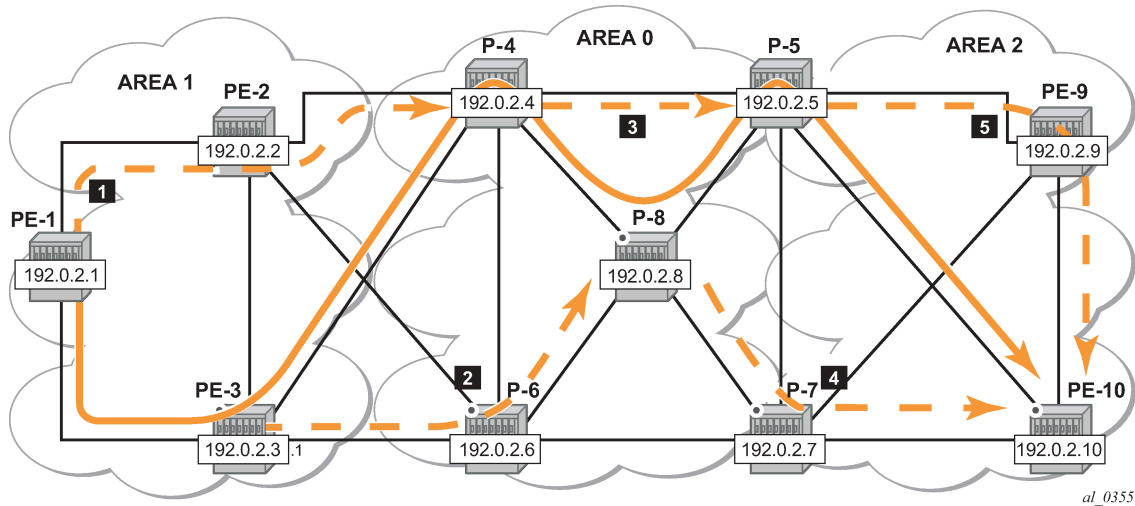


Figure 44: Protection of all nodes/links along the LSP path shows the complete picture of all the FRR protections and indicates each node/link protection in the setup.



Figure 44: Protection of all nodes/links along the LSP path



This can be seen in the detailed show output of the LSP path:

```
[/]
```

```
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-10" path detail
```

```
=====
```

```
MPLS LSP LSP-PE-1-PE-10 Path (Detail)
```

```
=====
```

```
Legend :
```

```
@ - Detour Available          # - Detour In Use
```

```
b - Bandwidth Protected      n - Node Protected
```

```
s - Soft Preemption
```

```
S - Strict                    L - Loose
```

```
A - ABR                       + - Inherited
```

```
=====
```

```
LSP LSP-PE-1-PE-10
```

```
Path empty_path
```

```
-----
```

```
LSP Name      : LSP-PE-1-PE-10
```

```
From          : 192.0.2.1
```

```
To            : 192.0.2.10
```

```
Admin State   : Up
```

```
Oper State    : Up
```

```
Path Name     : empty_path
```

```
Path LSP ID   : 24648
```

```
Path Type     : Primary
```

```
Path Admin    : Up
```

```
Path Oper     : Up
```

```
Out Interface : 1/1/2
```

```
Out Label     : 524287
```

```
Path Up Time  : 0d 00:19:36
```

```
Path Down Time : 0d 00:00:00
```

```
Retry Limit   : 0
```

```
Retry Timer   : 30 sec
```

```
Retry Attempt : 0
```

```
Next Retry In : 0 sec
```

```
---
```

```
Adspec        : Disabled
```

```
Oper Adspec   : Disabled
```

```
PathCompMethod : local-cspf
```

```
OperPathCompMethod : local-cspf
```

```
MetricType    : igp
```

```
Oper MetricType : igp
```

```
Least Fill    : Disabled
```

```
Oper LeastFill : Disabled
```

```
FRR           : Enabled
```

```
Oper FRR      : Enabled
```

```
FRR NodeProtect : Enabled
```

```
Oper FRR NP   : Enabled
```

```

FR Hop Limit      : 16                Oper FRHopLimit   : 16
FR Prop Admin Gr*: Disabled          Oper FRPropAdmGrp : Disabled
Propagate Adm Grp: Disabled          Oper Prop Adm Grp : Disabled
Inter-area       : True

---snip---

Adaptive         : Enabled            Oper Metric       : 15
Preference      : n/a
Path Trans      : 1                  CSPF Queries      : 37
Failure Code    : noError
Failure Node    : n/a
Explicit Hops   :
  No Hops Specified
Actual Hops     :
  192.168.13.1(192.0.2.1) @ n        Record Label     : N/A
-> 192.168.13.2(192.0.2.3) @ n        Record Label     : 524287
-> 192.168.34.2(192.0.2.4) @ n        Record Label     : 524287
-> 192.0.2.8(192.0.2.8) @ n          Record Label     : 524287
-> 192.168.48.2 @ n                  Record Label     : 524287
-> 192.0.2.5(192.0.2.5) @           Record Label     : 524287
-> 192.168.58.1 @                    Record Label     : 524287
-> 192.0.2.10(192.0.2.10)            Record Label     : 524287
-> 192.168.105.2                      Record Label     : 524287
Computed Hops   :
  192.168.13.1(S)
-> 192.168.13.2(S)
-> 192.168.34.2(SA)
-> 192.0.2.10(L)
Resignal Eligible: False
Last Resignal   : n/a                CSPF Metric       : 15
=====
* indicates that the corresponding row element may have been truncated.

```

- The first bypass originates in PE-1 and protects node PE-3.
- The second bypass originates in PE-3 and protects node P-4.
- The third bypass originates in P-4 and protects node P-8.
- The fourth bypass originates in P-8 and protects node P-5. There are two entries for P-8: hop 192.0.2.8 and hop 192.168.48.2.
- The fifth bypass originates in P-5 and protects the link between P-5 and PE-10. There are two entries for P-5: hop 192.0.2.5 and hop 192.168.58.1.

There are two entries for P-8, P-5 and PE-10 in the 'Actual Hops' section in the previous output: one for the interface IP address and one for the system IP address. This is a consequence of configuring **include-node-id-in-rro true** on P-8, P-5, and PE-10.

The **include-node-id-in-rro true** command is not mandatory for this example on ABR node P-5, but to be able to cover cases where a new LSP is established in the network and P-5 acts as an MP node while the corresponding PLR node for that new LSP is in another area. This RSVP command can be executed on all possible MP nodes in the network.

The following command shows the details of the bypass tunnel from PE-3 to PE-8, protecting PE-4:

```

[/]
A:admin@PE-3# show router mpls bypass-tunnel protected-lsp detail

=====
MPLS Bypass Tunnels (Detail)
=====

```

```

-----
bypass-node192.0.2.4-61442
-----
To           : 192.0.2.8           State           : Up
Out I/F      : 1/1/2              Out Label      : 524287
Up Time     : 0d 00:08:18         Active Time    : n/a
Reserved BW  : 0 Kbps             Protected LSP Count : 1
Type        : Dynamic            Bypass Path Cost : 100
Setup Priority : 7                Hold Priority   : 0
Class Type   : 0
Exclude Node : 192.0.2.4         Inter-Area     : True
Computed Hops :
  192.168.36.1(S)                Egress Admin Groups : None
-> 192.168.36.2(SA)              Egress Admin Groups : None
-> 192.0.2.8(L)                  Egress Admin Groups : None
Actual Hops  :
  192.168.36.1(192.0.2.3)        Record Label     : N/A
-> 192.168.36.2(192.0.2.6)        Record Label     : 524287
-> 192.0.2.8(192.0.2.8)          Record Label     : 524286
-> 192.168.68.2                  Record Label     : 524286
Last Resignal :
Attempted At : n/a               Resignal Reason  : n/a
Resignal Status: n/a            Reason           : n/a

Protected LSPs -
LSP Name     : LSP-PE-1-PE-10::empty_path
From         : 192.0.2.1         To               : 192.0.2.10
Avoid Node/Hop : 192.0.2.4     Downstream Label : 524287
Bandwidth    : 0 Kbps
=====

```

The LSP could be protected with one or more additional secondary paths, pre-sigaled or not, but this is outside the scope of this chapter.

When a link or node failure occurs along the LSP path, FRR protection kicks in and end-to-end path re-optimization is executed: a PATHERR message is forwarded to the head-end. Upon receiving the PATHERR message, the head-end calculates a new path.

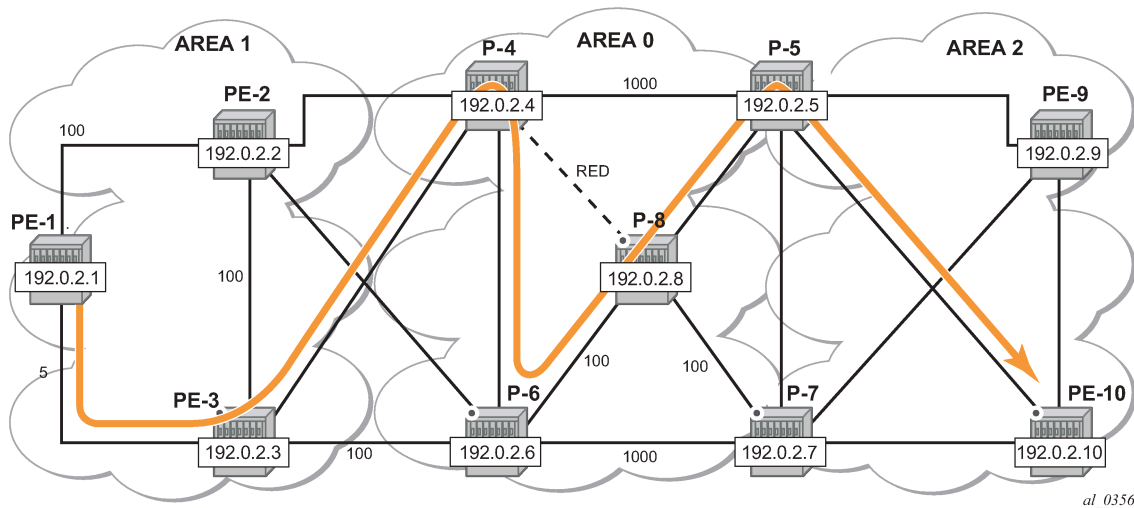
## Admin groups

The use of administrative groups is described in the [RSVP Point-to-Point LSPs](#) chapter.

To support admin groups for inter-area LSPs, the ingress node PE-1 must propagate the admin groups within the Session Attribute object (SA) of the PATH message so that the ABRs along the path receive the admin group restrictions they have to take into account when further expanding the ERO in the PATH message.

In [Figure 44: Protection of all nodes/links along the LSP path](#) the LSP path avoids the link between P-4 and P-8. This is implemented by assigning admin group "red" to the link between P-4 and P-8 and then configuring the LSP to exclude the admin group "red".

Figure 45: Admin group example



## Admin group configuration

On P-4, configure admin group "red" and assign a group value. In this example, group value 11 is used, but this can be any value between 0 and 31. Assign admin group "red" to the link to P-8.

This admin group configuration is required on P-4 and on iLER PE-1. However, it is good practice to configure the admin group on all the nodes.

```
# on all nodes:
configure {
  routing-options {
    if-attribute {
      admin-group "red" {
        value 11
      }
    }
  }
}
```

```
# on P-4:
configure {
  router "Base" {
    mpls {
      interface "int-P-4-P-8" {
        admin-group ["red"]
      }
    }
  }
}
```

On PE-1, change the LSP configuration as follows:

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-10" {
        exclude-admin-group "red"
        propagate-admin-group true
      }
    }
  }
}
```

It is possible to have the same admin group constraint applied to the FRR bypass tunnels in the PLRs, but that is not the case here. The bypass tunnels ignore any admin group constraint. The **propagate-admin-group** command is required to include the admin group properties in the SA object of the PATH message. The admin group value is mapped to a 32-bitmap. In this example, value 11 means that the 12th bit is set, which means in binary 100000000000 or hex 0x800.

```
# on PE-1:
68 2021/05/12 08:29:57.687 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: PATH Msg
Send PATH From:192.0.2.1, To:192.0.2.10
          TTL:255, Checksum:0x5319, Flags:0x0
Session   - EndPt:192.0.2.10, TunnId:1, ExtTunnId:192.0.2.1
SessAttr  - Name:LSP-PE-1-PE-10::empty_path
          SetupPri:7, HoldPri:0, Flags:0x17
          Ctype:RA, ExcAny:0x800, IncAny:0x0, IncAll:0x0
RSVPHop   - Ctype:1, Addr:192.168.13.1, LIH:3
TimeValue - RefreshPeriod:30
SendTempl - Sender:192.0.2.1, LspId:24650
SendTSpec - Ctype:QOS, CDR:0.000 bps, PBS:0.000 bps, PDR:infinity
          MPU:20, MTU:1564
LabelReq  - IfType:General, L3ProtID:2048
RR0       - IpAddr:192.168.13.1, Flags:0x0
ERO       - IPv4Prefix 192.168.13.2/32, Strict
          IPv4Prefix 192.168.34.2/32, Strict
          IPv4Prefix 192.0.2.10/32, Loose
FRR0bj    - SetupPri:7, HoldPri:0, HopLimit:16, BW:0.000 bps, Flags:0x2
          ExcAny:0x0, IncAny:0x0, IncAll:0x0
"
```

The following two sets of output show that when P-4 expands the ERO it now excludes the link to node P-8 for the path calculation and the path is set up through P-6, P-8 and P-5.

```
# on P-4:
147 2021/05/12 08:29:57.693 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: PATH Msg
Send PATH From:192.0.2.1, To:192.0.2.10
          TTL:253, Checksum:0xc1d2, Flags:0x0
Session   - EndPt:192.0.2.10, TunnId:1, ExtTunnId:192.0.2.1
SessAttr  - Name:LSP-PE-1-PE-10::empty_path
          SetupPri:7, HoldPri:0, Flags:0x17
          Ctype:RA, ExcAny:0x800, IncAny:0x0, IncAll:0x0
RSVPHop   - Ctype:1, Addr:192.168.46.1, LIH:3
TimeValue - RefreshPeriod:30
SendTempl - Sender:192.0.2.1, LspId:24650
SendTSpec - Ctype:QOS, CDR:0.000 bps, PBS:0.000 bps, PDR:infinity
          MPU:20, MTU:1564
LabelReq  - IfType:General, L3ProtID:2048
RR0       - IpAddr:192.168.46.1, Flags:0x0
          IpAddr:192.168.34.1, Flags:0x0
          IpAddr:192.168.13.1, Flags:0x0
ERO       - IPv4Prefix 192.168.46.2/32, Strict
          IPv4Prefix 192.168.68.2/32, Strict
          IPv4Prefix 192.168.58.1/32, Strict
          IPv4Prefix 192.0.2.10/32, Loose
FRR0bj    - SetupPri:7, HoldPri:0, HopLimit:16, BW:0.000 bps, Flags:0x2
          ExcAny:0x0, IncAny:0x0, IncAll:0x0
"
```

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-10" path detail
```

```

=====
MPLS LSP LSP-PE-1-PE-10 Path (Detail)
=====
Legend :
@ - Detour Available          # - Detour In Use
b - Bandwidth Protected      n - Node Protected
s - Soft Preemption
S - Strict                   L - Loose
A - ABR                      + - Inherited
=====

LSP LSP-PE-1-PE-10
Path empty_path
-----
LSP Name      : LSP-PE-1-PE-10
From          : 192.0.2.1
To           : 192.0.2.10
Admin State   : Up
Path Name    : empty_path
Path LSP ID  : 24650
Path Admin   : Up
Out Interface : 1/1/2
Path Up Time : 0d 00:26:29
Retry Limit  : 0
Retry Attempt : 0
Oper State    : Up
Path Type    : Primary
Path Oper    : Up
Out Label    : 524286
Path Down Time : 0d 00:00:00
Retry Timer  : 30 sec
Next Retry In : 0 sec

---snip---

Adspec       : Disabled
PathCompMethod : local-cspf
MetricType   : igp
Least Fill   : Disabled
FRR          : Enabled
FRR NodeProtect : Enabled
FR Hop Limit : 16
FR Prop Admin Gr* : Disabled
Propagate Adm Grp : Enabled
Inter-area   : True
Oper Adspec  : Disabled
OperPathCompMethod : local-cspf
Oper MetricType : igp
Oper LeastFill : Disabled
Oper FRR     : Enabled
Oper FRR NP  : Enabled
Oper FRHopLimit : 16
Oper FRPropAdmGrp : Disabled
Oper Prop Adm Grp : Enabled

---snip---

Include Groups :
None
Oper IncludeGroups:
None
Exclude Groups :
red
Oper ExcludeGroups:
red

Adaptive      : Enabled
Preference    : n/a
Path Trans    : 2
Failure Code  : noError
Failure Node  : n/a
Explicit Hops :
No Hops Specified
Actual Hops   :
192.168.13.1(192.0.2.1) @ n
-> 192.168.13.2(192.0.2.3) @ n
-> 192.168.34.2(192.0.2.4) @ n
-> 192.168.46.2 @ n
-> 192.0.2.8(192.0.2.8) @ n
-> 192.168.68.2 @ n
-> 192.0.2.5(192.0.2.5) @
-> 192.168.58.1 @
-> 192.0.2.10(192.0.2.10)
-> 192.168.105.2
Oper Metric   : 15
CSPF Queries : 38
Record Label  : N/A
Record Label  : 524286
Record Label  : 524285
Record Label  : 524286
Record Label  : 524285
Record Label  : 524285
Record Label  : 524285
Record Label  : 524285
Record Label  : 524284
Record Label  : 524284

```

```

Computed Hops      :
  192.168.13.1(S)
-> 192.168.13.2(S)
-> 192.168.34.2(SA)
-> 192.0.2.10(L)
Resignal Eligible: False
Last Resignal     : n/a                CSPF Metric      : 15
Last MBB         :
MBB Type         : ConfigChange        MBB State       : Success
Ended At        : 05/12/2021 08:29:59 Old Metric      : 15
Signaled BW     : 0 Mbps
Fail Code       : noError
=====
* indicates that the corresponding row element may have been truncated.

```

## Shared Risk Link Groups (SRLG)

Shared risk link groups are described in chapter [Shared Risk Link Groups for RSVP-Based LSPs](#).

SRLGs are also supported in the context of inter-area TE LSPs. SRLGs refer to situations where links in a network share a common fiber (or a common physical attribute). If one link fails, other links in the group may fail as well. Links in the group have fate sharing.

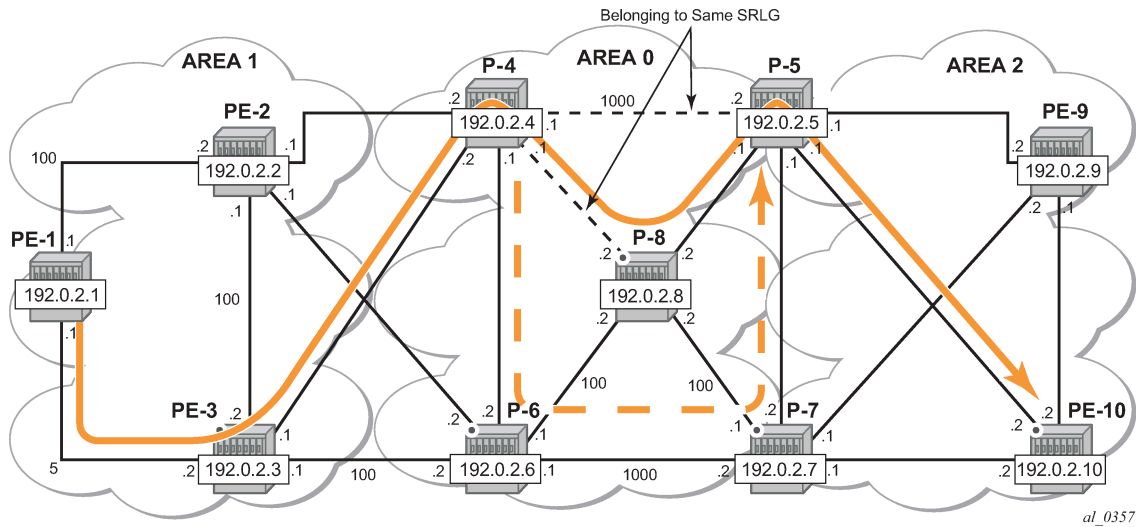
The MPLS TE SRLG feature enhances backup tunnel path selection so that a backup tunnel avoids using links that are in the same SRLG.

Consider the setup in [Figure 46: Share Risk Link Groups](#), where an inter-area LSP is set up from PE-1 to PE-10 and the path goes through P-8 because of a lower IGP metric. To protect against a node failure of P-8, P-4 (PLR) would normally set up an FRR backup directly to P-5 (MP), because of the lower IGP metric (P-4 to P-5:1000) compared to the IGP traffic via P-6 (P-4 to P-6 to P-7 to P-5:1020).

However, imagine that in this setup the link between P-4 and P-5 and the link between P-4 and P-8 are part of the same transmission bundle. In this case, a cut of that fiber bundle will bring down both the primary and the backup path.

This can be avoided by configuring these two links in the same SRLG group and enabling **srlg-frr strict** on P-4. In that case, the backup will be set up via P-6 as indicated by the dashed line in [Figure 46: Share Risk Link Groups](#).

Figure 46: Share Risk Link Groups



## SRLG configuration

On P-4, an SRLG group is configured, `srlg-frr strict` is enabled and the links to P-5 and to P-8 are added to this SRLG group.

The SRLG group configuration is required on all nodes that use SRLG groups and on the ABR used by the inter-area TE LSP. In this example, it is configured on all nodes.



### Note:

Enabling or disabling `srlg-frr` command only takes effect when the LSP primary path or the bypass path is resignaled. The user can either wait for the resignal timer to expire or cause the paths to be resignaled immediately by executing, at the ingress LER, the manual resignal command for the LSP primary path or for the bypass LSP path.

```
# on all nodes:
configure {
  routing-options {
    if-attribute {
      srlg-group "bundle-red" {
        value 1
      }
    }
  }
}

# on P-4:
configure {
  router "Base" {
    mpls {
      srlg-frr strict
      interface "int-P-4-P-5" {
        srlg-group "bundle-red" { }
      }
      interface "int-P-4-P-8" {
        srlg-group "bundle-red" { }
      }
    }
  }
}
```



Bounce RSVP to ensure that the srlg-frr command takes effect:

```
# on PE-1:
configure exclusive
router "Base" {
    rsvp {
        admin-state disable
        commit
        admin-state enable
        commit
    }
}
```

## LSP configuration

Remove the admin group restriction from the LSP.

```
# on PE-1:
configure {
    router "Base" {
        mpls {
            lsp "LSP-PE-1-PE-10" {
                delete exclude-admin-group "red"
                propagate-admin-group false
            }
        }
    }
}
```

Now check the LSP path on PE-1 and verify that FRR protection is in place.

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-10" path detail

=====
MPLS LSP LSP-PE-1-PE-10 Path (Detail)
=====
Legend :
  @ - Detour Available          # - Detour In Use
  b - Bandwidth Protected      n - Node Protected
  s - Soft Preemption
  S - Strict                    L - Loose
  A - ABR                       + - Inherited
=====
-----
LSP LSP-PE-1-PE-10
Path empty_path
-----
LSP Name      : LSP-PE-1-PE-10
From          : 192.0.2.1
To            : 192.0.2.10
Admin State   : Up
Path Name     : empty_path
Path LSP ID   : 24652
Path Admin    : Up
Out Interface : 1/1/2
Path Up Time  : 0d 00:31:05
Retry Limit   : 0
Retry Attempt : 0
Oper State    : Up
Path Type     : Primary
Path Oper     : Up
Out Label     : 524287
Path Down Time : 0d 00:00:00
Retry Timer   : 30 sec
Next Retry In : 0 sec

---snip---

Adspec        : Disabled
PathCompMethod : local-cspf
MetricType    : igp
Oper Adspec   : Disabled
OperPathCompMethod : local-cspf
Oper MetricType : igp
```

```

Least Fill      : Disabled          Oper LeastFill   : Disabled
FRR             : Enabled           Oper FRR         : Enabled
FRR NodeProtect : Enabled           Oper FRR NP      : Enabled
FR Hop Limit    : 16                Oper FRHopLimit  : 16
FR Prop Admin Gr*: Disabled         Oper FRPropAdmGrp : Disabled
Propagate Adm Grp: Disabled         Oper Prop Adm Grp : Disabled
Inter-area      : True

---snip---

Include Groups  :                   Oper IncludeGroups:
None                                                    None
Exclude Groups  :                   Oper ExcludeGroups:
None                                                    None

Adaptive        : Enabled           Oper Metric       : 15
Preference      : n/a
Path Trans      : 3                 CSPF Queries      : 39
Failure Code    : noError
Failure Node    : n/a
Explicit Hops   :
  No Hops Specified
Actual Hops     :
  192.168.13.1(192.0.2.1) @ n      Record Label      : N/A
  -> 192.168.13.2(192.0.2.3) @ n   Record Label      : 524287
  -> 192.168.34.2(192.0.2.4) @ n   Record Label      : 524287
  -> 192.0.2.8(192.0.2.8) @ n     Record Label      : 524287
  -> 192.168.48.2 @ n             Record Label      : 524287
  -> 192.0.2.5(192.0.2.5) @      Record Label      : 524287
  -> 192.168.58.1 @              Record Label      : 524287
  -> 192.0.2.10(192.0.2.10)      Record Label      : 524287
  -> 192.168.105.2               Record Label      : 524287
Computed Hops   :
  192.168.13.1(S)
  -> 192.168.13.2(S)
  -> 192.168.34.2(SA)
  -> 192.0.2.10(L)
Resignal Eligible: False
Last Resignal   : n/a              CSPF Metric       : 15
Last MBB        :
  MBB Type       : ConfigChange     MBB State         : Success
  Ended At       : 05/12/2021 08:37:49 Old Metric        : 15
  Signaled BW    : 0 Mbps
  Fail Code      : noError
=====
* indicates that the corresponding row element may have been truncated.

```

On P-4, the SRLG configuration is checked as follows:

```

[/]
A:admin@P-4# show router if-attribute srlg-group

=====
Interface Srlg Groups
=====
Group Name          Group Value    Penalty Weight
-----
bundle-red          1              0
-----
No. of Groups: 1
=====

[/]

```

```
A:admin@P-4# show router mpls interface

=====
MPLS Interfaces
=====
Interface                               Port-id           Adm  Opr(V4/V6)  TE-
-----                               -
metric
-----
system                                  system            Up   Up/Down     None
  Admin Groups                          None
  SRLG Groups                            None
int-P-4-P-5                             1/1/1            Up   Up/Down     None
  Admin Groups                          None
  SRLG Groups                          bundle-red
int-P-4-P-6                             1/1/3            Up   Up/Down     None
  Admin Groups                          None
  SRLG Groups                            None
int-P-4-P-8                             1/2/1            Up   Up/Down     None
  Admin Groups                          red
  SRLG Groups                          bundle-red
int-P-4-PE-2                             1/1/2            Up   Up/Down     None
  Admin Groups                          None
  SRLG Groups                            None
int-P-4-PE-3                             1/1/4            Up   Up/Down     None
  Admin Groups                          None
  SRLG Groups                            None
-----
Interfaces : 6
=====
```

On PE-4, it is verified that the bypass tunnel is set up via P-6 rather than via P-5, as follows:

```
[/]
A:admin@P-4# show router mpls bypass-tunnel protected-lsp detail

=====
MPLS Bypass Tunnels (Detail)
=====
-----
bypass-node192.0.2.8-61443
-----
To           : 192.168.57.1           State           : Up
Out I/F      : 1/1/3                   Out Label      : 524285
Up Time     : 0d 00:03:12         Active Time    : n/a
Reserved BW  : 0 Kbps              Protected LSP Count : 1
Type        : Dynamic             Bypass Path Cost : 1020
Setup Priority : 7                  Hold Priority   : 0
Class Type   : 0
Exclude Node : None                Inter-Area     : False
Computed Hops :
  192.168.46.1(S)                  Egress Admin Groups : None
  -> 192.168.46.2(S)                Egress Admin Groups : None
  -> 192.168.67.2(S)                Egress Admin Groups : None
  -> 192.168.57.1(S)                Egress Admin Groups : None
Actual Hops  :
  192.168.46.1(192.0.2.4)           Record Label     : N/A
  -> 192.168.46.2(192.0.2.6)         Record Label     : 524285
  -> 192.168.67.2(192.0.2.7)         Record Label     : 524287
  -> 192.168.57.1(192.0.2.5)         Record Label     : 524286
Last Resignal :
Attempted At : n/a                 Resignal Reason  : n/a
Resignal Status: n/a               Reason           : n/a
```

```
Protected LSPs -  
LSP Name      : LSP-PE-1-PE-10::empty_path  
From          : 192.0.2.1          To          : 192.0.2.10  
Avoid Node/Hop : 192.0.2.8        Downstream Label : 524287  
Bandwidth     : 0 Kbps  
=====
```

## Conclusion

Inter-area TE P2P LSPs can be set up based on ERO expansion. With this feature, the head-end does a partial CSPF calculation to its local ABR. On receiving a PATH message with a loose hop ERO, this ABR does a partial or full CSPF calculation to the next ABR to reach the final destination.

FRR protection within the area is available. FRR node protection of the ABR is possible through an MBT on the PLR (node just upstream of the ABR) to the MP (node just downstream of the ABR) or through a dynamically signaled bypass tunnel on the PLR. Dynamic ABR node protection requires that the node ID of the MP node is propagated in the RESV message and that an XRO object is included in the bypass PATH message which makes it possible for the ABR to calculate a path to an MP node.

TE features such as BW, path prioritization, path pre-emption, and graceful shutdown are supported, as well as propagation of the session attribute with affinity along the LSP path (admin groups) and SRLG.

# LDP FEC to BGP Label Route Stitching

This chapter provides information about LDP FEC to BGP label route stitching.

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

## Applicability

This chapter is applicable to SR OS routers and was initially written for SR OS Release 13.0.R7. The MD-CLI in the current edition corresponds to SR OS Release 21.2.R1. Label Distribution Protocol (LDP) Forwarding Equivalence Class (FEC) to Border Gateway Protocol (BGP) label route stitching was first implemented in SR OS Release 8.0.

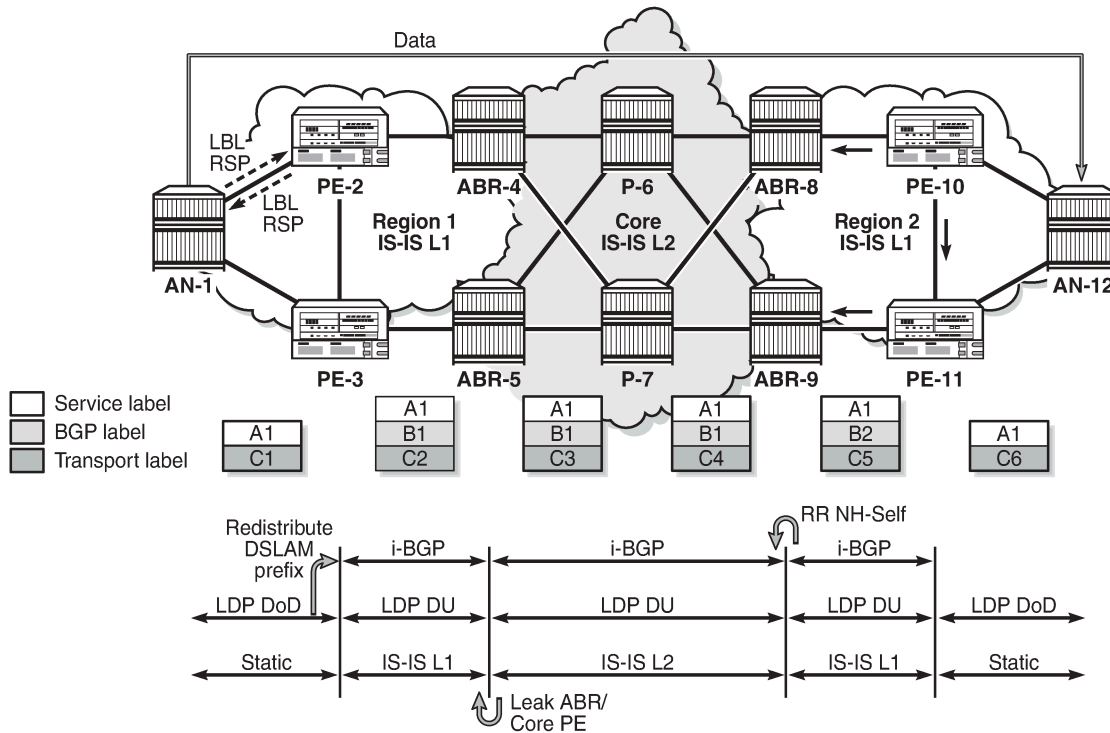
## Overview

Stitching of an LDP FEC to a BGP labeled route allows LDP-capable PE devices, such as Digital Subscriber Line Access Multiplexers (DSLAMs), to offer services to LDP-capable PE devices in other areas or domains without the need to support BGP labeled routes. This feature is used in a large network to provide services across multiple areas or Autonomous Systems (ASs).

When BGP is used to distribute a particular route, it can at the same time be used to distribute a Multi-Protocol Label Switching (MPLS) label that is mapped to that route. The label mapping information for a particular route is appended to the same BGP update message that is used to distribute the route. This is described in RFC 3701, *Carrying Label Information in BGPv4*.

[Figure 47: LDP FEC to BGP label route stitching](#) shows a network with a core area and regional areas. The components of the network are defined in the paragraphs that follow. For simplification, the control plane is displayed from right to left and the data plane from left to right.

Figure 47: LDP FEC to BGP label route stitching



25614

The Access Nodes (ANs) are DSLAMs that support LDP. In seamless MPLS networks, LDP Downstream-on-Demand (DoD) label advertisement can be used between the ANs and their next-hop PEs. Usually, MPLS routers implement LDP Downstream Unsolicited (DU) label distribution, advertising MPLS labels for all routes in their Routing Information Base (RIB). The ANs do not need to have LDP bindings for all prefixes in the network. The ANs will request the LDP labels they need. LDP DoD improves scalability in large networks.

BGP Route Reflectors (RRs) can be used to improve scalability. The RR can be any node; it does not need to be an Area Border Router (ABR) as in [Figure 47: LDP FEC to BGP label route stitching](#). If the RR is not in the forwarding path, it does not need to be capable of forwarding MPLS packets.

There are different areas for IS-IS: routers in the core network have level 2 (L2) capability, whereas the routers in the regional areas have level 1 (L1) capability and the ABRs have both. In each ABR, an IS-IS export policy is defined to leak the routes from the core to the regional networks.

Passing L1 routes (regional) into L2 (core) is inherent to IS-IS and cannot be controlled through policy. Passing L2 routes to L1 can be controlled through policy.

Only nodes within a regional area, and the ABR nodes in the same area, exchange LDP FECs. PE routers in a regional area learn the reachability of PE routers in other regional areas by way of RFC 3107 BGP labeled routes redistributed by the remote nodes.

The label stack contains three labels for packets sent in an Epipe service between the access nodes:

- The DSLAMs push a service label to the packets sent in the Epipe service. The service label remains unchanged end-to-end between the DSLAMs. The service label is popped by the remote DSLAM and is the inner label of the label stack.

- The BGP label is the middle label of the label stack and should be regarded as a transport label. The transport label stack contains two labels: BGP and LDP transport label. BGP labeled routes are not supported on the DSLAMs. The BGP label is pushed by the PE nearest to the local DSLAM and is swapped at the BGP next hop, which can be a BGP peer configured with next-hop-self or the PE that is the remote endpoint of the BGP tunnel. The BGP label is popped by the PE at the end of the BGP tunnel.
- The DSLAMs push an LDP transport label to the packets sent to the remote DSLAM. At the PE nearest to the local DSLAM, the LDP transport label is stitched to the BGP label. At the same time, that same PE pushes the LDP transport label to reach the BGP next hop. The LDP transport label is swapped in every Label Switching Router (LSR) and popped by the PE nearest to the remote DSLAM. That PE also pops the BGP label, which is stitched to the LDP transport label that is pushed to the packets sent to the remote DSLAM. This LDP label is the top label of the label stack.

When PE-2 is an ingress Label Edge Router (iLER) sending a service packet to the remote PE, PE-2 inserts the BGP route label to reach the remote PE and an LDP label to reach the next-hop router. In [Figure 47: LDP FEC to BGP label route stitching](#), this is the remote ABR because it has set next-hop-self (NH-Self).

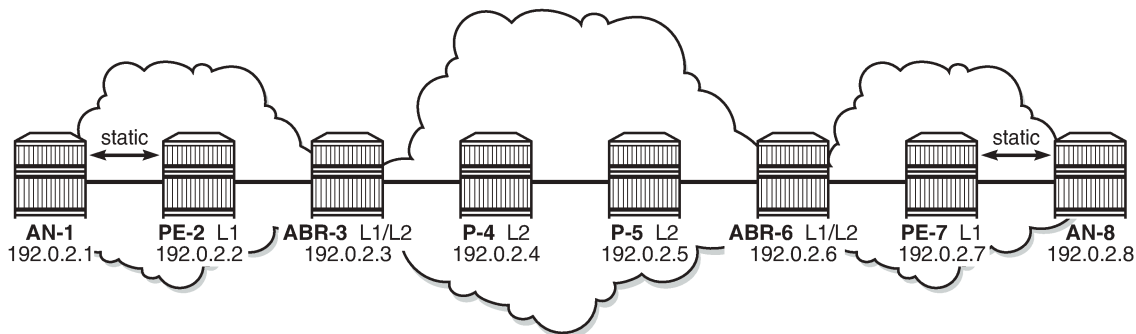
The access node AN-1, which is a DSLAM, can behave as a PE router for Epipe services. It will need to establish a pseudowire (PW) to a PE in a different regional area via LSR PE-2. In this case, PE-2 performs the following actions:

- Translates the LDP FEC it learned from AN-1 into a BGP labeled route and redistributes it using iBGP within its area. This is in addition to redistributing the FEC to its LDP neighbors in the same area.
- Translates the BGP labeled routes it learns through iBGP into an LDP FEC and redistributes it to its LDP neighbors in the same area. AN-1 requests the LDP FEC of the remote DSLAM (AN-12) using LDP DoD.
- When a data packet is received from AN-1 with destination AN-12, PE-2 swaps the LDP label into a BGP label and pushes the LDP label to reach the BGP next hop. When a data packet with destination AN-1 is received on PE-2 from the local ABR (ABR-4), the top transport label (LDP) is removed and the BGP label is swapped for the LDP label corresponding to AN-1.

## Configuration

[Figure 48: Example topology](#) shows the example topology that is used in this section. An Epipe will be established between the access nodes AN-1 and AN-8. PE-2 and PE-7 will stitch the LDP FECs to BGP label routes. In the regional areas, IS-IS L1 capability is used whereas in the core area, IS-IS L2 is used. The ABR nodes support both IS-IS L1 and L2 and export routes from L2 to L1. Static routes are configured between the access nodes and the next-hop PEs.

Figure 48: Example topology



25615

## Initial configuration



### Note:

In the example topology, all nodes are 7750 SRs, while the ANs should be access devices, such as DSLAMs. The limitation of this approach is that the ANs (SRs) in this setup can only request a label for the directly connected PE and not for their remote peer AN; however, DSLAMs do not have this limitation. Consequently, the Epipe service in this configuration will be operationally down because the transport tunnel is down.

All nodes have the following initial configuration:

- Cards, media dependent adapters (MDAs), ports
- Router interfaces



### Note:

The IP addresses for the link between node A and node B are in subnet 192.168.AB.0/0. The node with the lowest ID has IP address 192.168.AB.1/30 and the node with the highest ID has IP address 192.168.AB.2/30.

```
# on PE-2:
configure {
  router "Base" {
    interface "int-PE-2-ABR-3" {
      port 1/1/1
      ipv4 {
        primary {
          address 192.168.23.1
          prefix-length 30
        }
      }
    }
  }
  interface "int-PE-2-AN-1" {
    port 1/1/2
    ipv4 {
      primary {
        address 192.168.12.2
        prefix-length 30
      }
    }
  }
}
```



```

    }
    interface "system" {
        ipv4 {
            primary {
                address 192.0.2.2
                prefix-length 32
            }
        }
    }
}

```

- Static routes are configured between AN-1 and PE-2 and between PE-7 and AN-8:

```

# on AN-1:
configure {
    router "Base" {
        static-routes {
            route 0.0.0.0/0 route-type unicast {
                next-hop "192.168.12.2" {
                    admin-state enable
                }
            }
        }
    }
}

```

```

# on PE-2:
configure {
    router "Base" {
        static-routes {
            route 192.0.2.1/32 route-type unicast {
                next-hop "192.168.12.1" {
                    admin-state enable
                }
            }
        }
    }
}

```

- IS-IS (alternatively, OSPF could have been used)
  - PE-2 and PE-7 have L1 capability.

```

# on PE-2:
configure {
    router "Base" {
        isis 0 {
            admin-state enable
            level-capability 1
            area-address [49.0001]
            interface "int-PE-2-ABR-3" {
                interface-type point-to-point
            }
            interface "system" {
            }
        }
    }
}

```

- P-4 and P-5 have L2 capability.

- ABR-3 and ABR-6 have L1 capability on the interfaces toward the PE routers in the regional areas and L2 capability on the interfaces toward the P routers in the core area. A policy is applied to export the system IP addresses from L2 to L1:

```
# on ABR-3:
configure {
  router "Base" {
    isis 0 {
      admin-state enable
      export-policy ["export_L2_to_L1_policy"]
      area-address [49.0001]
      interface "int-ABR-3-P-4" {
        interface-type point-to-point
        level-capability 2
      }
      interface "int-ABR-3-PE-2" {
        interface-type point-to-point
        level-capability 1
      }
      interface "system" {
      }
    }
  }
}

configure {
  policy-options {
    prefix-list "system_IP_prefixes" {
      prefix 192.0.2.0/24 type longer {
      }
    }
  }
  policy-statement "export_L2_to_L1_policy" {
    entry 10 {
      from {
        level 2
        prefix-list ["system_IP_prefixes"]
        protocol {
          name [isis]
        }
      }
      action {
        action-type accept
      }
    }
  }
}
}
```

- LDP
  - Link LDP is enabled on all router interfaces on all nodes, including the ANs.
  - On PE-2 and PE-7, DoD is enabled in the session parameters for the peering sessions with the ANs:

```
# on PE-2:
configure {
  router "Base" {
    ldp {
      session-parameters {
        peer 192.0.2.1 {
          dod-label-distribution true
        }
      }
    }
  }
}
```

```

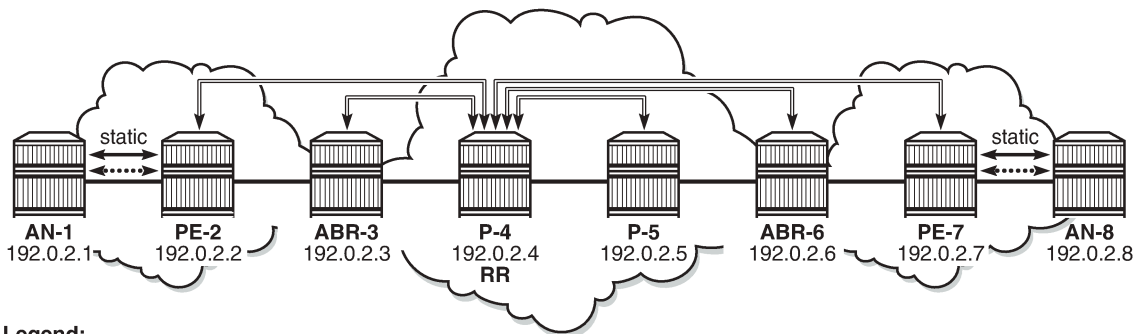
    }
    interface-parameters {
      interface "int-PE-2-ABR-3" {
        ipv4 {
        }
      }
      interface "int-PE-2-AN-1" {
        ipv4 {
        }
      }
    }
  }
}

```

## Configure BGP

BGP is configured on all nodes except the ANs. [Figure 49: BGP enabled with P-4 as RR](#) shows that P-4 is the RR.

Figure 49: BGP enabled with P-4 as RR



**Legend:**  
 ↔ iBGP  
 ↔ LDP DoD

25616

The initial BGP configuration on PE-2 is the following:

```

# on PE-2:
configure {
  router "Base" {
    autonomous-system 64496
    bgp {
      ebgp-default-reject-policy {
        import false
        export false
      }
      group "internal_group" {
        peer-as 64496
      }
      neighbor "192.0.2.4" {
        group "internal_group"
      }
    }
  }
}

```

```
}

```

The configuration is identical for ABR-3, P-5, ABR-6, and PE-7. The initial BGP configuration on the RR P-4 is:

```
# on P-4:
configure {
  router "Base" {
    autonomous-system 64496
    bgp {
      cluster {
        cluster-id 1.1.1.1
      }
      ebgp-default-reject-policy {
        import false
        export false
      }
      group "internal_group" {
        peer-as 64496
      }
      neighbor "192.0.2.2" {
        group "internal_group"
      }
      neighbor "192.0.2.3" {
        group "internal_group"
      }
      neighbor "192.0.2.5" {
        group "internal_group"
      }
      neighbor "192.0.2.6" {
        group "internal_group"
      }
      neighbor "192.0.2.7" {
        group "internal_group"
      }
    }
  }
}
```

This BGP configuration is incomplete: for labeled IPv4 BGP peering sessions, an additional address family will be configured on PE-2 and PE-7, as well as on RR P-4 for neighbors PE-2 and PE-7. The configuration is shown in the following section. The prefixes for AN-1 and AN-8 will be advertised in the labeled IPv4 BGP sessions only, not in IPv4 BGP sessions.

## Export policies for BGP and LDP

LDP FEC to BGP label route stitching is established by configuring separate tunnel table route export policies in both protocols. At the local next-hop PE, the LDP FEC of the local AN must be translated into a BGP label and at the remote PE, the BGP label must be translated into an LDP FEC.

An export policy for the export from LDP to BGP must be defined on the PE nodes.

```
# on PE-2:
configure {
  policy-options {
    prefix-list "local_AN_prefixes" {
      prefix 192.0.2.1/32 type exact {
      }
    }
  }
}
```

```
prefix-list "remote_AN_prefixes" {
  prefix 192.0.2.8/32 type exact {
  }
}
policy-statement "export_BGP_policy" {
  entry 10 {
    from {
      prefix-list ["local_AN_prefixes"]
      protocol {
        name [ldp]
      }
    }
    action {
      action-type accept
    }
  }
}
}
```

On PE-7, the policy statement is identical, but the prefix list is different.

This export policy must be applied in the **bgp** context: either in the general settings or per **group** or per **neighbor**.

```
# on PE-2:
configure {
  router "Base" {
    bgp {
      group "internal_group" {
        export {
          policy ["export_BGP_policy"]
        }
      }
    }
  }
}
```

In a similar way, BGP labels must be exported to LDP on the PE routers. The export policy is configured as follows, with a different prefix list:

```
# on PE-2:
configure {
  policy-options {
    prefix-list "remote_AN_prefixes" {
      prefix 192.0.2.8/32 type exact {
      }
    }
  }
  policy-statement "export_LDP_policy" {
    entry 10 {
      from {
        prefix-list ["remote_AN_prefixes"]
        protocol {
          name [bgp-label]
        }
      }
      action {
        action-type accept
      }
    }
  }
}
```

```
}
```

This export policy is applied in the **ldp** context, as follows:

```
# on PE-2:
configure {
  router "Base" {
    ldp {
      export-tunnel-table ["export_LDP_policy"]
    }
  }
}
```

## Advertise labels in BGP updates

BGP should evaluate the activated /32 LDP prefixes in the export policy. This needs to be configured on the endpoints of the BGP tunnel on PE-2 and PE-7, as follows:

```
# on PE-2 and PE-7:
configure {
  router "Base" {
    bgp {
      neighbor "192.0.2.4" {
        advertise-ldp-prefix true
        family {
          label-ipv4 true
        }
      }
    }
  }
}
```

On RR P-4, the family **label-ipv4** is enabled and the LDP prefix is advertised toward the clients PE-2 and PE-7, as follows.

```
# on RR P-4:
configure {
  router "Base" {
    bgp {
      neighbor "192.0.2.2" {
        advertise-ldp-prefix true
        family {
          label-ipv4 true
        }
      }
      neighbor "192.0.2.7" {
        advertise-ldp-prefix true
        family {
          label-ipv4 true
        }
      }
    }
  }
}
```

Configuring address family **label-ipv4 true** and **advertise-ldp-prefix true** implies that all activated /32 LDP FEC prefixes will be sent to the remote BGP peer as an RFC 3107 formatted label.

Configuring address family **label-ipv4 true** and **advertise-ldp-prefix false** or without `advertise-ldp-prefix` implies that only core IPv4 routes learned from the Route Table Manager (RTM) are advertised as RFC 3107 BGP labeled routes to this neighbor. No stitching of LDP FEC to the BGP labeled route will be performed for this neighbor, even if the same prefix was learned from LDP.

The BGP open messages contain address family AFI=1 and SAFI=1 between the RR and peers for address family IPv4, that is used for IPv4 unicast. See `Cap_Code MP-BGP`. Bytes 0x0 0x1 (AFI=1) 0x0 0x1 (SAFI=1).

```
# on ABR-3 (debugging is enabled in classic CLI):
A:ABR-3# show debug
debug
  router "Base"
  bgp
    open
    update
  exit
exit
exit
```

```
[/]
A:admin@ABR-3# show log log-id 2

=====
Event Log 2 log-name 2
=====
---snip---
2 2021/08/09 13:47:46.475 CEST MINOR: DEBUG #2001 Base BGP
"BGP: OPEN
Peer 1: 192.0.2.4 - Received BGP OPEN: Version 4
AS Num 64496: Holdtime 90: BGP_ID 192.0.2.4: Opt Length 20 (ExtOpt F)
Opt Para: Type CAPABILITY: Length = 18: Data:
  Cap_Code GRACEFUL-RESTART: Length 2
  Bytes: 0x0 0x78
Cap_Code MP-BGP: Length 4
Bytes: 0x0 0x1 0x0 0x1
  Cap_Code ROUTE-REFRESH: Length 0
  Cap_Code 4-OCTET-ASN: Length 4
  Bytes: 0x0 0x0 0xfb 0xf0
"
---snip---
```

Between peers that advertise the labels, AFI=1 and SAFI=4, the address family is labeled IPv4 unicast. The following BGP open message is seen on PE-2:

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

=====
Event Log 2 log-name 2
=====
Description : (Not Specified)
Memory Log contents [size=100 next event=11 (not wrapped)]

10 2021/08/09 13:51:13.041 CEST MINOR: DEBUG #2001 Base BGP
"BGP: OPEN
Peer 1: 192.0.2.4 - Received BGP OPEN: Version 4
AS Num 64496: Holdtime 90: BGP_ID 192.0.2.4: Opt Length 20 (ExtOpt F)
Opt Para: Type CAPABILITY: Length = 18: Data:
  Cap_Code GRACEFUL-RESTART: Length 2
  Bytes: 0x0 0x78
```

```

Cap_Code MP-BGP: Length 4
Bytes: 0x0 0x1 0x0 0x4
Cap_Code ROUTE-REFRESH: Length 0
Cap_Code 4-OCTET-ASN: Length 4
  Bytes: 0x0 0x0 0xfb 0xf0
"
---snip---

```

No BGP update messages are sent to ABR-3. Prefix 192.0.2.8 is advertised as a labeled IPv4 route from PE-7 to P-4 and forwarded by P-4 to its other labeled IPv4 client, PE-2, but it is not sent to BGP IPv4 clients, such as ABR-3.

The BGP update messages between labeled IPv4 peers contain label information, for example, for prefix 192.0.2.8/32. The address family is LBL-IPV4 (IPV4-Labeled) and the label is 524280. The following BGP update for prefix 192.0.2.8/32 is received on PE-2:

```

[/]
A:admin@PE-2# show log log-id 2

=====
Event Log 2 log-name 2
=====
Description : (Not Specified)
Memory Log contents [size=100  next event=14  (not wrapped)]

12 2021/08/09 13:52:10.511 CEST MINOR: DEBUG #2001 Base Peer 1: 192.0.2.4
"Peer 1: 192.0.2.4: UPDATE
Peer 1: 192.0.2.4 - Received BGP UPDATE:
  Withdrawn Length = 0
  Total Path Attr Length = 56
  Flag: 0x90 Type: 14 Len: 17 Multiprotocol Reachable NLRI:
    Address Family LBL-IPV4
    NextHop len 4 NextHop 192.0.2.7
    192.0.2.8/32 Label 524280
  Flag: 0x40 Type: 1 Len: 1 Origin: 0
  Flag: 0x40 Type: 2 Len: 0 AS Path:
  Flag: 0x80 Type: 4 Len: 4 MED: 1
  Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
  Flag: 0x80 Type: 9 Len: 4 Originator ID: 192.0.2.7
  Flag: 0x80 Type: 10 Len: 4 Cluster ID:
    1.1.1.1
"
---snip---

```

After applying the export policy from BGP to LDP, enabling the address family labeled IPv4 in BGP, and advertising labels for the LDP FEC prefixes, LDP will look for BGP route entries in the tunnel table. If a /32 BGP labeled route matches a prefix entry in the export policy, LDP originates an LDP FEC for this prefix, stitches it to the BGP labeled route, and redistributes the LDP FEC to its BGP neighbors. This can be shown on PE-7, as follows.

```

[/]
A:admin@PE-7# show router bgp inter-as-label

=====
BGP Inter-AS labels
Flags: B - entry has backup, P - entry is promoted
=====
NextHop                Received      Advertised    Label
                        Label         Label         Origin
-----
192.0.2.8              524287       524280       InternalLdp

```



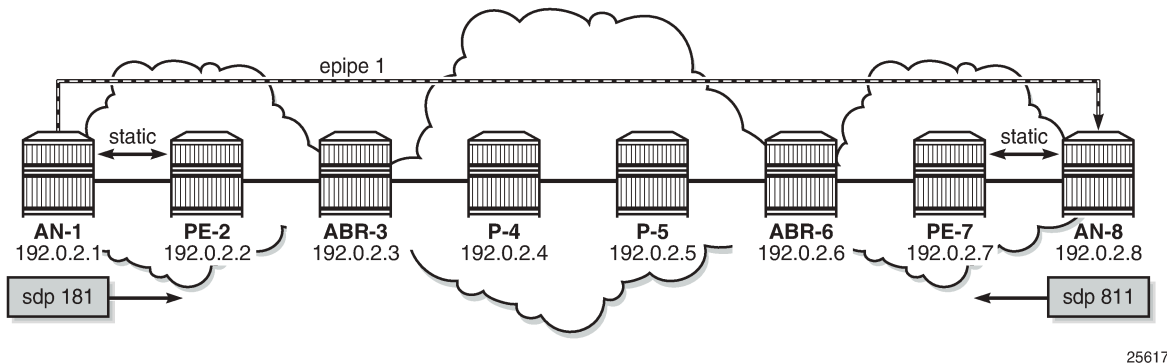
```
-----
Total Labels allocated: 1
-----
```

The label received from AN-8 is 524287. The label origin is **InternalLdp**. This LDP label is stitched to BGP label 524280 that will be advertised by PE-7 to its BGP labeled IPv4 peers: PE-7 advertises to RR P-4 and P-4 advertises this route to PE-2. Traffic sent from AN-1 toward AN-8 will be forwarded from PE-2 to its BGP NH PE-7 using BGP label 524280. In PE-7, the BGP label is stitched to LDP label 524287 that will be used to forward the packet to AN-8.

## Configure SDP and Epipe

An end-to-end Epipe service is established between AN-1 and AN-8, as shown in [Figure 50: End-to-end Epipe service](#).

Figure 50: End-to-end Epipe service



### Note:

In this setup, ANs are simulated by 7750 SRs. Due to this limitation, the SDP used by the Epipe service will not become operational. 7750 SR only supports single-hop DoD, which implies that AN-1 can only request a label for the LSR ID of the directly connected router, PE-2, not of remote nodes, such as AN-8. Similarly, AN-8 cannot request a label for AN-1. Therefore, it is not possible to have an LDP LSP between the ANs and the SDP will be down because there is no transport tunnel.

The SDP is configured on AN-1, as follows:

```
# on AN-1:
configure {
  service {
    sdp 181 {
      admin-state enable
      delivery-type mpls
      ldp true
      far-end {
        ip-address 192.0.2.8
      }
    }
  }
}
```

An Epipe is configured on AN-1, as follows:

```
# on AN-1:
configure {
  service {
    epipe "Epipe_1_name" {
      admin-state enable
      service-id 1
      customer "1"
      spoke-sdp 181:1 {
      }
      sap 1/2/1:1 {
      }
    }
  }
}
```

The configuration of the SDP and the Epipe on AN-8 is similar.

The SDP is down because there is no transport tunnel, which can be shown as follows:

```
[/]
A:admin@AN-1# show service sdp detail

=====
Services: Service Destination Points Details
=====
-----
Sdp Id 181 -192.0.2.8
-----
Description          : (Not Specified)
SDP Id               : 181
Admin Path MTU       : 0
Delivery             : MPLS
Far End              : 192.0.2.8
Oper Tunnel Far End  : 192.0.2.8
LSP Types            : LDP
SDP Source           : manual
Oper Path MTU       : 0
Tunnel Far End      :
Admin State          : Up
Signaling            : TLDP
Acct. Pol            : None
Oper State           : Down
Metric               : 0
Collect Stats        : Disabled
---snip---
Flags                : TranspTunnDown
---snip---
-----
Number of SDPs : 1
-----
=====
```

A targeted LDP session is established between AN-1 and AN-8, which can be shown as follows:

```
[/]
A:admin@AN-1# show router ldp session ipv4

=====
LDP IPv4 Sessions
=====
-----
Peer LDP Id      Adj Type  State      Msg Sent  Msg Recv  Up Time
-----
192.0.2.2:0     Link     Established 478       484       0d 00:20:59
192.0.2.8:0     Targeted Established 10        11        0d 00:00:23
-----
```

No. of IPv4 Sessions: 2  
=====

## LDP FEC resolution at PE-2 for traffic from AN-8 to AN-1

The following steps occur at PE-2 for the LDP FEC resolution for traffic from AN-1 toward AN-8. The situation is similar for PE-7.

1. After receiving an LDP label binding message for LDP FEC for the system address of AN-1 (192.0.2.1/32), PE-2 installs this prefix in the Label Forwarding Information Base (LFIB). PE-2 programs a push and a swap Next Hop Label Forwarding Entry (NHLFE) in the egress data path to forward packets to prefix 192.0.2.1/32.



**Note:**

PE-2 installs this LDP FEC in the LFIB only if there is an exact match of the prefix 192.0.2.1/32 in the routing table or a longest match of the prefix in the routing table, in case aggregate-prefix-match is configured on PE-2. The advertising LDP neighbor (AN-1) must be the next hop to reach the FEC prefix.

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

=====
LDP Bindings (IPv4 LSR ID 192.0.2.2)
              (IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
  (S) - Static           (M) - Multi-homed Secondary Support
  (B) - BGP Next Hop    (BU) - Alternate Next-hop for Fast Re-Route
  (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
  (C) - FEC resolved with class-based-forwarding
=====
LDP IPv4 Prefix Bindings (Active)
=====
Prefix          Op
IngLbl          EgrLbl
EgrNextHop      EgrIf/LspId
-----
192.0.2.1/32    Push
--             524287
192.168.12.1    1/1/2

192.0.2.1/32    Swap
524286          524287
192.168.12.1    1/1/2

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

2. PE-2 programs a tunnel entry for prefix 192.0.2.1/32 in the tunnel table.

```
[/]
A:admin@PE-2# show router tunnel-table 192.0.2.1/32

=====
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
-----
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
=====
```

3. PE-2 advertises a new FEC label binding for prefix 192.0.2.1/32 toward all its LDP neighbors. The result can be shown on ABR-3, as follows:

```
[/]
A:admin@ABR-3# show router ldp bindings prefixes prefix 192.0.2.1/32

=====
LDP Bindings (IPv4 LSR ID 192.0.2.3)
              (IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
=====
LDP IPv4 Prefix Bindings
=====
Prefix
Peer          FEC-Flags
IgrLbl        EgrLbl
EgrNextHop    EgrIntf/LspId
-----
192.0.2.1/32
192.0.2.2:0
  --          524286
  --          --
-----
No. of IPv4 Prefix Bindings: 1
=====
```

4. When BGP learns the LDP FEC via the tunnel table and the FEC prefix exists in the BGP route policy, PE-2 originates a BGP labeled route toward all its neighbors that have the advertise label for LDP FEC prefixes enabled. The following output shows the BGP labeled route to RR P-4 for prefix 192.0.2.1/32.

```
[/]
A:admin@PE-2# show router bgp routes label-ipv4 hunt

=====
BGP Router ID:192.0.2.2      AS:64496      Local AS:64496
=====
```

```

Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
               l - leaked, x - stale, > - best, b - backup, p - purge
Origin codes  : i - IGP, e - EGP, ? - incomplete

=====
BGP Routes
=====
-----
RIB In Entries
-----
---snip---
-----
RIB Out Entries
-----
Network       : 192.0.2.1/32
Nexthop       : 192.0.2.2
Path Id       : None
To            : 192.0.2.4
Res. Nexthop  : n/a
Local Pref.   : 100
Aggregator AS : None
Atomic Aggr.  : Not Atomic
AIGP Metric   : None
Connector     : None
Community     : No Community Members
Cluster       : No Cluster Members
Originator Id : None
IPv4 Label    : 524280
Lbl Allocation : NEXT-HOP
Origin        : IGP
AS-Path       : No As-Path
Route Tag     : 0
Neighbor-AS   : n/a
Orig Validation: NotFound
Source Class  : 0

Interface Name : NotAvailable
Aggregator     : None
MED            : 1
IGP Cost       : n/a

Peer Router Id : 192.0.2.4
Label Type     : SWAP

Dest Class    : 0

-----
Routes : 3
=====

```

## BGP labeled route resolution at PE-2 for traffic from AN-1 to AN-8

The following steps occur at PE-2 for the BGP labeled route resolution for traffic from AN-1 toward AN-8. The situation is similar for PE-7.

1. When there is an LDP LSP to the BGP neighbor advertising the route (PE-7) and PE-2 has received a BGP labeled route via iBGP for AN-8, PE-2 installs the prefix 192.0.2.8/32 in BGP. The LDP tunnel toward PE-7 is shown, then the BGP labeled IPv4 route toward AN-8, as advertised by PE-7.

```

[/]
A:admin@PE-2# show router tunnel-table 192.0.2.7

=====
IPv4 Tunnel Table (Router: Base)
=====
Destination      Owner      Encap TunnelId  Pref  Nexthop      Metric
  Color
-----
192.0.2.7/32     ldp       MPLS  65542      9    192.168.23.2  50

```

```

-----
Flags: B = BGP or MPLS backup hop available
      L = Loop-Free Alternate (LFA) hop available
      E = Inactive best-external BGP route
      k = RIB-API or Forwarding Policy backup hop
=====

```

```

[/]
A:admin@PE-2# show router bgp routes 192.0.2.8/32 label-ipv4
=====
BGP Router ID:192.0.2.2      AS:64496      Local AS:64496
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
              l - leaked, x - stale, > - best, b - backup, p - purge
Origin codes : i - IGP, e - EGP, ? - incomplete
=====
BGP Routes
=====
Flag  Network                               LocalPref  MED
      Nexthop (Router)                   Path-Id    IGP Cost
      As-Path                               Label
-----
u*>i  192.0.2.8/32                           100        1
      192.0.2.7                             None       50
      No As-Path                               524280
-----
Routes : 1
=====

```

The BGP label for traffic toward AN-8 is 524280. This is the middle label in the label stack. The next hop is PE-7.

2. PE-2 programs a swap NHLFE in the egress data path to forward packets to 192.0.2.8/32, as follows:

```

[/]
A:admin@PE-2# show router ldp bindings active prefixes prefix 192.0.2.8/32
=====
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
-----
192.0.2.8/32(B)                   Swap
524279                               524280

```

```
192.0.2.7                               LspId 65542
-----
No. of IPv4 Prefix Active Bindings: 1
=====
```

The (B) indicates that 192.0.2.8/32 is a BGP next hop. The ingress label is the LDP transport label from AN-1 for prefix 192.0.2.8/32. The LSP ID 65542 corresponds to the LDP LSP toward egress next-hop PE-7, as shown earlier in the tunnel table. The BGP egress label for traffic toward AN-8 is 524280.

3. PE-2 programs a tunnel table entry for 192.0.2.8/32.

```
[/]
A:admin@PE-2# show router tunnel-table
=====
IPv4 Tunnel Table (Router: Base)
=====
Destination      Owner      Encap TunnelId  Pref  Nexthop      Metric
  Color
-----
192.0.2.1/32     ldp       MPLS  65537     9    192.168.12.1  1
192.0.2.3/32     ldp       MPLS  65538     9    192.168.23.2 10
192.0.2.4/32     ldp       MPLS  65539     9    192.168.23.2 20
192.0.2.5/32     ldp       MPLS  65540     9    192.168.23.2 30
192.0.2.6/32     ldp       MPLS  65541     9    192.168.23.2 40
192.0.2.7/32     ldp       MPLS  65542     9    192.168.23.2 50
192.0.2.8/32     bgp       MPLS  262145    12    192.0.2.7    1000
-----
Flags: B = BGP or MPLS backup hop available
       L = Loop-Free Alternate (LFA) hop available
       E = Inactive best-external BGP route
       k = RIB-API or Forwarding Policy backup hop
=====
```

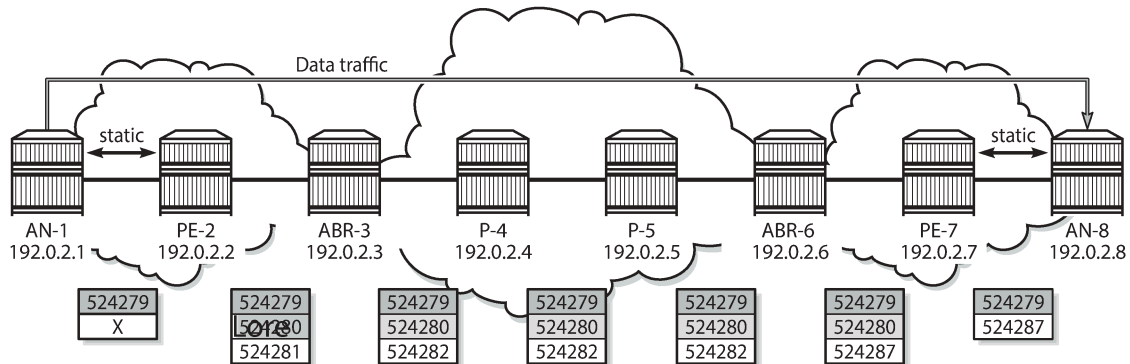
This is the only BGP tunnel in the tunnel table; all tunnels toward the other nodes are LDP tunnels. LDP routes have preference over BGP labeled routes, but there is no LDP route toward 192.0.2.8/32. Therefore, the BGP tunnel will be used for traffic destined to AN-8.

4. PE-2 advertises a new FEC label binding for prefix 192.0.2.8/32 toward AN-1. This is only done after AN-1 requests a label for prefix 192.0.2.8/32, because LDP DoD is enabled. This is possible if the ANs are DSLAMs, but not in this setup with SRs.

## Data plane overview for PE-2

[Figure 51: Label stacks for traffic from AN-1 to AN-8](#) shows the label stacks that are used for traffic from AN-1 to AN-8.

Figure 51: Label stacks for traffic from AN-1 to AN-8



25618



**Note:**

The LDP transport label that is pushed by AN-1 is not known because of the single-hop LDP DoD implementation in 7750 SR. AN-1 cannot request the LDP label for AN-8. Therefore, the LDP transport label is represented by "X".

The service label added for the Epipe on AN-1 for egress traffic to AN-8 is 524279. Ingress traffic on AN-1 has service label 524279. This can be shown as follows:

```
[/]
A:admin@AN-1# show service id 1 labels

=====
Martini Service Labels
=====
Svc Id      Sdp Binding      Type  I.Lbl      E.Lbl
-----
1           181:1           Spok  524279     524279
-----
Number of Bound SDPs : 1
=====
```

This service label remains unchanged end-to-end.

As shown earlier, the (middle) BGP label for traffic with destination AN-8 is 524280, as follows:

```
[/]
A:admin@PE-2# show router ldp bindings active prefixes prefix 192.0.2.8/32

=====
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
-----
192.0.2.8/32(B)       Swap
524279                524280
192.0.2.7             LspId 65542
-----
No. of IPv4 Prefix Active Bindings: 1
=====
  
```

The next hop is PE-7, which is the PE nearest to AN-8. The BGP label will not be swapped between PE-2 and PE-7 because there is no intermediate node that has set next-hop-self. An intermediate node with next-hop-self would become the next hop instead of PE-7. The BGP label is only added or removed by the next-hop PE.

On PE-2, when a service packet with destination AN-8 is received, the ingress LDP transport label X is swapped into BGP label 524280. To reach PE-7, which is the BGP next hop for traffic toward AN-8, another LDP transport label 524281 is pushed to the packet, as follows:

```

[/]
A:admin@PE-2# show router ldp bindings active prefixes prefix 192.0.2.7/32
=====
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
-----
192.0.2.7/32          Push
--                    524281
192.168.23.2          1/1/1
-----
192.0.2.7/32          Swap
524281                524281
192.168.23.2          1/1/1
-----
No. of IPv4 Prefix Active Bindings: 2
  
```

The next hop is ABR-3, where the ingress label 524281 is swapped to egress label 524282, as follows:

```
[/]
A:admin@ABR-3# show router ldp bindings active prefixes prefix 192.0.2.7/32

=====
LDP Bindings (IPv4 LSR ID 192.0.2.3)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
  (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
-----
192.0.2.7/32    Push
--             524282
192.168.34.2    1/1/1

192.0.2.7/32    Swap
524281          524282
192.168.34.2    1/1/1

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

In the subsequent LSRs, the transport label is swapped, as follows:

On P-4:

```
[/]
A:admin@P-4# show router ldp bindings active prefixes prefix 192.0.2.7/32
---snip---
192.0.2.7/32          Swap
524282              524282
192.168.45.2         1/1/1
---snip---
```

On P-5:

```
[/]
A:admin@P-5# show router ldp bindings active prefixes prefix 192.0.2.7/32
---snip---
192.0.2.7/32          Swap
524282              524282
192.168.56.2         1/1/1
```

---snip---

On ABR-6, the LDP label 524282 is swapped to 524287:

```
[/]
A:admin@ABR-6# show router ldp bindings active prefixes prefix 192.0.2.7/32
---snip---
192.0.2.7/32                               Swap
524282                                     524287
192.168.67.2                               1/1/1
---snip---
```

On PE-7, the LDP label 524287 is popped, as follows:

```
[/]
A:admin@PE-7# show router ldp bindings active prefixes prefix 192.0.2.7/32
---snip---
192.0.2.7/32                               Pop
524287                                       --
--                                           --
---snip---
```

The BGP label is also popped and mapped onto LDP label 524287 that will be pushed by PE-7 on packets toward AN-8.

```
[/]
A:admin@PE-7# show router ldp bindings active prefixes prefix 192.0.2.8/32

=====
LDP Bindings (IPv4 LSR ID 192.0.2.7)
(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
-----
192.0.2.8/32                         Push
--                                     524287
192.168.78.2                         1/1/1

192.0.2.8/32                         Swap
524281                                 524287
192.168.78.2                         1/1/1

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

## OAM

The following operations, administration, and maintenance (OAM) commands—in classic CLI—can be launched to validate an LDP FEC stitched to a BGP IPv4 labeled route and vice versa.

```
A:PE-2# oam lsp-ping bgp-label prefix 192.0.2.8/32
LSP-PING 192.0.2.8/32: 80 bytes MPLS payload
Seq=1, send from intf int-PE-2-ABR-3, reply from 192.0.2.8
    udp-data-len=32 ttl=255 rtt=7.18ms rc=4 (NoFECMapping)

---- LSP 192.0.2.8/32 PING Statistics ----
1 packets sent, 1 packets received, 0.00% packet loss
round-trip min = 7.18ms, avg = 7.18ms, max = 7.18ms, stddev = 0.000ms
```

In a similar way, LSP trace can validate LDP FEC to BGP label route stitching:

```
A:PE-2# oam lsp-trace bgp-label prefix 192.0.2.8/32
lsp-trace to 192.0.2.8/32: 0 hops min, 0 hops max, 104 byte packets
1 192.0.2.3 rtt=1.59ms rc=8(DSRtrMatchLabel)
2 192.0.2.4 rtt=2.99ms rc=8(DSRtrMatchLabel)
3 192.0.2.5 rtt=4.02ms rc=8(DSRtrMatchLabel)
4 192.0.2.6 rtt=5.40ms rc=8(DSRtrMatchLabel)
5 192.0.2.7 rtt=6.24ms rc=8(DSRtrMatchLabel) rsc=1
6 192.0.2.8 rtt=7.20ms rc=4(NoFECMapping) rsc=1
```

The detailed output includes the BGP label to LDP label mapping information at the PE:

```
A:PE-2# oam lsp-trace bgp-label prefix 192.0.2.8/32 detail
lsp-trace to 192.0.2.8/32: 0 hops min, 0 hops max, 104 byte packets
1 192.0.2.3 rtt=1.56ms rc=8(DSRtrMatchLabel)
2 192.0.2.4 rtt=2.86ms rc=8(DSRtrMatchLabel)
3 192.0.2.5 rtt=3.61ms rc=8(DSRtrMatchLabel)
4 192.0.2.6 rtt=5.23ms rc=8(DSRtrMatchLabel)
5 192.0.2.7 rtt=6.38ms rc=8(DSRtrMatchLabel) rsc=1
    DS 1: ipaddr=192.168.78.2 ifaddr=192.168.78.2 iftype=ipv4Numbered MRU=1560
    label[1]=524287 protocol=3(LDP)
6 192.0.2.8 rtt=7.12ms rc=4(NoFECMapping) rsc=1
```

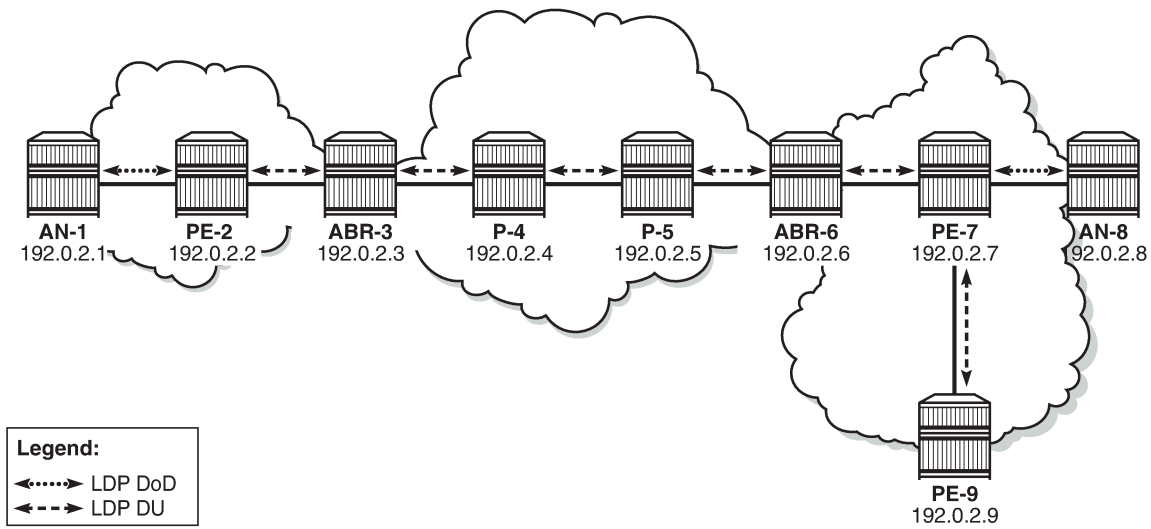
## Block BGP label bindings to LDP DU peers

On a PE, labeled BGP prefixes are exported to LDP to allow LDP DoD peers to request these labels. LDP DU peers will also get all labeled BGP prefixes if not explicitly blocked by an LDP export policy, based on prefix lists. This can result in a high administrative and operational effort in large networks.

Blocking BGP label bindings to LDP DU peers is less labor-intensive because per-peer export policies are re-evaluated on NH type change (such as from BGP to LDP or to "unresolved state"), not only on a configuration change.

**Figure 52: Block BGP label bindings to LDP DU peer PE-9** shows the extended topology used for this configuration. The additional PE router, PE-9, does not need to know the BGP labeled prefixes. LDP DU is used between PE-7 and PE-9.

Figure 52: Block BGP label bindings to LDP DU peer PE-9



25619

Blocking BGP label bindings to LDP DU peers can be achieved in two ways:

1. LDP export policy based on prefix list.
2. LDP export policy based on BGP NH type change. No prefix list is required.

To compare the two, both are described.

### LDP export policy based on prefix list

Before applying the policy to block BGP label bindings from PE-7 to PE-9, the LDP bindings on PE-9 for prefix 192.0.2.1 are the following:

```
[/]
A:admin@PE-9# show router ldp bindings prefixes prefix 192.0.2.1/32

=====
LDP Bindings (IPv4 LSR ID 192.0.2.9)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
=====
LDP IPv4 Prefix Bindings
=====
Prefix
Peer                               FEC-Flags
IgrLbl                             EgrLbl
EgrNextHop                         EgrIntf/LspId
-----
192.0.2.1/32
```

```

192.0.2.7:0
--
--
--
-----
No. of IPv4 Prefix Bindings: 1
=====

```

The following policy created on PE-7 is based on a prefix list that only contains the system address of the remote AN: 192.0.2.1.

```

# on PE-7:
configure {
  policy-options {
    prefix-list "remote_AN_prefixes" {
      prefix 192.0.2.1/32 type exact {
      }
    }
  }
  policy-statement "block_BGP_bindings_remote_AN_pol" {
    entry 10 {
      from {
        prefix-list ["remote_AN_prefixes"]
      }
      action {
        action-type reject
      }
    }
  }
}

```

The policy is applied on PE-7 in the LDP session-parameters context for peer 192.0.2.9.

```

# on PE-7:
configure {
  router "Base" {
    ldp {
      session-parameters {
        peer 192.0.2.9 {
          export-prefixes ["block_BGP_bindings_remote_AN_pol"]
        }
      }
    }
  }
}

```

After the policy is applied, there are no LDP bindings for prefix 192.0.2.1 on PE-9:

```

[/]
A:admin@PE-9# show router ldp bindings prefixes prefix 192.0.2.1/32

=====
LDP Bindings (IPv4 LSR ID 192.0.2.9)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC

```

```

=====
LDP IPv4 Prefix Bindings
=====
Prefix
Peer                               FEC-Flags
IgrLbl                             EgrLbl
EgrNextHop                         EgrIntf/LspId
-----
No Matching Entries Found
=====

```

The original situation is restored by removing the export prefixes in the LDP session-parameters context on PE-7.

```

# on PE-7:
configure {
  router "Base"
  ldp {
    session-parameters {
      peer 192.0.2.9 {
        delete export-prefixes ["block_BGP_bindings_remote_AN_pol"]
      }
    }
  }
}

```

```

[/]
A:admin@PE-9# show router ldp bindings prefixes prefix 192.0.2.1/32

=====
LDP Bindings (IPv4 LSR ID 192.0.2.9)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
=====
LDP IPv4 Prefix Bindings
=====
Prefix
Peer                               FEC-Flags
IgrLbl                             EgrLbl
EgrNextHop                         EgrIntf/LspId
-----
192.0.2.1/32
192.0.2.7:0
--
--                               524279
--                               --
-----
No. of IPv4 Prefix Bindings: 1
=====

```

## LDP export policy based on BGP NH type change

The **from protocol name [bgp]** argument will have a different meaning in the context of per-peer and targeted export policies. For those types of policies, policies are re-evaluated on NH type change; for example, from BGP to LDP or from LDP to "unresolved state". This requires less configuration because no prefix list needs to be specified. The following policy is configured on PE-7.

```
# on PE-7:
configure {
  policy-options {
    policy-statement "block_BGP_to_LDP_DU_policy" {
      entry 10 {
        from {
          protocol {
            name [bgp]
          }
        }
        action {
          action-type reject
        }
      }
    }
  }
}
```

The policy is applied in the LDP session-parameter context for peer 192.0.2.9.

```
# on PE-7:
configure {
  router "Base" {
    ldp {
      session-parameters {
        peer 192.0.2.9 {
          export-prefixes ["block_BGP_to_LDP_DU_policy"]
        }
      }
    }
  }
}
```

PE-7 will not send BGP label mapping information for prefix 192.0.2.1/32 to PE-9, or for any other prefix of a remote AN. In this example, AN-1 with prefix 192.0.2.1/32 is the only remote AN for PE-7.

```
[/]
A:admin@PE-9# show router ldp bindings prefixes prefix 192.0.2.1/32

=====
LDP Bindings (IPv4 LSR ID 192.0.2.9)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
=====
LDP IPv4 Prefix Bindings
```



```
=====
Prefix
Peer                               FEC-Flags
IgrLbl                             EgrLbl
EgrNextHop                         EgrIntf/LspId
-----
No Matching Entries Found
=====
```

## Conclusion

LDP FEC to BGP label route stitching allows LDP-capable PE devices, such as DSLAMs, to offer services to LDP-capable PE devices in other areas or domains without the need to support BGP labeled routes. This feature can be used in a seamless MPLS environment.

# LDP over RSVP Using OSPF as IGP

This chapter provides information about label distribution protocol (LDP) over resource reservation protocol with traffic engineering (RSVP-TE), also called LDP over RSVP, that uses RSVP label switched paths (LSPs) as a transport vehicle to carry the packets using LDP LSPs.

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

## Applicability

This chapter was initially written for SR OS Release 7.0.R5, but the configuration in this edition is based on MD-CLI in SR OS Release 21.2.R1. There are no prerequisites.

## Overview

Only user packets are tunneled over RSVP LSPs; targeted LDP (T-LDP) control messages are still sent unlabeled using the interior gateway protocol (IGP) shortest path. Because LDP does not have traffic engineering (TE), it can benefit from the RSVP-TE features. LDP fast reroute (FRR) is loopfree alternate (LFA), but with LDP over RSVP, it can use RSVP FRR detour or bypass tunnels.

The main advantage of LDP over RSVP is seen in large networks. A full mesh of intra-area RSVP LSPs between PE nodes (which in some cases is not scalable) is not needed anymore. While a label edge router (LER) may not have that many tunnels, any transit node may have thousands of LSPs, and if each transit node also has to deal with detour tunnels or bypass tunnels, this number can make the label switching router (LSR) overly burdened.

LDP over RSVP can be configured in an intra-area domain and an inter-area domain. Any router in an area can be a stitching point for LDP over RSVP. LDP over RSVP introduces a tunnel-in-tunnel tunnel type (in addition to the existing LDP tunnel type and RSVP tunnel type). If multiple tunnel types match the destination PE forwarding equivalence class (FEC) lookup, LDP prefers an LDP tunnel over an LDP over RSVP tunnel by default.

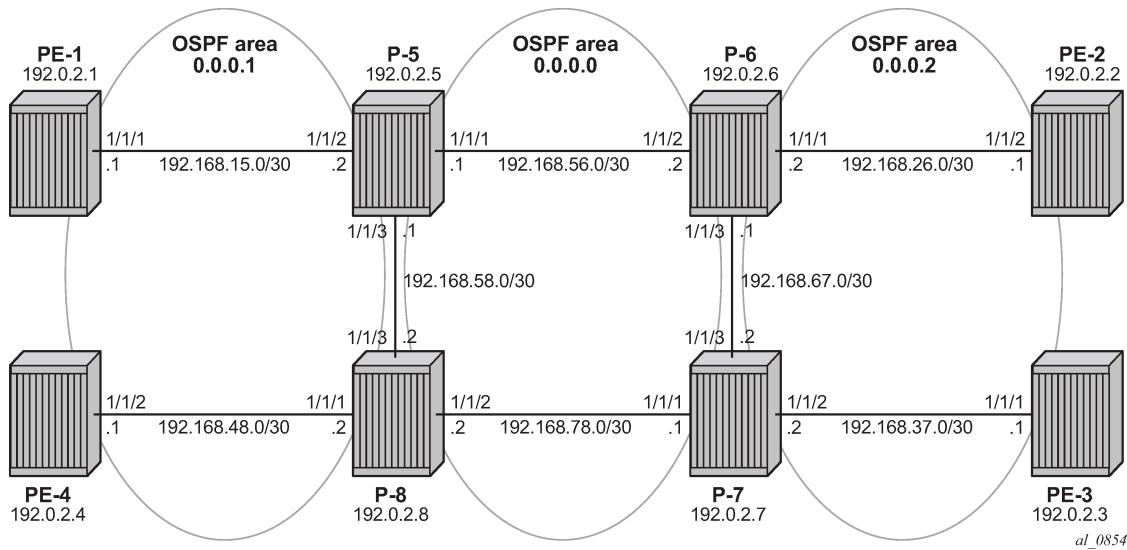
First, it is important to understand how LDP FEC resolution is working (with LDP over RSVP enabled). A more detailed description can be found later on in this chapter. The ingress LER receives an LDP label message including a FEC with prefix **P** and label **L** from a peer by a T-LDP session. LDP tries to resolve prefix **P** by performing a lookup in the Routing Table Manager (RTM). The result of this is a Next Hop (NH) to the destination PE, either an intra-area PE (intra-area context) or an Area Border Router (ABR) (inter-area context). When the NH matches the targeted LDP peer, LDP performs a second lookup for that NH in the tunnel table which returns a user configured RSVP LSP with the best metric. If there are multiple configured RSVP LSPs with the best metric, LDP selects the first available RSVP LSP. If all user configured RSVP LSPs are down, no more action is taken. If the user did not configure any RSVP

LSPs under the T-LDP context, the lookup in the tunnel table returns the first available RSVP LSP which terminates on the ABR (inter-area) or intra-area PE with the lowest metric.

If the lookup in the tunnel table results in no RSVP LSP, the system can fall back to link-level interface LDP (iLDP). In that way, it is possible that the NH is reachable using iLDP. Accordingly, the egress label is then installed on the ingress LER.

Figure 53: Initial example topology shows the example topology with four PE routers and four P routers.

Figure 53: Initial example topology



OSPF area 0.0.0.1 and OSPF area 0.0.0.2 are two metro areas, connected to each other via a core area, represented by OSPF backbone area (area 0.0.0.0). Therefore, P-5, P-6, P-7, and P-8 are all acting as area border routers (ABRs). LDP over RSVP principles will be shown for intra-area PE communication (between PE-1 and PE-4) and inter-area communication (between PE-1 and PE-2).

## Configuration

### 1. Configuring the IP/MPLS network.

The system addresses and IP interface addresses are configured according to Figure 53: Initial example topology. An interior gateway protocol (IGP) is needed to distribute routing information on all routers. In this case, the IGP is Open Shortest Path First (OSPF) using the backbone area 0.0.0.0 in the core and normal areas (area 0.0.0.1 and area 0.0.0.2) in the two metro regions, connected toward the backbone area via ABRs. A configuration example is shown for PE-1 and P-5. A similar configuration can be derived for the other P and PE nodes.

```
# on PE-1:
configure {
  router "Base" {
    interface "int-PE-1-P-5" {
      port 1/1/1
      ipv4 {
        primary {
          address 192.168.15.1
          prefix-length 30
        }
      }
    }
  }
}
```

```
    }
  }
}
interface "system" {
  ipv4 {
    primary {
      address 192.0.2.1
      prefix-length 32
    }
  }
}
ospf 0 {
  admin-state enable
  traffic-engineering
  area 0.0.0.1 {
    interface "int-PE-1-P-5" {
      interface-type point-to-point
    }
    interface "system" {
    }
  }
}
```

```
# on P-5:
configure {
  router "Base" {
    interface "int-P-5-P-6" {
      port 1/1/1
      ipv4 {
        primary {
          address 192.168.56.1
          prefix-length 30
        }
      }
    }
    interface "int-P-5-P-8" {
      port 1/1/3
      ipv4 {
        primary {
          address 192.168.58.1
          prefix-length 30
        }
      }
    }
    interface "int-P-5-PE-1" {
      port 1/1/2
      ipv4 {
        primary {
          address 192.168.15.2
          prefix-length 30
        }
      }
    }
    interface "system" {
      ipv4 {
        primary {
          address 192.0.2.5
          prefix-length 32
        }
      }
    }
  }
  ospf 0 {
    admin-state enable
  }
}
```

```

traffic-engineering
area 0.0.0.0 {
    interface "int-P-5-P-6" {
        interface-type point-to-point
    }
    interface "int-P-5-P-8" {
        interface-type point-to-point
    }
    interface "system" {
    }
}
area 0.0.0.1 {
    interface "int-P-5-PE-1" {
        interface-type point-to-point
    }
}
}

```

Because Fast Reroute (FRR) will be enabled on the RSVP LSPs in the core area, Traffic Engineering (TE) is needed on the IGP. By doing this, OSPF will generate opaque link state advertisements (LSAs) which are collected in a Traffic Engineering Database (TED), separate from the traditional OSPF topology database. OSPF interfaces are set up as type point-to-point to improve convergence, no Designated Router/Backup Designated Router (DR/BDR) election process is performed. Convergence is beyond the scope of this chapter.

On all nodes originating and terminating a T-LDP session, an explicit **ldp-over-rsvp** parameter must be configured to enable this OSPF instance for LDP over RSVP, as follows:

```

# on PE-1, PE-2, PE-3, PE-4, P-5, P-6, P-7, P-8:
configure {
    router "Base" {
        ospf 0 {
            ldp-over-rsvp true
        }
    }
}

```

To verify that OSPF neighbors are up (state: Full), the **show router ospf neighbor** command is executed. To check if IP interface addresses/subnets are known on all PEs, **show router route-table** or **show router fib <IOM-card-slot>** displays the content of the forwarding information base (FIB).

```

[/]
A:admin@PE-1# show router ospf neighbor
=====
Rtr Base OSPFv2 Instance 0 Neighbors
=====
Interface-Name          Rtr Id          State    Pri  RetxQ  TTL
Area-Id
-----
int-PE-1-P-5           192.0.2.5      Full     1    0      33
0.0.0.1
-----
No. of Neighbors: 1
=====

```

```

[/]
A:admin@PE-1# show router route-table
=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]          Type    Proto    Age    Pref
-----

```

```

Next Hop[Interface Name]                               Metric
-----
192.0.2.1/32                                           Local Local 00h02m37s 0
  system
192.0.2.2/32                                           Remote OSPF 00h01m15s 10
  192.168.15.2
192.0.2.3/32                                           Remote OSPF 00h00m57s 10
  192.168.15.2
192.0.2.4/32                                           Remote OSPF 00h00m35s 10
  192.168.15.2
192.0.2.5/32                                           Remote OSPF 00h01m31s 10
  192.168.15.2
192.0.2.6/32                                           Remote OSPF 00h01m17s 10
  192.168.15.2
192.0.2.7/32                                           Remote OSPF 00h00m57s 10
  192.168.15.2
192.0.2.8/32                                           Remote OSPF 00h00m37s 10
  192.168.15.2
192.168.15.0/30                                         Local Local 00h02m37s 0
  int-PE-1-P-5
192.168.26.0/30                                         Remote OSPF 00h01m15s 10
  192.168.15.2
192.168.37.0/30                                         Remote OSPF 00h00m57s 10
  192.168.15.2
192.168.48.0/30                                         Remote OSPF 00h00m35s 10
  192.168.15.2
192.168.56.0/30                                         Remote OSPF 00h01m31s 10
  192.168.15.2
192.168.58.0/30                                         Remote OSPF 00h01m31s 10
  192.168.15.2
192.168.67.0/30                                         Remote OSPF 00h01m11s 10
  192.168.15.2
192.168.78.0/30                                         Remote OSPF 00h00m37s 10
  192.168.15.2
-----
No. of Routes: 16
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
=====

```

```

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

=====
FIB Display
=====
Prefix [Flags]                                         Protocol
  NextHop
-----
192.0.2.1/32                                           LOCAL
  192.0.2.1 (system)
192.0.2.2/32                                           OSPF
  192.168.15.2 (int-PE-1-P-5)
192.0.2.3/32                                           OSPF
  192.168.15.2 (int-PE-1-P-5)
192.0.2.4/32                                           OSPF
  192.168.15.2 (int-PE-1-P-5)
192.0.2.5/32                                           OSPF
  192.168.15.2 (int-PE-1-P-5)
192.0.2.6/32                                           OSPF
  192.168.15.2 (int-PE-1-P-5)

```

```

192.0.2.7/32                                OSPF
  192.168.15.2 (int-PE-1-P-5)
192.0.2.8/32                                OSPF
  192.168.15.2 (int-PE-1-P-5)
192.168.15.0/30                             LOCAL
  192.168.15.0 (int-PE-1-P-5)
192.168.26.0/30                             OSPF
  192.168.15.2 (int-PE-1-P-5)
192.168.37.0/30                             OSPF
  192.168.15.2 (int-PE-1-P-5)
192.168.48.0/30                             OSPF
  192.168.15.2 (int-PE-1-P-5)
192.168.56.0/30                             OSPF
  192.168.15.2 (int-PE-1-P-5)
192.168.58.0/30                             OSPF
  192.168.15.2 (int-PE-1-P-5)
192.168.67.0/30                             OSPF
  192.168.15.2 (int-PE-1-P-5)
192.168.78.0/30                             OSPF
  192.168.15.2 (int-PE-1-P-5)
-----
Total Entries : 16
=====

```

The next step in the process of setting up the IP/MPLS network, is enabling the IP interfaces in the **mpls** and **rsvp** context on all involved nodes (PE and P nodes). Explicit enabling of **mpls** and **rsvp** context is done by the **admin-state enable** command. The following output displays the MPLS/RSVP configuration for PE-1.

```

# on PE-1:
configure {
  router "Base" {
    mpls {
      admin-state enable
      interface "int-PE-1-P-5" {
      }
    }
    rsvp {
      admin-state enable
      interface "int-PE-1-P-5" {
      }
    }
  }
}

```

## 2. Configure the RSVP LSPs.

In both metro areas, RSVP LSPs are set up from all PEs toward the ABRs, no intra-area PE-PE RSVP LSPs are needed. In the core/backbone, a full RSVP LSP mesh is required. To simplify the RSVP LSP configuration, no FRR is enabled on the RSVP LSPs in the metro areas, only in the backbone area. All RSVP paths are configured as **strict paths**. As an example, the configuration for PE-1 and P-5 is as follows:

```

# on PE-1:
configure {
  router "Base" {
    mpls {
      admin-state enable
      path "path-PE-1-P-5" {
        admin-state enable
        hop 10 {
          ip-address 192.168.15.2
        }
      }
    }
  }
}

```

```
        type strict
    }
}
path "path-PE-1-P-5-P-8" {
    admin-state enable
    hop 10 {
        ip-address 192.168.15.2
        type strict
    }
    hop 20 {
        ip-address 192.168.58.2
        type strict
    }
}
lsp "LSP-PE-1-P-5" {
    admin-state enable
    to 192.0.2.5
    type p2p-rsvp
    primary "path-PE-1-P-5" {
    }
}
lsp "LSP-PE-1-P-8" {
    admin-state enable
    to 192.0.2.8
    type p2p-rsvp
    primary "path-PE-1-P-5-P-8" {
    }
}
}
```

```
# on P-5:
configure {
    router "Base" {
        mpls {
            admin-state enable
            path "path-P-5-P-6" {
                admin-state enable
                hop 10 {
                    ip-address 192.168.56.2
                    type strict
                }
            }
            path "path-P-5-P-8" {
                admin-state enable
                hop 10 {
                    ip-address 192.168.58.2
                    type strict
                }
            }
            path "path-P-5-P-6-P-7" {
                admin-state enable
                hop 10 {
                    ip-address 192.168.56.2
                    type strict
                }
                hop 20 {
                    ip-address 192.168.67.2
                    type strict
                }
            }
            path "path-P-5-PE-1" {
                admin-state enable
                hop 10 {
                    ip-address 192.168.15.1
```



```
        type strict
    }
}
path "path-P-5-P-8-PE-4" {
    admin-state enable
    hop 10 {
        ip-address 192.168.58.2
        type strict
    }
    hop 20 {
        ip-address 192.168.48.1
        type strict
    }
}
lsp "LSP-P-5-PE-1" {
    admin-state enable
    to 192.0.2.1
    type p2p-rsvp
    primary "path-P-5-PE-1" {
    }
}
lsp "LSP-P-5-PE-4" {
    admin-state enable
    to 192.0.2.4
    type p2p-rsvp
    primary "path-P-5-P-8-PE-4" {
    }
}
lsp "LSP-P-5-P-6" {
    admin-state enable
    to 192.0.2.6
    path-computation-method local-cspf
    fast-reroute {
        frr-method facility
    }
    type p2p-rsvp
    primary "path-P-5-P-6" {
    }
}
lsp "LSP-P-5-P-7" {
    admin-state enable
    to 192.0.2.7
    path-computation-method local-cspf
    fast-reroute {
        frr-method facility
    }
    type p2p-rsvp
    primary "path-P-5-P-6-P-7" {
    }
}
lsp "LSP-P-5-P-8" {
    admin-state enable
    to 192.0.2.8
    path-computation-method local-cspf
    fast-reroute {
        frr-method facility
    }
    type p2p-rsvp
    primary "path-P-5-P-8" {
    }
}
}
```

The following command on PE-1 lists the RSVP LSPs:

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

=====
MPLS LSPs (Originating)
=====
LSP Name          Tun   Fastfail  Adm  Opr
  To              Id     Config
-----
LSP-PE-1-P-5     1      No        Up   Up
  192.0.2.5
LSP-PE-1-P-8     2      No        Up   Up
  192.0.2.8
-----
LSPs : 2
=====
```

The following command on PE-1 shows the tunnel table for the RSVP signaling protocol. By default, RSVP LSPs have preference 7.

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

=====
IPv4 Tunnel Table (Router: Base)
=====
Destination      Owner   Encap TunnelId  Pref  Nexthop      Metric
  Color
-----
192.0.2.5/32     rsvp   MPLS  1           7     192.168.15.2 16777215
192.0.2.8/32     rsvp   MPLS  2           7     192.168.15.2 16777215
-----
Flags: B = BGP or MPLS backup hop available
       L = Loop-Free Alternate (LFA) hop available
       E = Inactive best-external BGP route
       k = RIB-API or Forwarding Policy backup hop
=====
```

On ABR P-5:

```
[/]
A:admin@P-5# show router mpls lsp

=====
MPLS LSPs (Originating)
=====
LSP Name          Tun   Fastfail  Adm  Opr
  To              Id     Config
-----
LSP-P-5-P-6     1      Yes        Up   Up
  192.0.2.6
LSP-P-5-P-7     2      Yes        Up   Up
  192.0.2.7
LSP-P-5-P-8     3      Yes        Up   Up
  192.0.2.8
LSP-P-5-PE-1    4      No         Up   Up
  192.0.2.1
LSP-P-5-PE-4    5      No         Up   Up
  192.0.2.4
-----
```

```
-----
LSPs : 5
=====
```

```
[/]
A:admin@P-5# show router tunnel-table
```

```
=====
IPv4 Tunnel Table (Router: Base)
=====
```

Destination Color	Owner	Encap	TunnelId	Pref	Nexthop	Metric
192.0.2.1/32	rsvp	MPLS	4	7	192.168.15.1	16777215
192.0.2.4/32	rsvp	MPLS	5	7	192.168.58.2	16777215
192.0.2.6/32 [B]	rsvp	MPLS	1	7	192.168.56.2	10
192.0.2.7/32 [B]	rsvp	MPLS	2	7	192.168.56.2	20
192.0.2.8/32 [B]	rsvp	MPLS	3	7	192.168.58.2	10

```
-----
Flags: B = BGP or MPLS backup hop available
       L = Loop-Free Alternate (LFA) hop available
       E = Inactive best-external BGP route
       k = RIB-API or Forwarding Policy backup hop
=====
```

By default, the metric for strict LSPs configured without Constrained Shortest Path First (CSPF) (RSVP LSPs in metro areas) is infinite (value = 16777215). The LSP metric for CSPF LSPs (RSVP LSPs in the core area) follows the IGP cost. LSP metrics can be explicitly set on the LSP level, see also in the [Additional topics](#) section.

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-P-5" {
        metric ?

metric <number>
<number>          - <0..16777215>
Dynamic Default   - 0

LSP metric that forces to a constant value
```

Whenever an RSVP LSP comes up, it is by default eligible for LDP over RSVP, meaning that RSVP signals to the relevant IGP (OSPF in this case) that the LSP should be included in the IGP Shortest Path First (SPF) run. The destination of the LSP (192.0.2.5) is considered as a potential endpoint in the Forwarding Equivalence Class (FEC) resolution. With the **info detail** command, all default settings of a context are shown, but the output is filtered for **ldp-over-rsvp**, as follows:

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-P-5" {
        info detail | match ldp-over-rsvp

[ex:/configure router "Base" mpls lsp "LSP-PE-1-P-5"]
A:admin@PE-1# info detail | match ldp-over-rsvp
```

ldp-over-rsvp true

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-P-5" detail

=====
MPLS LSPs (Originating) (Detail)
=====
Legend :
+ - Inherited
=====
-----
Type : Originating
-----
LSP Name      : LSP-PE-1-P-5
LSP Type      : RegularLsp           LSP Tunnel ID      : 1
LSP Index     : 1                   TTM Tunnel Id      : 1
From          : 192.0.2.1
To            : 192.0.2.5
Adm State     : Up                   Oper State          : Up
LSP Up Time   : 0d 00:05:42          LSP Down Time      : 0d 00:00:00
Transitions   : 1                   Path Changes        : 1
Retry Limit   : 0                   Retry Timer         : 30 sec
Signaling     : RSVP                 Resv. Style         : SE
Hop Limit     : 255                 Negotiated MTU     : 1564
Adaptive      : Enabled              ClassType           : 0
FastReroute   : Disabled            Oper FR             : Disabled
PathCompMethod : none                ADSPEC              : Disabled
FallbkPathComp : not-applicable
Metric        : N/A
Load Bal Wt   : N/A                  ClassForwarding    : Disabled
Include Grps  :                      Exclude Grps        :
None
Least Fill    : Disabled
BFD Template  : None                 BFD Ping Intvl     : 60
BFD Enable    : False                BFD Failure-action  : None
WaitForUpTimer : 4

Revert Timer  : Disabled              Next Revert In     : N/A
Entropy Label : Enabled+              Oper Entropy Label : Enabled
Negotiated EL : Disabled
Auto BW       : Disabled
LdpOverRsvp : Enabled
VprnAutoBind : Enabled
IGP Shortcut  : Enabled              BGP Shortcut       : Enabled
IGP LFA       : Disabled            IGP Rel Metric     : Disabled
BGPTransTun  : Enabled
Oper Metric   : 16777215
Prop Adm Grp  : Disabled
PCE Report    : Disabled+
PCE Control   : Disabled
Path Profile  : None
Admin Tags    : None
Lsp Self Ping : Disabled+              Self Ping Timeouts : 0
SelfPingOAMFail* : 0

Primary(a)    : path-PE-1-P-5
Up Time       : 0d 00:05:42
Bandwidth     : 0 Mbps
=====
* indicates that the corresponding row element may have been truncated.
```

The following command makes a specific RSVP LSP ineligible for LDP over RSVP:

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-P-5" {
        ldp-over-rsvp false
      }
    }
  }
}
```

### 3. Create T-LDP sessions according to RSVP LSPs.

It is a must that when configuring an RSVP LSP eligible for LDP over RSVP, also a T-LDP session is initiated. This must be done on all PE and P nodes.

```
# on PE-1:
configure {
  router "Base" {
    ldp {
      targeted-session {
        peer 192.0.2.5 {
        }
        peer 192.0.2.8 {
        }
      }
    }
  }
}
```

```
[/]
A:admin@PE-1# show router ldp session family ipv4
```

#### LDP IPv4 Sessions

Peer LDP Id	Adj Type	State	Msg Sent	Msg Recv	Up Time
192.0.2.5:0	<b>Targeted</b>	Established	25	27	0d 00:01:47
192.0.2.8:0	<b>Targeted</b>	Established	14	16	0d 00:00:58

No. of IPv4 Sessions: 2

### 4. Enable LDP over RSVP.

This is done using the **tunneling** keyword inside the T-LDP session context. This configuration is needed on all PE and ABR nodes.

```
# on PE-1:
configure {
  router "Base" {
    ldp {
      targeted-session {
        peer 192.0.2.5 {
          tunneling {
          }
        }
        peer 192.0.2.8 {
          tunneling {
          }
        }
      }
    }
  }
}
```

As a result of the **tunneling** command, the LDP over RSVP process of FEC resolving is initiated. As already stated in the introduction, FEC resolution is a three-step process. First run an SPF calculation to the destination, then select an endpoint close to that destination followed by a tunnel to that endpoint. The next two steps go more into detail on this FEC resolution process. Step 5 will handle inter-area FEC resolving and Step 6 will handle intra-area FEC resolving.

**5. Inter-area FEC resolving (ingress LER is PE-1, egress LER is PE-2)**

**a. Verification endpoint nodes and associated RSVP tunnels.**

The first thing to do in the inter-area FEC resolving process is for PE-1 to perform an SPF calculation toward PE-2 with the purpose to search for an eligible endpoint, as close as possible to PE-2. An endpoint is eligible when:

- a T-LDP session exists between PE-1 and the endpoint node
- tunneling is configured on the endpoint node
- PE-1 received a label for the destination FEC from the endpoint node
- and an RSVP LSP that can be used for LDP over RSVP exists between PE-1 and the endpoint node

Endpoint node in OSPF area 1 can be either P-5 or P-8 (only those nodes have a T-LDP session toward PE-1). The following command shows that P-5 is the endpoint node (EgrNextHop). The RSVP LSP that is used has ID 1.

```
[/]
A:admin@PE-1# show router ldp bindings active prefixes prefix 192.0.2.2/32

=====
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 (0) - SR-OSPF Next Hop
  (C) - FEC resolved with class-based-forwarding
=====
LDP IPv4 Prefix Bindings (Active)
=====
Prefix                               Op
IngLbl                               EgrLbl
EgrNextHop                           EgrIf/LspId
-----
192.0.2.2/32                          Push
--                                     524268
192.0.2.5                             LspId 1

192.0.2.2/32                          Swap
524281                                 524268
192.0.2.5                             LspId 1
-----
No. of IPv4 Prefix Active Bindings: 2
```

The following command shows that RSVP LSP with ID 1 is LSP-PE-1-P-5:

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

=====
MPLS LSPs (Originating)
=====
LSP Name          Tun   Fastfail  Adm  Opr
To                Id     Config
-----
LSP-PE-1-P-5      1     No      Up Up
  192.0.2.5
LSP-PE-1-P-8      2      No        Up   Up
  192.0.2.8
-----
LSPs : 2
=====
```

The following command shows that the RSVP tunnel toward P-5 has tunnel ID 1 and next-hop 192.168.15.2:

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

=====
IPv4 Tunnel Table (Router: Base)
=====
Destination      Owner   Encap TunnelId  Pref  Nexthop      Metric
Color
-----
192.0.2.2/32     ldp    MPLS  65540   9     192.0.2.5    30
192.0.2.3/32     ldp    MPLS  65541   9     192.0.2.5    40
192.0.2.4/32     ldp    MPLS  65538   9     192.0.2.5    30
192.0.2.5/32    rsvp  MPLS  1     7     192.168.15.2 16777215
192.0.2.5/32     ldp    MPLS  65537   9     192.0.2.5    10
192.0.2.6/32     ldp    MPLS  65539   9     192.0.2.5    20
192.0.2.7/32     ldp    MPLS  65542   9     192.0.2.5    30
192.0.2.8/32     rsvp   MPLS  2       7     192.168.15.2 16777215
192.0.2.8/32     ldp    MPLS  65543   9     192.0.2.8    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
=====
```

Endpoint node in OSPF area 0 can be either P-6, P-7, or P-8 (only those nodes have a T-LDP session toward P-5). The following command on P-5 shows that P-6 is the endpoint node (EgrNextHop). The RSVP LSP that is used on P-5 has ID 1.

```
[/]A:admin@P-5# show router ldp bindings active prefixes prefix 192.0.2.2/32

=====
LDP Bindings (IPv4 LSR ID 192.0.2.5)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
```

```

WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
e - Label ELC
FEC Flags:
LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
BA - ASBR Backup FEC
(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
-----
192.0.2.2/32                          Push
--
192.0.2.6                            LspId 1

192.0.2.2/32                          Swap
524268                                524270
192.0.2.6                            LspId 1
-----
No. of IPv4 Prefix Active Bindings: 2
=====

```

The following command shows that the RSVP LSP with ID 1 is LSP-P-5-P-6:

```

[/]
A:admin@P-5# show router mpls lsp
=====
MPLS LSPs (Originating)
=====
LSP Name                               Tun  Fastfail  Adm  Opr
To                                     Id    Config
-----
LSP-P-5-P-6                            1    Yes    Up  Up
192.0.2.6
LSP-P-5-P-7                             2     Yes      Up   Up
192.0.2.7
LSP-P-5-P-8                             3     Yes      Up   Up
192.0.2.8
LSP-P-5-PE-1                             4     No       Up   Up
192.0.2.1
LSP-P-5-PE-4                             5     No       Up   Up
192.0.2.4
-----
LSPs : 5
=====

```

The following command shows that the RSVP tunnel with ID 1 and destination 192.0.2.6 has next-hop 192.168.56.2 and metric 10:

```

[/]
A:admin@P-5# show router tunnel-table
=====
IPv4 Tunnel Table (Router: Base)
=====

```



Destination Color	Owner	Encap	TunnelId	Pref	Nexthop	Metric
192.0.2.1/32	rsvp	MPLS	4	7	192.168.15.1	16777215
192.0.2.1/32	ldp	MPLS	65537	9	192.0.2.1	10
192.0.2.2/32	ldp	MPLS	65540	9	192.0.2.6	20
192.0.2.3/32	ldp	MPLS	65541	9	192.0.2.7	30
192.0.2.4/32	rsvp	MPLS	5	7	192.168.58.2	16777215
192.0.2.4/32	ldp	MPLS	65538	9	192.0.2.4	20
<b>192.0.2.6/32 [B]</b>	<b>rsvp</b>	<b>MPLS</b>	<b>1</b>	<b>7</b>	<b>192.168.56.2</b>	<b>10</b>
192.0.2.6/32	ldp	MPLS	65539	9	192.0.2.6	10
192.0.2.7/32 [B]	rsvp	MPLS	2	7	192.168.56.2	20
192.0.2.7/32	ldp	MPLS	65542	9	192.0.2.7	20
192.0.2.8/32 [B]	rsvp	MPLS	3	7	192.168.58.2	10
192.0.2.8/32	ldp	MPLS	65543	9	192.0.2.8	10

Flags: B = BGP or MPLS backup hop available  
 L = Loop-Free Alternate (LFA) hop available  
 E = Inactive best-external BGP route  
 k = RIB-API or Forwarding Policy backup hop

On node P-6, the following commands are launched for the final destination node PE-2. Also there, an RSVP LSP toward PE-2 is used as transport tunnel for user packets.

```
[/]
A:admin@P-6# show router ldp bindings active prefixes prefix 192.0.2.2/32

=====
LDP Bindings (IPv4 LSR ID 192.0.2.6)
(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 (0) - SR-OSPF Next Hop
  (C) - FEC resolved with class-based-forwarding
=====
LDP IPv4 Prefix Bindings (Active)
=====
Prefix                               Op
IngLbl                               EgrLbl
EgrNextHop                           EgrIf/LspId
-----
192.0.2.2/32                          Push
--                                     524285
192.0.2.2                             LspId 4

192.0.2.2/32                          Swap
524270                                524285
192.0.2.2                             LspId 4
-----
No. of IPv4 Prefix Active Bindings: 2
```

```

=====
[/]
A:admin@P-6# show router mpls lsp
=====
MPLS LSPs (Originating)
=====
LSP Name          Tun   Fastfail  Adm  Opr
To                Id    Config
-----
LSP-P-6-P-5      1     Yes      Up   Up
  192.0.2.5
LSP-P-6-P-7      2     Yes      Up   Up
  192.0.2.7
LSP-P-6-P-8      3     Yes      Up   Up
  192.0.2.8
LSP-P-6-PE-2    4     No       Up   Up
  192.0.2.2
LSP-P-6-PE-3     5     No       Up   Up
  192.0.2.3
-----
LSPs : 5
=====

```

```

[/]
A:admin@P-6# show router tunnel-table
=====
IPv4 Tunnel Table (Router: Base)
=====
Destination      Owner   Encap TunnelId  Pref  Nexthop      Metric
Color
-----
192.0.2.1/32     ldp    MPLS  65539    9     192.0.2.5    20
192.0.2.2/32    rsvp   MPLS  4    7     192.168.26.1 16777215
192.0.2.2/32     ldp    MPLS  65537    9     192.0.2.2    10
192.0.2.3/32     rsvp   MPLS  5        7     192.168.67.2 16777215
192.0.2.3/32     ldp    MPLS  65538    9     192.0.2.3    20
192.0.2.4/32     ldp    MPLS  65542    9     192.0.2.8    30
192.0.2.5/32 [B]  rsvp   MPLS  1        7     192.168.56.1 10
192.0.2.5/32     ldp    MPLS  65540    9     192.0.2.5    10
192.0.2.7/32 [B]  rsvp   MPLS  2        7     192.168.67.2 10
192.0.2.7/32     ldp    MPLS  65541    9     192.0.2.7    10
192.0.2.8/32 [B]  rsvp   MPLS  3        7     192.168.67.2 20
192.0.2.8/32     ldp    MPLS  65543    9     192.0.2.8    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
=====

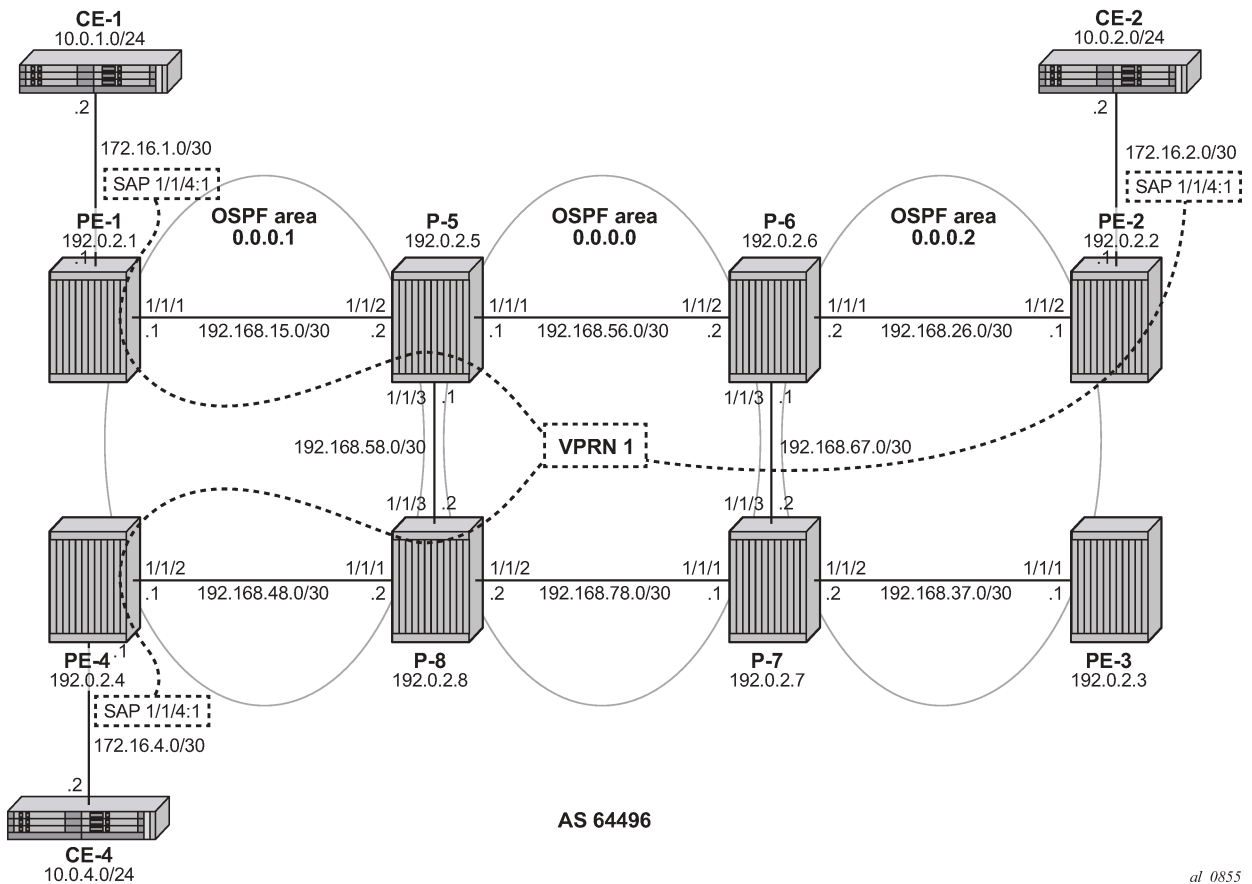
```

Nodes P-5 and P-6 behave as stitching nodes to stitch RSVP LSPs. P-5 stitches LSP-PE-1-P-5 and LSP-P-5-P-6 together while P-6 node stitches LSP-P-5-P-6 and LSP-P-6-PE-2 together.

When the endpoints are defined, one corresponding RSVP LSP to those endpoints is chosen (when ECMP equals 1). Selection criteria are as follows. When RSVP LSPs are configured under the T-LDP **tunneling** command (maximum 4), the one with the lowest LSP metric is selected. When no RSVP LSPs are configured under the T-LDP **tunneling** command, LDP checks the tunnel table for all available RSVP LSPs. The RSVP LSP with the lowest metric and operational state up is selected.

b. Traffic verification using a virtual private routed network (VPRN) service.

Figure 54: VPRN 1 with LDP over RSVP and no intra-area PE connectivity



al\_0855

VPRN service 1 is set up between three PE nodes (PE-1, PE-2, and PE-4) using the **auto-bind-tunnel resolution-filter ldp resolution filter**. See also [Figure 54: VPRN 1 with LDP over RSVP and no intra-area PE connectivity](#) for the exact addressing scheme.

```
# on PE-1:
configure {
  service {
    vprn "VPRN 1" {
      admin-state enable
      service-id 1
      customer "1"
      autonomous-system 64496
      bgp-ipvpn {
        mpls {
          admin-state enable
          route-distinguisher "64496:1"
          vrf-target {
            community "target:64496:1"
          }
        }
      }
      auto-bind-tunnel {
        resolution filter
        resolution-filter {
          ldp true
        }
      }
    }
  }
}
```

```

    }
  }
}
interface "int-PE-1-CE-1" {
  ipv4 {
    primary {
      address 172.16.1.1
      prefix-length 30
    }
  }
  sap 1/1/4:1 {
  }
}
static-routes {
  route 10.0.1.0/24 route-type unicast {
    next-hop "172.16.1.2" {
      admin-state enable
    }
  }
}
}

```

In order to distribute VPRN information (VPN-IPv4 routes and VPRN service labels) across the service provider network, Multi-Protocol Border Gateway Protocol (MP-BGP) is needed. MP-BGP is configured on PE-1, PE-2, and PE-4 with P-5 (192.0.2.5) being the Route Reflector (RR). In this way, no full BGP mesh between the three PE-nodes is needed, only a BGP peering toward the RR.

```

# on PE-1:
configure {
  router "Base" {
    autonomous-system 64496
    bgp {
      admin-state enable
      group "internal" {
        type internal
        family {
          vpn-ipv4 true
        }
      }
      neighbor "192.0.2.5" {
        group "internal"
      }
    }
  }
}

```

```

# on P-5:
configure {
  router "Base" {
    autonomous-system 64496
    bgp {
      admin-state enable
      group "internal" {
        type internal
        family {
          vpn-ipv4 true
        }
      }
      cluster {
        cluster-id 5.5.5.5
      }
    }
    neighbor "192.0.2.1" {
      group "internal"
    }
  }
}

```

```

neighbor "192.0.2.2" {
    group "internal"
}
neighbor "192.0.2.4" {
    group "internal"
}
}

```

If user traffic is monitored between PE-1 (ingress LER) and PE-2 (egress LER), three labels are seen. The outer label is the transport label distributed using the RSVP protocol, the inner label is the service label distributed using MP-BGP. LDP over RSVP adds an extra MPLS transport label between the outer transport and the service label (distributed using LDP). This middle label is used to tell the endpoint nodes (P-5 and P-6 acting as ABR) what to do. The transport label stack contains two labels: an RSVP label and an LDP label.

The following command shows that RSVP transport label 524287 is added as the outer label on each user packet sent on the link from PE-1 to P-5:

```

[/]
A:admin@PE-1# show router rsvp session lsp-name "LSP-PE-1-P-5::path-PE-1-P-5" detail
=====
RSVP Sessions (Detailed)
=====
-----
LSP : LSP-PE-1-P-5::path-PE-1-P-5
-----
From          : 192.0.2.1          To          : 192.0.2.5
Tunnel ID     : 1              LSP ID      : 9728
Style         : SE             State        : Up
Session Type  : Originate
In Interface  : n/a            Out Interface : 1/1/1
In IF Name    : n/a
Out IF Name   : int-PE-1-P-5
In Label      : n/a            Out Label    : 524287
Previous Hop  : n/a            Next Hop     : 192.168.15.2
Hops          :
              192.168.15.2(S)
SetupPriority : 7              Hold Priority : 0
Class Type    : 0
SubGrpOrig ID : 0              SubGrpOrig Addr:
P2MP ID      : 0
FrrAvailType : N/A
FrrSrlgStrict : N/A          SrlgDisjoint : N/A

Path Recd     : 0              Path Sent    : 41
Resv Recd     : 46            Resv Sent    : 0
Summary msgs  :
SPath Recd    : 0              SPath Sent   : 0
SResv Recd    : 0              SResv Sent   : 0
LSP Attr Flags : N/A
=====

```

The following command shows that LDP label 524268 is added as the middle label on each user packet:

```

[/]
A:admin@PE-1# show router ldp bindings active prefixes prefix 192.0.2.2/32
=====
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
-----
192.0.2.2/32                         Push
--
192.0.2.5                             524268
                                       LspId 1

192.0.2.2/32                         Swap
524281                               524268
192.0.2.5                             LspId 1
-----
No. of IPv4 Prefix Active Bindings: 2
=====

```

Service label 524277 is added as the inner MP-BGP label on each user packet.



**Note:**

This label will not change at endpoint nodes (P-5 and P-6). Ingress LER (PE-1) will push the service label to the user packet while the egress LER (PE-2) will pop the service label.

```

[/]
A:admin@PE-1# show router bgp routes vpn-ipv4
=====
BGP Router ID:192.0.2.1      AS:64496      Local AS:64496
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
               l - leaked, x - stale, > - best, b - backup, p - purge
Origin codes  : i - IGP, e - EGP, ? - incomplete
=====
BGP VPN-IPv4 Routes
=====
Flag  Network                               LocalPref  MED
      Nexthop (Router)                   Path-Id    IGP Cost
      As-Path                             Label
-----
i     64496:1:10.0.1.0/24                  100        None
      192.0.2.1                          None       0
      No As-Path                          524277
u*>i 64496:1:10.0.2.0/24                  100        None
      192.0.2.2                          None       30
      No As-Path                          524277
u*>i 64496:1:10.0.4.0/24                  100        None

```

```

192.0.2.4          None          30
No As-Path        524277
i 64496:1:172.16.1.0/30 100          None
192.0.2.1        None          0
No As-Path        524277
u*>i 64496:1:172.16.2.0/30 100          None
192.0.2.2        None          30
No As-Path        524277
u*>i 64496:1:172.16.4.0/30 100          None
192.0.2.4        None          30
No As-Path        524277
-----
Routes : 6
=====

```

The following command shows that RSVP transport label 524284 is added as the top label on each user packet for traffic sent from P-5 to P-6:

```

[/]
A:admin@P-5# show router rsvp session lsp-name "LSP-P-5-P-6::path-P-5-P-6" detail

=====
RSVP Sessions (Detailed)
=====
-----
LSP : LSP-P-5-P-6::path-P-5-P-6
-----
From           : 192.0.2.5           To           : 192.0.2.6
Tunnel ID      : 1             LSP ID       : 39936
Style          : SE           State        : Up
Session Type   : Originate
In Interface   : n/a         Out Interface : 1/1/1
In IF Name     : n/a
Out IF Name    : int-P-5-P-6
In Label       : n/a         Out Label   : 524284
Previous Hop   : n/a         Next Hop     : 192.168.56.2
Hops           :
              : 192.168.56.2(S)
SetupPriority  : 7             Hold Priority : 0
Class Type     : 0
SubGrpOrig ID : 0             SubGrpOrig Addr:
P2MP ID       : 0
FrrAvailType  : Facility
FrrBypassLspName: bypass-link192.168.56.2-61442
FrrSrlgStrict : N/A         SrlgDisjoint : N/A

Path Recd     : 0             Path Sent    : 44
Resv Recd     : 43          Resv Sent    : 0
Summary msgs  :
SPath Recd    : 0             SPath Sent   : 0
SResv Recd    : 0             SResv Sent   : 0
LSP Attr Flags : N/A
=====

```

LDP label 524270 is added as the middle label on each user packet.

```

[/]A:admin@P-5# show router ldp bindings active prefixes prefix 192.0.2.2/32

=====
LDP Bindings (IPv4 LSR ID 192.0.2.5)
(IPv6 LSR ID ::)
=====

```

```

Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
  (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
-----
192.0.2.2/32                          Push
--                                     524270
192.0.2.6                              LspId 1

192.0.2.2/32                          Swap
524268                                524270
192.0.2.6                              LspId 1
-----
No. of IPv4 Prefix Active Bindings: 2
=====

```

Service label 524277 is added as the inner MP-BGP label on each user packet.

The following command shows that RSVP transport label 524287 is added as the outer label on each user packet sent from P-6 to PE-2.

```

[/]
A:admin@P-6# show router rsvp session lsp-name "LSP-P-6-PE-2::path-P-6-PE-2" detail
=====
RSVP Sessions (Detailed)
=====
LSP : LSP-P-6-PE-2::path-P-6-PE-2
-----
From          : 192.0.2.6          To          : 192.0.2.2
Tunnel ID    : 4                LSP ID     : 1536
Style        : SE              State      : Up
Session Type : Originate
In Interface : n/a             Out Interface : 1/1/1
In IF Name   : n/a
Out IF Name  : int-P-6-PE-2
In Label     : n/a             Out Label  : 524287
Previous Hop : n/a             Next Hop   : 192.168.26.1
Hops         :
              192.168.26.1(S)
SetupPriority : 7                Hold Priority : 0
Class Type   : 0
SubGrpOrig ID : 0              SubGrpOrig Addr:
P2MP ID     : 0
FrrAvailType : N/A
FrrSrlgStrict : N/A           SrlgDisjoint : N/A
Path Recd   : 0                Path Sent   : 43

```



```

Resv Recd      : 44                Resv Sent       : 0
Summary msgs   :                   SPath Sent      : 0
SPath Recd    : 0                SResv Sent     : 0
SResv Recd    : 0
LSP Attr Flags : N/A
=====

```

c. LDP label 524285 is added as the middle label on each user packet.

```

[/]
A:admin@P-6# show router ldp bindings active prefixes prefix 192.0.2.2/32

=====
LDP Bindings (IPv4 LSR ID 192.0.2.6)
              (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
-----
192.0.2.2/32    Push
--              524285
192.0.2.2       LspId 4

192.0.2.2/32    Swap
524270         524285
192.0.2.2       LspId 4
-----
No. of IPv4 Prefix Active Bindings: 2
=====

```

Service label 524277 is added as the inner MP-BGP label on each user packet.

6. Intra-area FEC resolving (ingress LER is PE-1, egress LER is PE-4).

a. Verification endpoint node and associated RSVP tunnel.

The first thing to do in the intra-area FEC resolving process is for PE-1 to perform an SPF calculation toward PE-4 to search for an eligible endpoint, as close as possible to PE-4. An endpoint is eligible when:

- a T-LDP session exists between PE-1 and the endpoint node
- tunneling is configured on the endpoint node
- PE-1 received a label for the destination FEC from the endpoint node
- and an RSVP LSP that can be used for LDP over RSVP exists between PE-1 and the endpoint node

First endpoint node in OSPF area 1 can be either P-5 or P-8 (only those nodes have a T-LDP session toward PE-1). With **show router ldp bindings active prefixes prefix 192.0.2.4/32**, it can be concluded that P-5 is the endpoint node. Furthermore, LSP ID 1 indicates that an RSVP LSP toward P-5 is used.

```
[/]
A:admin@PE-1# show router ldp bindings active prefixes prefix 192.0.2.4/32

=====
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
-----
192.0.2.4/32                         Push
--
192.0.2.5                          LspId 1
-----
192.0.2.4/32                         Swap
524283                               524269
192.0.2.5                             LspId 1
-----
No. of IPv4 Prefix Active Bindings: 2
=====
```

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

=====
MPLS LSPs (Originating)
=====
LSP Name                               Tun   Fastfail  Adm  Opr
To                                     Id     Config
-----
LSP-PE-1-P-5                          1     No      Up Up
192.0.2.5
LSP-PE-1-P-8                          2     No        Up  Up
192.0.2.8
-----
LSPs : 2
=====
```

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

```

=====
IPv4 Tunnel Table (Router: Base)
=====
Destination          Owner    Encap TunnelId  Pref  NextHop      Metric
  Color
-----
192.0.2.2/32         ldp     MPLS  65540    9    192.0.2.5    30
192.0.2.3/32         ldp     MPLS  65541    9    192.0.2.5    40
192.0.2.4/32         ldp     MPLS  65538    9    192.0.2.5    30
192.0.2.5/32       rsvp   MPLS  1    7    192.168.15.2 16777215
192.0.2.5/32         ldp     MPLS  65537    9    192.0.2.5    10
192.0.2.6/32         ldp     MPLS  65539    9    192.0.2.5    20
192.0.2.7/32         ldp     MPLS  65542    9    192.0.2.5    30
192.0.2.8/32         rsvp    MPLS  2        7    192.168.15.2 16777215
192.0.2.8/32         ldp     MPLS  65543    9    192.0.2.8    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
=====

```

On node P-5, the same commands can be repeated for the final destination node (PE-4). Also there, an RSVP LSP toward PE-4 is used as transport tunnel for user packets.

```

[/]
A:admin@P-5# show router ldp bindings active prefixes prefix 192.0.2.4/32

=====
LDP Bindings (IPv4 LSR ID 192.0.2.5)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
  (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
-----
192.0.2.4/32    Push
--             524285
192.0.2.4       LspId 5

192.0.2.4/32    Swap
524269          524285
192.0.2.4     LspId 5
-----
No. of IPv4 Prefix Active Bindings: 2

```

```

=====
[/]
A:admin@P-5# show router mpls lsp to 192.0.2.4
=====
MPLS LSPs (Originating)
=====
LSP Name          Tun   Fastfail  Adm  Opr
To                Id    Config
-----
LSP-P-5-PE-4    5    No      Up Up
192.0.2.4
-----
LSPs : 1
=====

```

```

[/]
A:admin@P-5# show router tunnel-table
=====
IPv4 Tunnel Table (Router: Base)
=====
Destination      Owner   Encap TunnelId  Pref  Nexthop      Metric
Color
-----
192.0.2.1/32     rsvp   MPLS  4           7     192.168.15.1 16777215
192.0.2.1/32     ldp    MPLS  65537      9     192.0.2.1     10
192.0.2.2/32     ldp    MPLS  65540      9     192.0.2.6     20
192.0.2.3/32     ldp    MPLS  65541      9     192.0.2.7     30
192.0.2.4/32    rsvp  MPLS  5    7    192.168.58.2 16777215
192.0.2.4/32     ldp    MPLS  65538      9     192.0.2.4     20
192.0.2.6/32 [B]  rsvp   MPLS  1           7     192.168.56.2 10
192.0.2.6/32     ldp    MPLS  65539      9     192.0.2.6     10
192.0.2.7/32 [B]  rsvp   MPLS  2           7     192.168.56.2 20
192.0.2.7/32     ldp    MPLS  65542      9     192.0.2.7     20
192.0.2.8/32 [B]  rsvp   MPLS  3           7     192.168.58.2 10
192.0.2.8/32     ldp    MPLS  65543      9     192.0.2.8     10
-----
Flags: B = BGP or MPLS backup hop available
       L = Loop-Free Alternate (LFA) hop available
       E = Inactive best-external BGP route
       k = RIB-API or Forwarding Policy backup hop
=====

```

P-5 node acts as a stitching node to stitch RSVP LSPs. P-5 stitches LSP-PE-1-P-5 and LSP-P-5-PE-4 together.

When the endpoint node (P-5) is defined, the corresponding RSVP LSP to this endpoint is chosen. Selection criteria are as follows (when ECMP=1). When RSVP LSPs are configured under the T-LDP **tunneling** command (maximum 4), the one with the lowest LSP metric is selected. When no RSVP LSPs are configured under the T-LDP **tunneling** command, LDP checks the tunnel table for all available RSVP LSPs. The RSVP LSP with the lowest metric and operational state **up** is selected.

- b.** Traffic verification using a VPRN service (see [Figure 54: VPRN 1 with LDP over RSVP and no intra-area PE connectivity](#)).

If user traffic between PE-1 (ingress LER) and PE-4 (egress LER) is monitored, three labels are seen. The outer label is the transport label (distributed using RSVP protocol), the inner label is the service label (distributed using MP-BGP). LDP over RSVP adds an extra MPLS transport label

between outer and inner label (distributed using LDP). This middle label is used to tell the endpoint node (P-5) what to do.

The following command shows that transport label 524287 is added as the top RSVP label on each user packet sent from PE-1 to P-5.

```
[/]
A:admin@PE-1# show router rsvp session lsp-name "LSP-PE-1-P-5::path-PE-1-P-5" detail

=====
RSVP Sessions (Detailed)
=====
-----
LSP : LSP-PE-1-P-5::path-PE-1-P-5
-----
-----
From           : 192.0.2.1           To           : 192.0.2.5
Tunnel ID      : 1               LSP ID       : 9728
Style          : SE              State        : Up
Session Type   : Originate
In Interface   : n/a            Out Interface : 1/1/1
In IF Name     : n/a
Out IF Name    : int-PE-1-P-5
In Label       : n/a            Out Label   : 524287
Previous Hop   : n/a            Next Hop     : 192.168.15.2
Hops           :
                192.168.15.2(S)
SetupPriority  : 7               Hold Priority : 0
Class Type    : 0
SubGrpOrig ID : 0               SubGrpOrig Addr:
P2MP ID       : 0
FrrAvailType  : N/A
FrrSrlgStrict : N/A            SrlgDisjoint : N/A

Path Recd     : 0               Path Sent    : 45
Resv Recd     : 51            Resv Sent    : 0
Summary msgs  :
SPath Recd    : 0               SPath Sent   : 0
SResv Recd    : 0               SResv Sent   : 0
LSP Attr Flags : N/A
=====
```

LDP over RSVP label 524269 is added as the middle LDP label on each user packet.

```
[/]
A:admin@PE-1# show router ldp bindings active prefixes prefix 192.0.2.4/32

=====
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
-----
192.0.2.4/32                         Push
--
192.0.2.5                             LspId 1

192.0.2.4/32                         Swap
524283                               524269
192.0.2.5                             LspId 1

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

```

Service label 524277 is added as the inner MP-BGP label on each user packet.



**Note:**

This label will not change at endpoint node (P-5). Ingress LER (PE-1) will push the service label to the user packet while the egress LER (PE-4) will pop the service label.

```

[/]
A:admin@PE-1# show router bgp neighbor 192.0.2.5 received-routes vpn-ipv4

=====
BGP Router ID:192.0.2.1      AS:64496      Local AS:64496
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
               l - leaked, x - stale, > - best, b - backup, p - purge
Origin codes  : i - IGP, e - EGP, ? - incomplete

=====
BGP VPN-IPv4 Routes
=====
Flag  Network                               LocalPref  MED
      Nexthop (Router)                    Path-Id    IGP Cost
      As-Path                               Label
-----
i     64496:1:10.0.1.0/24                   100        None
      192.0.2.1                             None        0
      No As-Path                             524277
u*>i 64496:1:10.0.2.0/24                   100        None
      192.0.2.2                             None        30
      No As-Path                             524277
u*>i 64496:1:10.0.4.0/24                   100        None
      192.0.2.4                             None        30
      No As-Path                             524277
i     64496:1:172.16.1.0/30                100        None
      192.0.2.1                             None        0
      No As-Path                             524277
u*>i 64496:1:172.16.2.0/30                100        None
      192.0.2.2                             None        30
      No As-Path                             524277
u*>i 64496:1:172.16.4.0/30                100        None
      192.0.2.4                             None        30
      No As-Path                             524277
-----
Routes : 6

```

The following command shows that P-5 pushes RSVP transport label 524283 as the top label on each user packet sent on path-P-5-P-8-PE-4. This RSVP transport label is swapped by P-8 to label 524287.

```
[/]
A:admin@P-5# show router mpls lsp "LSP-P-5-PE-4" path detail

=====
MPLS LSP LSP-P-5-PE-4 Path (Detail)
=====
Legend :
  @ - Detour Available          # - Detour In Use
  b - Bandwidth Protected       n - Node Protected
  s - Soft Preemption           L - Loose
  S - Strict                    + - Inherited
  A - ABR

-----
LSP LSP-P-5-PE-4
Path path-P-5-P-8-PE-4
-----
LSP Name      : LSP-P-5-PE-4
From          : 192.0.2.5
To            : 192.0.2.4
Admin State   : Up
Path Name     : path-P-5-P-8-PE-4
Path LSP ID   : 36352
Path Admin    : Up
Out Interface : 1/1/3
Path Up Time  : 0d 00:23:13
Retry Limit   : 0
Retry Attempt : 0
Oper State    : Up
Path Type     : Primary
Path Oper     : Up
Out Label   : 524283
Path Down Time : 0d 00:00:00
Retry Timer   : 30 sec
Next Retry In : 0 sec

BFD Configuration and State
Template      : None
Enable       : False
WaitForUpTimer : 4 sec
WaitForUpTmLeft : 0 sec
Ping Interval : 60
State        : notApplicable
OperWaitForUpTimer: N/A

Adspec       : Disabled
PathCompMethod : none
MetricType   : igp
Least Fill   : Disabled
FRR          : Disabled
Propagate Adm Grp: Disabled
Inter-area   : False
Oper Adspec  : Disabled
OperPathCompMethod: none
Oper MetricType : igp
Oper LeastFill : Disabled
Oper FRR      : Disabled
Oper Prop Adm Grp : Disabled

PCE Report   : Disabled+
PCE Control  : Disabled
PCE Update ID : 0
Oper PCE Report : Disabled
Oper PCE Control : Disabled

Neg MTU      : 1564
Bandwidth    : No Reservation
Hop Limit    : 255
Record Route : Record
Record Label : Record
Setup Priority : 7
Hold Priority : 0
Class Type   : 0
Backup CT    : None
MainCT Retry : n/a
Oper MTU     : 1564
Oper Bandwidth : 0 Mbps
Oper HopLimit : 255
Oper Record Route : Record
Oper Record Label : Record
Oper SetupPriority: 7
Oper HoldPriority : 0
Oper CT      : 0
```

```

Rem          :
MainCT Retry : 0
  Limit      :
Include Groups :                               Oper IncludeGroups:
None                                                None
Exclude Groups :                               Oper ExcludeGroups:
None                                                None

Adaptive      : Enabled                       Oper Metric      : 16777215
Preference    : n/a
Path Trans    : 1                             CSPF Queries     : 0
Failure Code  : noError
Failure Node  : n/a
Explicit Hops :
              192.168.58.2(S)
              -> 192.168.48.1(S)
Actual Hops   :
              192.168.58.1(192.0.2.5)          Record Label     : N/A
-> 192.168.58.2(192.0.2.8)                    Record Label     : 524283
-> 192.168.48.1                                Record Label     : 524287
Resignal Eligible: False
Last Resignal : n/a                           CSPF Metric      : 0
=====

```



**Note:**

**show router rsvp session lsp-name LSP-P-5-PE-4::path-P-5-P-8-PE-4 detail** cannot be used because it only shows the outgoing RSVP label toward node P-8. On node P-8, RSVP transport label 524283 will be swapped into RSVP transport label 524287 for the link from P-8 to PE-4.

LDP label 524285 is added as the middle label on each user packet.

```

[/]
A:admin@P-5# show router ldp bindings active prefixes prefix 192.0.2.4/32
=====
LDP Bindings (IPv4 LSR ID 192.0.2.5)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
  (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
-----
192.0.2.4/32    Push
  - -          524285
192.0.2.4      LspId 5
192.0.2.4/32    Swap

```



```

524269                               524285
192.0.2.4                             LspId 5

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

```

Service label 524277 is added as the inner MP-BGP label on each user packet.

```

[/]A:admin@P-5# show router bgp neighbor 192.0.2.4 received-routes vpn-ipv4
=====
BGP Router ID:192.0.2.5          AS:64496          Local AS:64496
=====
Legend -
Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                l - leaked, x - stale, > - best, b - backup, p - purge
Origin codes  : i - IGP, e - EGP, ? - incomplete

=====
BGP VPN-IPv4 Routes
=====
Flag  Network                               LocalPref  MED
      Nexthop (Router)                       Path-Id    IGP Cost
      As-Path                                Path-Id    Label
-----
*>i  64496:1:10.0.4.0/24                      100        None
      192.0.2.4                               None        20
      No As-Path                               None        524277
*>i  64496:1:172.16.4.0/30                   100        None
      192.0.2.4                               None        20
      No As-Path                               None        524277
-----
Routes : 2
=====

```

## Additional topics

### prefer-tunnel-in-tunnel

If the next-hop router advertised the same FEC over link-level LDP (iLDP), LDP prefers the iLDP tunnel by default unless the user explicitly changed the default preference using the **prefer-tunnel-in-tunnel** command. When **prefer-tunnel-in-tunnel** is set to false, an LDP over RSVP tunnel has precedence.

Until now, no RSVP LSPs are configured inside the **ldp targeted-session peer tunneling** context. Therefore, two additional strict non-CSPF RSVP LSPs are added between ingress LER PE-1 and egress LER P-5. Both LSPs have an explicit metric setting and are applied inside the **ldp tunneling** context. On the Layer 3 interface between PE-1 and P-5, iLDP is enabled.

```

# on PE-1:
configure {
  router "Base" {
    ldp {
      interface-parameters {
        interface "int-PE-1-P-5" {
          ipv4 {
            admin-state enable
          }
        }
      }
    }
  }
}

```

```

    }
}

# on P-5:
configure {
  router "Base" {
    ldp {
      interface-parameters {
        interface "int-P-5-PE-1" {
          ipv4 {
            admin-state enable
          }
        }
      }
    }
  }
}

```

```

# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-P-5-metric100" {
        admin-state enable
        type p2p-rsvp
        to 192.0.2.5
        metric 100
        primary "path-PE-1-P-5" {
        }
      }
      lsp "LSP-PE-1-P-5-metric200" {
        admin-state enable
        type p2p-rsvp
        to 192.0.2.5
        metric 200
        primary "path-PE-1-P-5" {
        }
      }
    }
  }
}

```

```

# on PE-1:
configure
  router "Base" {
    ldp {
      targeted-session {
        peer 192.0.2.5 {
          tunneling {
            lsp "LSP-PE-1-P-5-metric100" { }
            lsp "LSP-PE-1-P-5-metric200" { }
          }
        }
      }
    }
  }
}

```

The following tunnel table on node PE-1 contains four tunnels toward P-5: one LDP tunnel and three RSVP tunnels:

```

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

=====
IPv4 Tunnel Table (Router: Base)
=====
Destination          Owner      Encap TunnelId  Pref  Nexthop      Metric
Color

```

```
-----
192.0.2.2/32      ldp      MPLS  65540   9      192.168.15.2  30
192.0.2.3/32      ldp      MPLS  65541   9      192.168.15.2  40
192.0.2.4/32      ldp      MPLS  65538   9      192.168.15.2  30
192.0.2.5/32      rsvp     MPLS   3       7      192.168.15.2  100
192.0.2.5/32      rsvp     MPLS   4       7      192.168.15.2  200
192.0.2.5/32      rsvp     MPLS   1       7      192.168.15.2  16777215
192.0.2.5/32      ldp      MPLS  65537   9      192.168.15.2  10
192.0.2.6/32      ldp      MPLS  65539   9      192.168.15.2  20
192.0.2.7/32      ldp      MPLS  65542   9      192.168.15.2  30
192.0.2.8/32      rsvp     MPLS   2       7      192.168.15.2  16777215
192.0.2.8/32      ldp      MPLS  65543   9      192.168.15.2  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
=====
```

Tunnel ID 1 is a reference to LSP-PE-1-P-5. Tunnel ID 3 is a reference to LSP-PE-1-P-5-metric100. Tunnel ID 4 is a reference to LSP-PE-1-P-5-metric200 and owner LDP is a reference to iLDP.

Taken into account the FEC resolution rules, iLDP prevails, so no LDP over RSVP tunnel is used. The following command shows that the egress interface is 1/1/1; no RSVP LSP is used, so no LSP ID is present:

```
[/]
A:admin@PE-1# show router ldp bindings active prefixes prefix 192.0.2.5/32

=====
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
-----
192.0.2.5/32    Push
--             524271
192.168.15.2    1/1/1

192.0.2.5/32    Swap
524284          524271
192.168.15.2    1/1/1

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

This behavior can be changed by setting the **prefer-tunnel-in-tunnel** command in the **ldp** context to false. Now, the LDP over RSVP tunnel with the best (= lowest) metric is taken.

```
# on PE-1:
configure {
  router "Base" {
    ldp {
      prefer-tunnel-in-tunnel false
    }
  }
}
```

The following command shows that LSP ID 3 is used:

```
[/]
A:admin@PE-1# show router ldp bindings active prefixes prefix 192.0.2.5/32

=====
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
-----
192.0.2.5/32                         Push
--                                   524271
192.0.2.5                             LspId 3

192.0.2.5/32                         Swap
524284                               524271
192.0.2.5                             LspId 3

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

The following command shows that LSP ID 3 corresponds to LSP-PE-1-P-5-metric100:

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

=====
MPLS LSPs (Originating)
=====
LSP Name                               Tun   Fastfail  Adm  Opr
To                                     Id     Config
-----
LSP-PE-1-P-5                           1     No        Up   Up
192.0.2.5
```

```

LSP-PE-1-P-8                2      No      Up      Up
 192.0.2.8
LSP-PE-1-P-5-metric100    3      No      Up      Up
 192.0.2.5
LSP-PE-1-P-5-metric200     4      No      Up      Up
 192.0.2.5
-----
LSPs : 4
=====

```

When the LSP-PE-1-P-5-metric100 is disabled, then the LSP-PE-1-P-5-metric200 becomes active.

```

# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-P-5-metric100" {
        admin-state disable
      }
    }
  }
}

```

The following command shows that LSP ID 4 is used for traffic toward 192.0.2.5:

```

[/]
A:admin@PE-1# show router ldp bindings active prefixes prefix 192.0.2.5/32

=====
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
-----
192.0.2.5/32                          Push
--                                     524271
192.0.2.5                              LspId 4

192.0.2.5/32                          Swap
524284                                 524271
192.0.2.5                              LspId 4

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

```

The following command shows that LSP ID 4 corresponds to LSP-PE-1-P-5-metric200:

```

[/]

```

```
A:admin@PE-1# show router mpls lsp
```

```
=====
MPLS LSPs (Originating)
=====
LSP Name                               Tun   Fastfail Adm  Opr
To                                     Id    Config
-----
LSP-PE-1-P-5                           1     No       Up  Up
  192.0.2.5
LSP-PE-1-P-8                           2     No       Up  Up
  192.0.2.8
LSP-PE-1-P-5-metric100                 3     No       Dwn Dwn
  192.0.2.5
LSP-PE-1-P-5-metric200                4     No       Up  Up
  192.0.2.5
-----
LSPs : 4
=====
```

When LSP-PE-1-P-5-metric200 is disabled too, iLDP resumes.

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-P-5-metric200" {
        admin-state disable
      }
    }
  }
}
```

The following command shows that iLDP is used and the egress interface is port 1/1/1:

```
[/]
A:admin@PE-1# show router ldp bindings active prefixes prefix 192.0.2.5/32
=====
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
-----
192.0.2.5/32                         Push
  --                                524271
192.168.15.2                         1/1/1
192.0.2.5/32                         Swap
524284                                524271
```

```

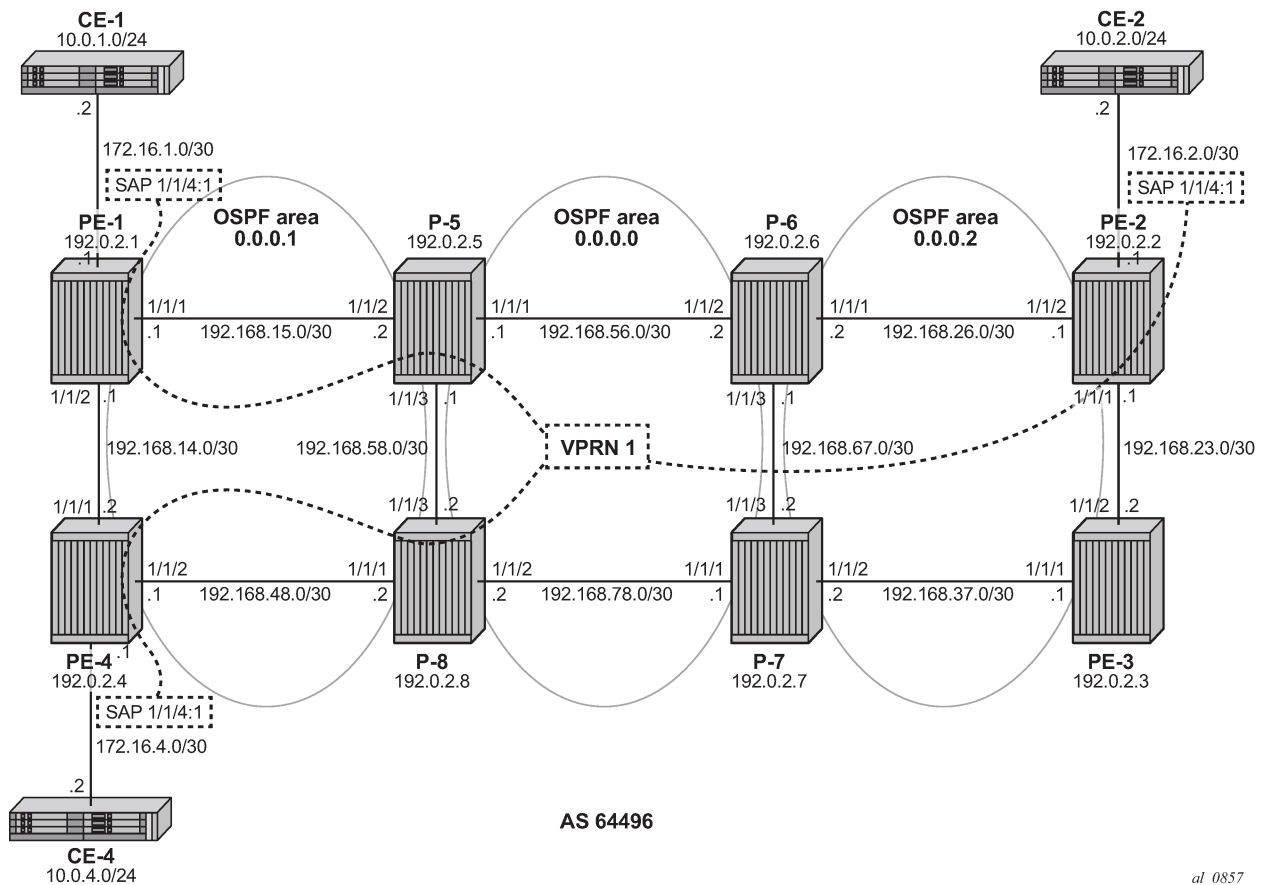
192.168.15.2                               1/1/1
-----
No. of IPv4 Prefix Active Bindings: 2
=====

```

### Intra-PE connectivity changes LDP over RSVP behavior

Figure 55: VPRN 1 with LDP over RSVP and intra-area PE connectivity shows two metro areas; both of the intra PEs are physically connected with each other. Compared with the previous figures, PE-1 is directly connected to PE-4 and PE-2 is directly connected to PE-3 (up to the OSPF level).

Figure 55: VPRN 1 with LDP over RSVP and intra-area PE connectivity



The SPF path calculation on PE-1 toward destination (PE-4) does not point to node P-5 anymore (as was seen before), but now points directly to PE-4 (shortest, lowest IGP metric). As a conclusion, it can be said that when possible intra-area endpoint nodes are not part of the calculated SPF path, LDP over RSVP is not preferred anymore. For this situation, it is advisable to configure iLDP on the intra-PE interfaces to have a fallback mechanism.

This is configured on PE-1 and PE-4 as follows:

```

# on PE-1:
configure {

```

```

router "Base" {
  interface "int-PE-1-PE-4" {
    port 1/1/2
    ipv4 {
      primary {
        address 192.168.14.1
        prefix-length 30
      }
    }
  }
  ospf 0 {
    area 0.0.0.1 {
      interface "int-PE-1-PE-4" {
        interface-type point-to-point
      }
    }
  }
}

```

```

# on PE-4:
configure {
  router "Base" {
    interface "int-PE-4-PE-1" {
      port 1/1/1
      ipv4 {
        primary {
          address 192.168.14.2
          prefix-length 30
        }
      }
    }
  }
  ospf 0 {
    area 0.0.0.1 {
      interface "int-PE-4-PE-1" {
        interface-type point-to-point
      }
    }
  }
}

```

LDP is configured on the interfaces between PE-1 and PE-4, as follows:

```

# on PE-1:
configure {
  router "Base" {
    ldp {
      interface-parameters {
        interface "int-PE-1-PE-4" {
          ipv4 {
            admin-state enable
          }
        }
      }
    }
  }
}

```

```

# on PE-4:
configure {
  router "Base" {
    ldp {
      interface-parameters {
        interface "int-PE-4-PE-1" {
          ipv4 {
            admin-state enable
          }
        }
      }
    }
  }
}

```



```
}

```

From the moment iLDP is configured, an LDP LSP is set up. Intra-area PE traffic will flow over this LDP LSP.

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

=====
IPv4 Tunnel Table (Router: Base)
=====
Destination          Owner      Encap TunnelId  Pref  Nexthop      Metric
  Color
-----
192.0.2.4/32         ldp        MPLS  65544    9    192.168.14.2  10
-----
Flags: B = BGP or MPLS backup hop available
       L = Loop-Free Alternate (LFA) hop available
       E = Inactive best-external BGP route
       k = RIB-API or Forwarding Policy backup hop
=====
```

If user traffic is monitored between ingress LER PE-1 and egress LER PE-4, only two labels are seen. The outer label is the transport label distributed using LDP; the inner label is the service label distributed using MP-BGP. No LDP over RSVP label is present anymore. The following command shows that LDP transport label 524285 is pushed by PE-1 as the outer label on packets destined to PE-4:

```
[/]A:admin@PE-1# show router ldp bindings active prefixes prefix 192.0.2.4/32

=====
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
-----
192.0.2.4/32    Push
--             524285
192.168.14.2    1/1/2

192.0.2.4/32    Swap
524283          524285
192.168.14.2    1/1/2
-----
No. of IPv4 Prefix Active Bindings: 2
```

Service label 524277 is added as the inner MP-BGP label on each user packet.

```
[/]
A:admin@PE-1# show router bgp routes vpn-ipv4
=====
BGP Router ID:192.0.2.1      AS:64496      Local AS:64496
=====
Legend -
Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                l - leaked, x - stale, > - best, b - backup, p - purge
Origin codes  : i - IGP, e - EGP, ? - incomplete
=====
BGP VPN-IPv4 Routes
=====
Flag  Network                               LocalPref  MED
      Nexthop (Router)                     Path-Id    IGP Cost
      As-Path                               Label
-----
i     64496:1:10.0.1.0/24                    100        None
      192.0.2.1                             None       0
      No As-Path                             524277
u*>i 64496:1:10.0.2.0/24                    100        None
      192.0.2.2                             None       30
      No As-Path                             524277
u*>i 64496:1:10.0.4.0/24                    100        None
      192.0.2.4                             None       10
      No As-Path                             524277
i     64496:1:172.16.1.0/30                 100        None
      192.0.2.1                             None       0
      No As-Path                             524277
u*>i 64496:1:172.16.2.0/30                 100        None
      192.0.2.2                             None       30
      No As-Path                             524277
u*>i 64496:1:172.16.4.0/30                 100        None
      192.0.2.4                             None       10
      No As-Path                             524277
-----
Routes : 6
=====
```

## Conclusion

LDP over RSVP allows tunneling of user packets toward an LDP far-end destination inside an RSVP LSP (with the benefits of RSVP LSPs, fast-reroute (FRR) and traffic engineering (TE)). The main application of this feature is for deployment of MPLS based services, for example, VPRN, virtual leased line (VLL), and virtual private LAN service (VPLS) services, in large networks where a full mesh of LSPs reaches the limits of scalability.

# LDP Point-to-Point LSPs

This chapter provides information about label distribution protocol (LDP) point-to-point label switched paths (LSPs)

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

## Applicability

This chapter is applicable to SR OS and was originally written for SR OS Release 7.0.R5. The MD-CLI output in the current edition corresponds to SR OS Release 21.2.R1. There are no prerequisites or conditions on the hardware for this configuration.

## Overview

Due to the connectionless nature of the network layer protocol IP, packets travel through the network on a hop-by-hop basis with routing decisions made at each node. As a result, hyperaggregation of data on certain links may occur and it may impact the provider's ability to provide guaranteed service levels across the network end-to-end. To address these shortcomings, Multi-Protocol Label Switching (MPLS) was developed.

MPLS provides the capability to establish connection-oriented paths, called Label Switched Paths (LSPs), over a connectionless (IP) network. The LSP offers a mechanism to engineer network traffic independently from the underlying network routing protocol (mostly IP) to improve the network resiliency and recovery options and to permit delivery of services that are not readily supported by conventional IP routing techniques, such as Layer 2 IP Virtual Private Networks (VPNs). These benefits are essential for today's communication network explaining the wide deployment base of the MPLS technology.

RFC 3031, *Multiprotocol Label Switching Architecture*, specifies the MPLS architecture whereas this chapter describes the configuration and troubleshooting of point-to-point LSPs on SR OS.

## Packet forwarding

When a packet of a connectionless network layer protocol travels from one router to the next, each router in the network makes an independent forwarding decision by performing the following basic tasks: first analyzing the packet's header, then referencing the local routing table to find the longest match based on the destination address in the IP header, and finally sending out the packet on the selected interface.

In other terms, the first function partitions the entire set of possible packets into a set of Forwarding Equivalence Classes (FECs). All packets associated to a particular FEC will be forwarded along the same

logical path to the same destination. The second function maps each FEC to a next hop destination router. Each router along the packet's path performs these actions.

In MPLS, the assignment of a packet to a particular FEC is done just once, when the packet enters the network. In turn, the FEC is mapped to an LSP, which is established prior to packet forwarding.

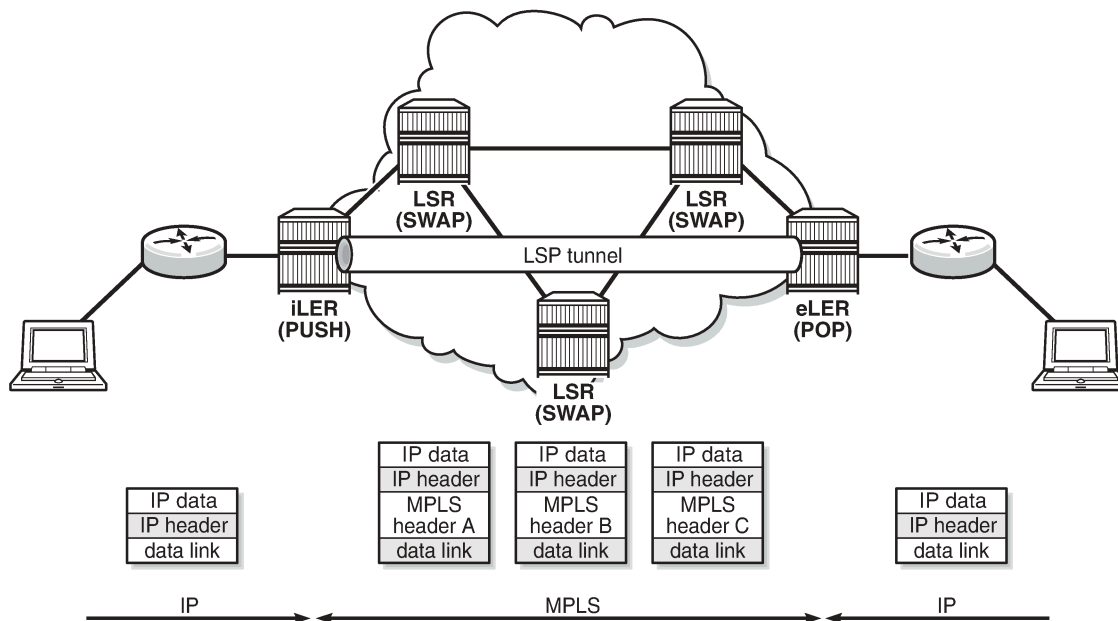
An MPLS label, representing the FEC to which the packet is assigned, is attached to the packet (push operation) and once labeled, the packet is forwarded to the next hop router along that LSP path.

At subsequent hops, no analysis of the packet's network layer header is needed. Instead, the label is used as an index into a table which specifies the next hop and a new label. The old label is replaced with the new label (swap operation), and the packet is forwarded to its next hop.

At the MPLS network egress, the label is removed from the packet (pop operation). If this router is the destination (based on the remaining packet), the packet is handed to the receiving application, such as a Virtual Private LAN Service (VPLS). If this router is not the destination of the packet, the packet will be sent into a new MPLS tunnel or forwarded by conventional IP forwarding toward the Layer 3 destination.

## Terminology

Figure 56: Generic MPLS network, MPLS label operations



25762

Figure 56: Generic MPLS network, MPLS label operations shows a general network topology clarifying the MPLS-related terms. A Label Edge Router (LER) is a device at the edge of an MPLS network, with at least one interface outside the MPLS domain. A router is usually defined as an LER based on its position relative to a particular LSP. The MPLS router at the head-end of an LSP is called the ingress Label Edge Router (iLER). The MPLS router at the tail-end of an LSP is called the egress Label Edge Router (eLER).

The iLER receives unlabeled packets from outside the MPLS domain, then applies MPLS labels to the packets, and forwards the labeled packets into the MPLS domain.

The eLER receives labeled packets from the MPLS domain, then removes the labels, and forwards unlabeled packets outside the MPLS domain. The eLER can signal an implicit-null label (numeric value 3). This informs the previous hop to send MPLS packets without an outer label and so is known as Penultimate Hop Popping (PHP).

A Label Switching Router (LSR) is a device internal to an MPLS network, with all interfaces inside the MPLS domain. These devices switch labeled packets inside the MPLS domain. In the core of the network, LSRs ignore the packet's network layer (IP) header and simply forward the packet using the MPLS label swapping mechanism.

## LSP establishment

Prior to packet forwarding, the LSP must be established. In order to do so, labels need to be distributed for the path. Labels are usually distributed by a downstream router in the upstream direction (relative to the data flow). There are a number of ways used for label distribution: static, LDP, and RSVP. For static P2P LSPs, see chapter [Static Point-to-Point LSPs](#); for RSVP-TE P2P LSPs, see chapter [RSVP Point-to-Point LSPs](#).

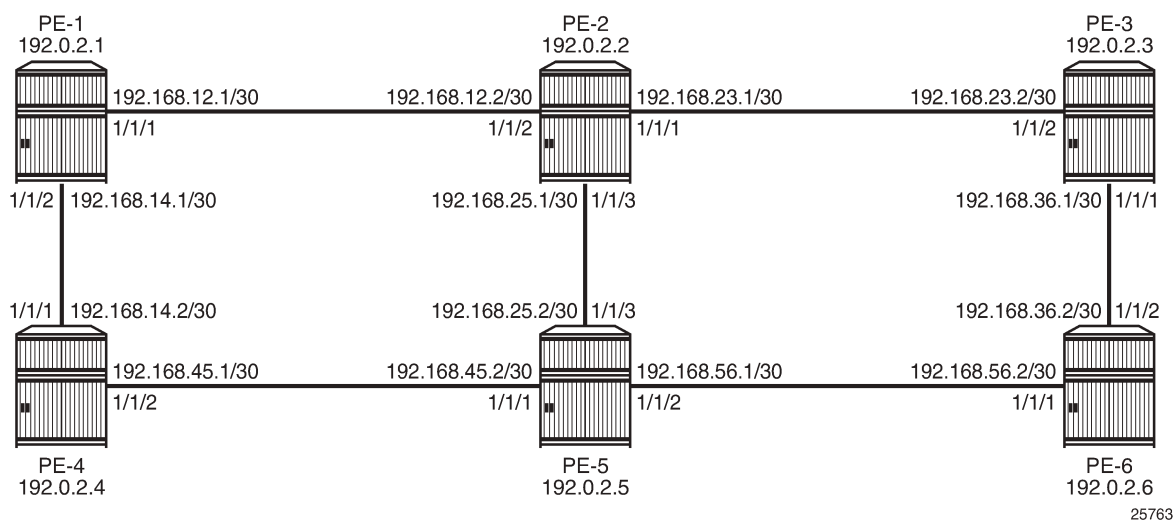
LDP (RFC 5036, *LDP Specification*) can be considered as an extension to the network interior gateway protocol (IGP). As routers become aware of new destination networks, they advertise labels in the upstream direction that will allow upstream routers to reach the destination.

Fast reroute (FRR) allows for computing backup paths and advertising the backup labels before a failure takes place. This way traffic can flow almost continuously, without waiting for routing protocol convergence; see chapter [MPLS LDP FRR using ISIS as IGP](#).

## Example topology

[Figure 57: MPLS example topology](#) shows the example topology consisting of six SR OS nodes located in a single autonomous system.

Figure 57: MPLS example topology



## Configuration

As a general prerequisite for the configuration of MPLS LSPs, a correctly working Interior Gateway Protocol (IGP) is required. Open Shortest Path First (OSPF) or Intermediate System to Intermediate System (IS-IS) can be used as IGP.

LDP is a simple label distribution protocol with basic MPLS functionality (no traffic engineering). Fast Reroute is supported; see chapter [MPLS LDP FRR using ISIS as IGP](#). LDP relies on the underlying routing information provided by an IGP in order to forward labeled packets. Each LDP configured LSR will originate a label for its system address and a label for each FEC for which it has a next hop that is external to the MPLS domain, without the explicit need to manually configure the LSPs. When deviations from this default behavior are desired, import and export policies can be applied.

The configuration is as simple as enabling the LDP protocol instance and adding all network interfaces, for each node. The configuration on node PE-1 is as follows; similar configurations apply on the other nodes.

```
# on PE-1:
configure {
  router "Base" {
    ldp {
      interface-parameters {
        interface "int-PE-1-PE-2" {
          ipv4 {
          }
        }
        interface "int-PE-1-PE-4" {
          ipv4 {
          }
        }
      }
    }
  }
}
```

The **show router ldp discovery** and **show router ldp session** commands can be used to verify the LDP hello adjacencies and sessions. The adjacency type (AdjType) needs to be **Link** while the state should be **Established**. In this example, only IPv4 addresses are used, so the output can be limited to IPv4 only by adding the keyword `ipv4`.

```
[/]
A:admin@PE-1# show router ldp discovery ipv4

=====
LDP IPv4 Hello Adjacencies
=====
Interface Name          Local Addr          State
AdjType                 Peer Addr
-----
int-PE-1-PE-2           192.0.2.1:0        Estab
link                   192.0.2.2:0
int-PE-1-PE-4           192.0.2.1:0        Estab
link                   192.0.2.4:0
-----
No. of IPv4 Hello Adjacencies: 2
=====
```

```
[/]
A:admin@PE-1# show router ldp session family ipv4
```

```

=====
LDP IPv4 Sessions
=====
Peer LDP Id      Adj Type  State      Msg Sent  Msg Recv  Up Time
-----
192.0.2.2:0     Link     Established 84         85        0d 00:03:14
192.0.2.4:0     Link     Established 74         76        0d 00:02:50
-----
No. of IPv4 Sessions: 2
=====

```

The **show router ldp bindings prefixes** command displays the contents of the LIB (Label Information Base) and contains all labels locally generated (IngrLbl) and those received from any LDP neighbors (EgrLbl), whether they are in use or not. The following output is for IPv4 prefixes:

```

[/]
A:admin@PE-1# show router ldp bindings 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
=====
LDP IPv4 Prefix Bindings
=====
Prefix
Peer          FEC-Flags
IngrLbl      EgrLbl
EgrNextHop   EgrIntf/LspId
-----
192.0.2.1/32
192.0.2.2:0
524287U      --
--          --

192.0.2.1/32
192.0.2.4:0
524287U      --
--          --

192.0.2.2/32
192.0.2.2:0
--          524287
192.168.12.2 1/1/1

192.0.2.2/32
192.0.2.4:0
524286U      524285
--          --

192.0.2.3/32
192.0.2.2:0
524285N      524285
192.168.12.2 1/1/1

```

```

192.0.2.3/32
192.0.2.4:0
524285U                               524284
--                                     --

192.0.2.4/32
192.0.2.2:0
524284U                               524284
--                                     --

192.0.2.4/32
192.0.2.4:0
--                                     524287
192.168.14.2                          1/1/2

192.0.2.5/32
192.0.2.2:0
524283N                               524283
192.168.12.2                          1/1/1

192.0.2.5/32
192.0.2.4:0
524283U                               524283
--                                     --

192.0.2.6/32
192.0.2.2:0
524282N                               524282
192.168.12.2                          1/1/1

192.0.2.6/32
192.0.2.4:0
524282U                               524282
--                                     --

-----
No. of IPv4 Prefix Bindings: 12
=====

```

The **show router ldp bindings active prefixes** command displays the content of the Label Forwarding Information Base (LFIB) and contains all active labels and the associated label actions used for label switching packets. The active LDP bindings for IPv4 prefixes are the following:

```

[/]
A:admin@PE-1# show router ldp bindings active prefixes ipv4

=====
LDP Bindings (IPv4 LSR ID 192.0.2.1)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
  (S) - Static           (M) - Multi-homed Secondary Support
  (B) - BGP Next Hop     (BU) - Alternate Next-hop for Fast Re-Route
  (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
  (C) - FEC resolved with class-based-forwarding
=====
LDP IPv4 Prefix Bindings (Active)

```



```

=====
Prefix                               Op
IngLbl                               EgrLbl
EgrNextHop                           EgrIf/LspId
-----
192.0.2.1/32                         Pop
524287                                --
--                                    --

192.0.2.2/32                         Push
--                                    524287
192.168.12.2                          1/1/1

192.0.2.2/32                         Swap
524286                                524287
192.168.12.2                          1/1/1

192.0.2.3/32                         Push
--                                    524285
192.168.12.2                          1/1/1

192.0.2.3/32                         Swap
524285                                524285
192.168.12.2                          1/1/1

192.0.2.4/32                         Push
--                                    524287
192.168.14.2                          1/1/2

192.0.2.4/32                         Swap
524284                                524287
192.168.14.2                          1/1/2

192.0.2.5/32                         Push
--                                    524283
192.168.12.2                          1/1/1

192.0.2.5/32                         Swap
524283                                524283
192.168.12.2                          1/1/1

192.0.2.6/32                         Push
--                                    524282
192.168.12.2                          1/1/1

192.0.2.6/32                         Swap
524282                                524282
192.168.12.2                          1/1/1

-----
No. of IPv4 Prefix Active Bindings: 11
=====
  
```

In the tunnel table, there are LDP LSPs to all other nodes:

```

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

=====
IPv4 Tunnel Table (Router: Base)
=====
Destination      Owner      Encap TunnelId  Pref  Nexthop      Metric
Color
  
```

```
-----
192.0.2.2/32      ldp      MPLS  65537   9      192.168.12.2  10
192.0.2.3/32      ldp      MPLS  65538   9      192.168.12.2  20
192.0.2.4/32      ldp      MPLS  65539   9      192.168.14.2  10
192.0.2.5/32      ldp      MPLS  65540   9      192.168.12.2  20
192.0.2.6/32      ldp      MPLS  65541   9      192.168.12.2  30
-----
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
=====
```

In order to signal PHP with LDP, implicit-null must be configured on the eLER.

```
# on PE-6:
configure {
  router "Base" {
    ldp {
      implicit-null-label true
    }
  }
}
```

The implicit-null is signaled immediately, all related labels are withdrawn and re-advertised with label value of 3. The new label would show up on PE-5 as a swap from the ingress label to an egress label of 3, although label 3 is not pushed on to the frame.

```
[/]
A:admin@PE-5# show router ldp bindings active prefixes prefix 192.0.2.6/32
=====
LDP Bindings (IPv4 LSR ID 192.0.2.5)
      (IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
  (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
--             3
192.168.56.2    1/1/2

192.0.2.6/32    Swap
524282         3
192.168.56.2    1/1/2
-----
No. of IPv4 Prefix Active Bindings: 2
```

---

## Import and export policies

The default label handling behavior is to originate label bindings for the system address and to propagate all FECs received. If this is not the desired behavior, an import/export policy can be applied. An LDP import policy impacts inbound filtering; an LDP export policy impacts outbound filtering. An export policy may be configured to control the set of LDP label bindings advertised by the LER (sending to LDP peers). As such, export policies are used to include additional FECs rather than filtering FECs from those advertised. An import policy can be used to control for which FECs a router will generate labels (accepting from LDP peers). This functionality is not unique to LDP; it can be used for RSVP-TE, OSPF, and IS-IS as well as others.

The policy can be global or LDP peer FEC prefix filtering, both for import and export. LDP peer FEC prefix filtering uses a similar policy context as the LDP global policies and works in addition to these global policies.

```
[/]  
A:admin@PE-1# tree flat detail | match import-pref  
configure groups group <string> router <string> ldp session-parameters peer <string | ipv4-  
address | ipv6-address> import-prefixes <string>  
configure groups group <string> router <string> ldp targeted-session import-prefixes <string>  
configure router <string> ldp session-parameters peer <ipv4-address | ipv6-address> import-  
prefixes <reference>  
configure router <string> ldp targeted-session import-prefixes <reference>
```

```
[/]  
A:admin@PE-1# tree flat detail | match export-pref  
configure groups group <string> router <string> ldp session-parameters peer <string | ipv4-  
address | ipv6-address> export-prefixes <string>  
configure groups group <string> router <string> ldp targeted-session export-prefixes <string>  
configure router <string> ldp session-parameters peer <ipv4-address | ipv6-address> export-  
prefixes <reference>  
configure router <string> ldp targeted-session export-prefixes <reference>
```

By default, no labels are generated for directly connected (local) interfaces. To change this behavior, an export policy is created and applied to the LDP instance. There is no configuration difference in defining an import and export policy.

A policy contains a list of entries (of which each has a number). An entry typically contains matching criteria (however, it is not required in cases where everything matches) and a corresponding action. Entries without an action are considered incomplete and are rendered inactive. When processing the policy, the router executes the specified action on the first matching statement; it does not process any further matches. For this reason, entries must be sequenced correctly from most to least specific.

The configuration of the LDP export policy for local interfaces is as follows:

```
# on all PEs:  
configure {  
  policy-options {  
    policy-statement "LDP-export" {  
      entry 10 {  
        from {  
          protocol {  
            name [direct]  
          }  
        }  
      }  
    }  
  }  
}
```

```

    }
    action {
        action-type accept
    }
}
}
}

```

There are 11 active LDP bindings before applying the export policy, as shown earlier.

The LDP export or import policy is applied to the LDP instance on the router, with the export or import keyword.

```

# on all PEs:
configure {
    router "Base" {
        ldp {
            export-policy ["LDP-export"]
        }
    }
}

```

When the export policy is applied, the active LDP binding table contains additional entries: the local interfaces of PE-x. In the following output, the 11 entries for the system prefixes are snipped; only the 7 additional prefixes are shown:

```

[/]
A:admin@PE-1# show router ldp bindings active prefixes ipv4

=====
LDP Bindings (IPv4 LSR ID 192.0.2.1)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
  (S) - Static          (M) - Multi-homed Secondary Support
  (B) - BGP Next Hop    (BU) - Alternate Next-hop for Fast Re-Route
  (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
  (C) - FEC resolved with class-based-forwarding
=====
LDP IPv4 Prefix Bindings (Active)
=====
Prefix                               Op
IngLbl                               EgrLbl
EgrNextHop                           EgrIf/LspId
-----
---snip---

192.168.12.0/30                       Pop
524281                                 --
--                                     --

192.168.14.0/30                       Pop
524280                                 --
--                                     --

192.168.23.0/30                       Swap
524279                                 524279
192.168.12.2                           1/1/1

```

```

192.168.25.0/30          Swap
524278                 524278
192.168.12.2           1/1/1

192.168.36.0/30       Swap
524277                 524277
192.168.12.2           1/1/1

192.168.45.0/30       Swap
524276                 524276
192.168.14.2           1/1/2

192.168.56.0/30       Swap
524275                 524275
192.168.12.2           1/1/1

-----
No. of IPv4 Prefix Active Bindings: 18
=====

```

## OAM

The following operations, administration, and maintenance operations can be launched on an LDP LSP (in classic CLI):

- oam lsp-ping
- oam lsp-trace

As an example, an LSP ping is sent from PE-1 to PE-6:

```

A:PE-1# oam lsp-ping prefix 192.0.2.6/32
LSP-PING 192.0.2.6/32: 80 bytes MPLS payload
Seq=1, send from intf int-PE-1-PE-2, reply from 192.0.2.6
      udp-data-len=32 ttl=255 rtt=4.29ms rc=3 (EgressRtr)

---- LSP 192.0.2.6/32 PING Statistics ----
1 packets sent, 1 packets received, 0.00% packet loss
round-trip min = 4.29ms, avg = 4.29ms, max = 4.29ms, stddev = 0.000ms

```

An LSP trace is sent from PE-1 to PE-6:

```

A:PE-1# oam lsp-trace prefix 192.0.2.6/32
lsp-trace to 192.0.2.6/32: 0 hops min, 0 hops max, 104 byte packets
1 192.0.2.2 rtt=3.58ms rc=8(DSRtrMatchLabel) rsc=1
2 192.0.2.3 rtt=3.89ms rc=8(DSRtrMatchLabel) rsc=1
3 192.0.2.6 rtt=4.63ms rc=3(EgressRtr) rsc=1

```

The return code (**rc**) is 8 for the LSRs and 3 for the eLER.

The detailed output for this LSP trace includes the interface IP address, the interface type, maximum receive unit (MRU), label and protocol; as follows:

```

A:PE-1# oam lsp-trace prefix 192.0.2.6/32 detail
lsp-trace to 192.0.2.6/32: 0 hops min, 0 hops max, 104 byte packets
1 192.0.2.2 rtt=2.44ms rc=8(DSRtrMatchLabel) rsc=1
   DS 1: ipaddr=192.168.23.2 ifaddr=192.168.23.2 iftype=ipv4Numbered MRU=1564
       label[1]=524282 protocol=3(LDP)
2 192.0.2.3 rtt=3.67ms rc=8(DSRtrMatchLabel) rsc=1
   DS 1: ipaddr=192.168.36.2 ifaddr=192.168.36.2 iftype=ipv4Numbered MRU=1564

```

```
label[1]=3 protocol=3(LDP)
3 192.0.2.6 rtt=4.51ms rc=3(EgressRtr) rsc=1
```

## LDP statistics

LDP-related statistics can be collected in files. On PE-1, file 1 is configured as follows:

```
# on PE-1:
configure {
  log {
    file 1 {
      rollover 5
      retention 1
      compact-flash-location {
        primary cf1
      }
    }
  }
}
```

The following accounting policy defines which statistics should be recorded:

```
# on PE-1:
configure {
  log {
    accounting-policy 1 {
      admin-state enable
      record combined-ldp-lsp-egress
      destination {
        file 1
      }
    }
  }
}
```

The collection of statistics for prefix 192.0.2.6/32 is enabled on PE-1 in the **ldp** context, as follows:

```
# on PE-1:
configure {
  router "Base" {
    ldp {
      egress-statistics {
        fec-prefix 192.0.2.6/32 {
          admin-state enable
          accounting-policy 1
          collect-stats true
        }
      }
    }
  }
}
```

FEC egress statistics can be displayed with the following **show** command:

```
[/]
A:admin@PE-1# show router ldp fec-egress-stats
```

The FEC egress stats for prefix 192.0.2.6/32 can be retrieved as follows:

```
[/]
A:admin@PE-1# show router ldp fec-egress-stats 192.0.2.6/32
```

```
=====
LDP IPv4 FEC Egress Statistics
=====
```

```

-----
FEC Prefix/Mask      : 192.0.2.6/32
-----
Collect Stats       : Enabled           Accounting Plcy.   : 1
Admin State        : Up
FC BE
InProf Pkts        : 0                 OutProf Pkts      : 7
InProf Octets      : 0                 OutProf Octets    : 858
FC L2
InProf Pkts        : 0                 OutProf Pkts      : 0
InProf Octets      : 0                 OutProf Octets    : 0
FC AF
InProf Pkts        : 0                 OutProf Pkts      : 0
InProf Octets      : 0                 OutProf Octets    : 0
FC L1
InProf Pkts        : 0                 OutProf Pkts      : 0
InProf Octets      : 0                 OutProf Octets    : 0
FC H2
InProf Pkts        : 0                 OutProf Pkts      : 0
InProf Octets      : 0                 OutProf Octets    : 0
FC EF
InProf Pkts        : 0                 OutProf Pkts      : 0
InProf Octets      : 0                 OutProf Octets    : 0
FC H1
InProf Pkts        : 0                 OutProf Pkts      : 0
InProf Octets      : 0                 OutProf Octets    : 0
FC NC
InProf Pkts        : 0                 OutProf Pkts      : 0
InProf Octets      : 0                 OutProf Octets    : 0

Aggregate Packets   : 7
Aggregate Octets    : 858
=====
LDP IPv4 FEC Egress Statistics: 1
=====

```

Statistics can be cleared as follows on PE-1:

```

[/]
A:admin@PE-1# clear router ldp fec-egress-statistics 192.0.2.6/32

```

## Debug

LDP debugging can be configured—in classic CLI—per LDP interface or per LDP peer, as follows:

```

# on PE-1:
debug
  router "Base"
    ldp ?
  - ldp
  - no ldp

[no] interface      + Enable/disable and configure debugging for an LDP interface
[no] peer           + Enable/disable and configure debugging for an LDP peer

```

A particular peer is specified by its IPv4 or IPv6 address. It is possible to configure debugging for specific LDP events: bindings or messages, as follows:

```

# on PE-1:

```

```

debug
  router "Base"
    ldp
      peer 192.0.2.2 ?
  - no peer <ip-address>
  - peer <ip-address>

<ip-address>      : ipv4-address   - a.b.c.d
                   ipv6-address  - x:x:x:x:x:x:x:x (eight 16-bit pieces)
                                     x:x:x:x:x:x:d.d.d.d
                                     x - [0..FFFF]H
                                     d - [0..255]D

[no] event          + Configure debugging for specific LDP events
[no] packet         + Enable/disable debugging for specific LDP packets

```

```

# on PE-1:
debug
  router "Base"
    ldp
      peer 192.0.2.2
      event ?
  - event
  - no event

[no] bindings      - Enable/disable debugging for LDP bindings
[no] messages      - Enable/disable debugging for LDP messages

```

It is also possible to configure debugging for specific packets, such as label packets:

```

# on PE-1:
debug
  router "Base"
    ldp
      peer 192.0.2.2
      packet ?
  - no packet
  - packet

[no] hello         - Enable/disable debugging for LDP Hello packets
[no] init          - Enable/disable debugging for LDP Init packets
[no] keepalive     - Enable/disable debugging for LDP Keepalive packets
[no] label         - Enable/disable debugging for LDP Label packets

```

The following debugging is configured on PE-1:

```

# on PE-1:
debug
  router "Base"
    ldp
      peer 192.0.2.2
      packet
      label detail
    exit
  exit
exit

```

Some label mapping packets sent to peer 192.0.2.2 are the following, with the label mapping for prefixes 192.0.2.1/32 and 192.0.2.4/32:

```

---snip---

```



```
5 2021/02/18 15:52:15.241 CET MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Label Mapping packet (msgId 14) to 192.0.2.2:0
Protocol version = 1
Label 524284 advertised for the following FECs
Prefix Address Family = 1 Prefix = 192.0.2.4/32
"
4 2021/02/18 15:52:04.040 CET MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Label Mapping packet (msgId 12) to 192.0.2.2:0
Protocol version = 1
Label 524285 advertised for the following FECs
Prefix Address Family = 1 Prefix = 192.0.2.3/32
"
3 2021/02/18 15:52:04.040 CET MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Recv Label Mapping packet (msgId 12) from 192.0.2.2:0
Protocol version = 1
Label 524285 advertised for the following FECs
Prefix Address Family = 1 Prefix = 192.0.2.3/32
"
2 2021/02/18 15:51:48.931 CET MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Recv Label Mapping packet (msgId 6) from 192.0.2.2:0
Protocol version = 1
Label 524287 advertised for the following FECs
Prefix Address Family = 1 Prefix = 192.0.2.2/32
"
1 2021/02/18 15:51:48.681 CET MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Label Mapping packet (msgId 6) to 192.0.2.2:0
Protocol version = 1
Label 524287 advertised for the following FECs
Prefix Address Family = 1 Prefix = 192.0.2.1/32
"
```

## Conclusion

MPLS provides the capability to establish connection-oriented paths over a connectionless network. LDP point-to-point LSPs are dynamically signaled and FRR is supported. This can greatly improve network resiliency. In this chapter, the configuration of several LDP point-to-point LSP features is given together with the associated show output which can be used to verify and troubleshoot.

# LDP-IGP Synchronization

This chapter provides information about LDP-IGP synchronization

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

## Applicability

This chapter was initially written for SR OS Release 14.0.R6, but the MD-CLI in the current edition is based on SR OS Release 21.2.R1.

Label Distribution Protocol - Interior Gateway Protocol (LDP-IGP) synchronization based on RFC 5443 is supported in SR OS Release 6.0, and later. LDP end-of-Label Information Base (LIB), as described in RFC 5919, is supported in SR OS Release 14.0.R1, and later.

## Overview

Within an MPLS network using LDP, it is common practice to enable a synchronization timer between LDP and the IGP to give both the IGP and LDP time to converge after a link is restored. Without LDP-IGP synchronization, the IGP and LDP converge independently. Because the IGP converges before LDP, traffic can be black-holed until LDP has converged. When the IGP converges after link restoration and a new next hop is available, this change in next hop causes LDP to stop using the LDP labels for the alternate path. After the adjacency with the new next hop is established, labels are allocated for the new shortest (primary) path. These new labels are not yet signaled by LDP, causing the traffic to be black-holed for all or part of the FECs until LDP converges.

LDP-IGP synchronization based on RFC 5443 consists of temporarily setting the run-time IGP cost of a restored link to infinity to give time for both IGP and LDP to converge. When the LDP synchronization timer expires, the runtime IGP cost is restored to the configured IGP cost and IGP will re-advertise it and use this for the next shortest path first (SPF) computation. The value for infinity of the IGP cost for a router interface depends on the IGP: 0xFFFF (65535) for OSPF, 0x3F (63) for IS-IS regular metric, and 0xFFFFFE (16777214) for IS-IS wide metric. LDP-IGP synchronization is not supported on RIP interfaces.

When the system converges, the IGP starts the LDP synchronization timer when the LDP session to the neighbor is established over the interface. The LDP synchronization timer is running during the exchange of label FEC bindings over the interface. When the LDP synchronization timer expires, the IGP announces the new best next hop and LDP uses this next hop if the label bindings for the neighbor's FEC are available. However, the LDP synchronization timer does not guarantee that all FEC bindings will be exchanged when the timer expires. Operators do not want to configure very large timers on every node, which may result in long synchronization times. The end-of-lib option (RFC 5919) reduces the

synchronization time; therefore, operators can configure large synchronization timers that will be aborted when the end-of-lib notification has been received from a downstream node.

By default, LDP-IGP synchronization is enabled for OSPF and for IS-IS, as follows:

```
[/]
A:admin@PE-1# configure {
  router "Base" {
    ospf {
      info detail | match ldp-sync
    }
  }
}

shows:    ldp-sync true
```

```
[/]
A:admin@PE-1# configure {
  router "Base" {
    isis 0 {
      info detail | match ldp-sync
    }
  }
}

shows:    ldp-sync true
```

By default, LDP synchronization is disabled (out-of-service) on each interface, as follows:

```
[/]
A:admin@PE-1# show router ospf interface "int-PE-1-P-2" detail | match Ldp
Ldp Sync      : outOfService      Ldp Sync Wait   : Disabled
Ldp Timer State : Disabled        Ldp Tm Left     : 0
```

```
[/]
A:admin@PE-1# show router isis interface "int-PE-1-P-2" detail | match Ldp
Ldp Sync      : outOfService      Ldp Sync Wait   : Disabled
Ldp Timer State : Disabled        Ldp Tm Left     : 0
```

LDP end-of-lib, as defined in RFC 5919, allows a downstream node to notify its upstream peer that the node has advertised its entire LIB to its upstream peer, which can terminate the LDP synchronization timer. LDP end-of-lib notifications use a FEC TLV with the type wildcard FEC element for all negotiated FEC types. LDP end-of-lib is sent even if the system has no label bindings to advertise. Each node notifies its peer nodes that it is safe to send LDP end-of-lib notifications even if the node is not configured to process them. The node sends an unrecognized notification capability TLV (RFC 5919) in the initialization message, indicating that it will ignore notification messages that carry status TLV with a non-fatal status code unknown to it.

The LDP synchronization timer is configured in seconds with a maximum of 1800 seconds on a per interface basis, as follows:

```
[/]
A:admin@PE-1# configure {
  router "Base" {
    interface "int-PE-1-P-2" {
      ldp-sync-timer ?
    }
  }
}

ldp-sync-timer
```

```
end-of-lib          - Terminate LDP sync timer
seconds            - Value for the LDP SYNC timer
```

As an example, an LDP synchronization timer of 300 seconds can be configured on interface "int-PE-1-P-2", with or without the LDP end-of-lib option, as follows:

```
# on PE-1:
[/]
A:admin@PE-1# configure {
  router "Base" {
    interface "int-PE-1-P-2" {
      ldp-sync-timer {
        seconds 300
      }
    }
  }
}
```

```
# on PE-1:
[/]
A:admin@PE-1# configure {
  router "Base" {
    interface "int-PE-1-P-2" {
      ldp-sync-timer {
        seconds 300
        end-of-lib true
      }
    }
  }
}
```

- When the end-of-lib option is not configured, the LDP synchronization timer is started when the LDP hello adjacency comes up over the interface. Any received LDP end-of-lib message is ignored.
- When the end-of-lib option is configured, the receiving node behaves as follows:
  - The LDP synchronization timer is started when the LDP hello adjacency comes up over the interface.
  - When LDP end-of-lib type wildcard FEC messages have been received for all negotiated FEC types for a certain session to an LDP peer for the IGP interface, the LDP synchronization timer is terminated and the system restores the IGP link cost.
  - If the LDP synchronization timer expires before the LDP end-of-lib messages are received for all negotiated FEC types, the system restores the IGP link cost.
  - All unexpected LDP end-of-lib messages are dropped.
- When the end-of-lib option is configured, the sending node will advertise an LDP end-of-lib message for all FECs (prefix and P2MP FECs) after all FECs are sent for all peers that have advertised the unrecognized notification capability TLV.

When a user changes the IGP cost of an interface, the new value is advertised at the next flooding of link attributes by the IGP. If the LDP synchronization timer is running, the new cost value will only be advertised after the timer expires. However, the following **tools** or **configure** commands can be used to terminate the LDP-IGP synchronization, causing the new IGP cost value to be advertised instantly.

The following two **tools** commands do not modify the configuration; they terminate the LDP synchronization timer and restore the actual cost of the IGP interface:

```
[/]
A:admin@PE-1# tools perform router ospf ldp-sync-exit

[/]
A:admin@PE-1# tools perform router isis ldp-sync-exit
```

The following three commands disable the LDP-IGP synchronization entirely, either from the interface or globally for the IGP (OSPF or IS-IS):

```
# on PE-1:
[/]
A:admin@PE-1# configure {
  router "Base" {
    interface "int-PE-1-P-2" {
      delete ldp-sync-timer
    }
  }
}
```

```
[/]
A:admin@PE-1# configure {
  router "Base" {
    ospf 0 {
      ldp-sync false
    }
  }
}
```

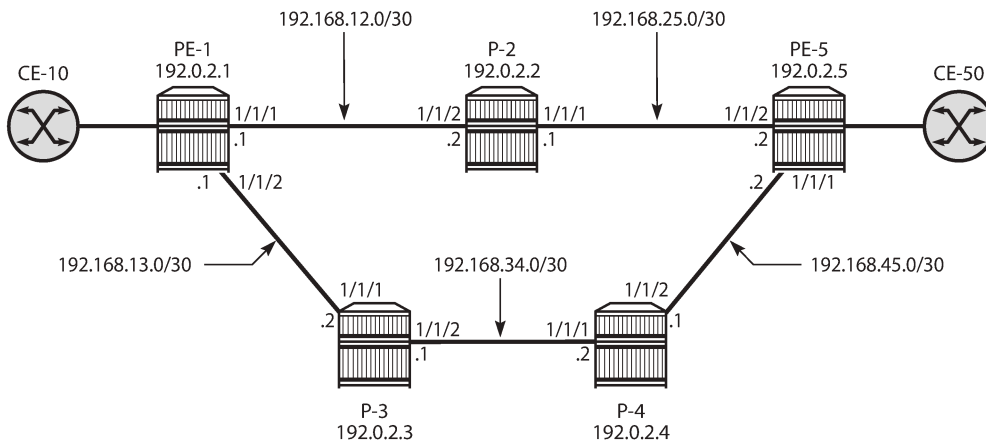
```
[/]
A:admin@PE-1# configure {
  router "Base" {
    isis 0 {
      ldp-sync false
    }
  }
}
```

If the user changes the value of the LDP synchronization timer parameter, the new value will take effect at the next synchronization event. If the timer is still running, it will continue to use the previous value.

## Configuration

[Figure 58: Example topology](#) shows the example topology.

Figure 58: Example topology



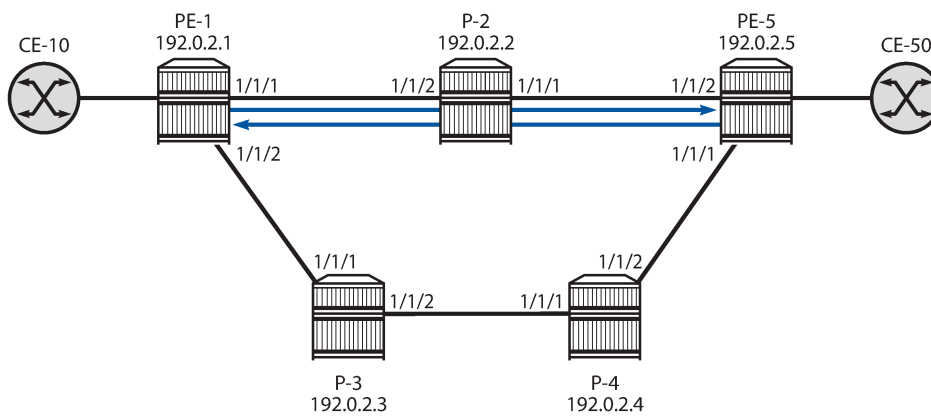
26227

The initial configuration on these nodes includes the following:

- Cards, MDAs, ports
- Router interfaces
- IGP: OSPF on all interfaces between the five P/PE routers (alternatively, IS-IS can be configured)
- LDP on all interfaces (LDP link adjacencies)
- Services on the PEs; for example, an Epipe between PE-1 and PE-5 (LDP targeted adjacencies)
- In this example topology, CE-10 and CE-50 correspond to VPRN\_10\_name on PE-1 and PE-5 using a hairpin to loop the traffic back to the node.

Default IGP metrics are used on the interfaces and, under normal conditions, traffic between CE-10 and CE-50 is sent over the shortest path via P-2, as shown in [Figure 59: Shortest path between PE-1 and PE-5](#).

Figure 59: Shortest path between PE-1 and PE-5



26228

## LDP-IGP synchronization without LDP end-of-lib

LDP-IGP synchronization is, by default, globally enabled for OSPF and IS-IS, but disabled on every interface. In this example, LDP-IGP synchronization will be configured with an LDP synchronization timer of 300 seconds on all the interfaces in all the nodes, as follows:

```
# on PE-1:
[/]
A:admin@PE-1# configure {
  router "Base" {
    interface "int-PE-1-P-2" {
      ldp-sync-timer {
        seconds 300
      }
    }
    interface "int-PE-1-P-3" {
      ldp-sync-timer {
        seconds 300
      }
    }
  }
}
```

```
# on P-2:
[/]
A:admin@P-2# configure {
  router "Base" {
    interface "int-P-2-PE-1" {
      ldp-sync-timer {
        seconds 300
      }
    }
    interface "int-P-2-PE-5" {
      ldp-sync-timer {
        seconds 300
      }
    }
  }
}
```

The configuration is similar on the other nodes. With this configuration, a restored interface will temporarily get an IGP cost of infinity; therefore, the link will not be used for data traffic until the LDP synchronization timer terminates (when it expires after 300 seconds or when it is terminated manually). To simulate a link failure, port 1/1/1 is disabled and re-enabled on PE-1, as follows:

```
# on PE-1:
[/]
A:admin@PE-1# configure {
  port 1/1/1 {
    admin-state disable
  }
}

[/]
A:admin@PE-1# configure {
  port 1/1/1 {
    admin-state enable
  }
}
```

```
}

```

The LDP synchronization timer is not started before the LDP hello adjacency is established. The following output shows the port re-enabled, but before the LDP adjacency is established (Ldp Timer State = Wait for Ldp Adj.):

```
[/]
A:admin@PE-1# show router ospf interface "int-PE-1-P-2" detail | match Ldp
Ldp Sync      : inService          Ldp Sync Wait   : Disabled
Ldp Timer State : Wait for Ldp Adj.   Ldp Tm Left     : 0

```

The following debug messages for OSPF show that the OSPF interface state is up (point-to-point), the LDP Sync Timer state is updated to "WAIT\_FOR\_ADJ", and afterward the LDP state is updated to "LDP\_INTF\_HAS\_ADJ", as follows:

```
21 2021/07/30 16:36:32.769 CEST MINOR: DEBUG #2001 Base OSPFv2
"OSPFv2: INTF
IF 192.168.12.1 Idx 2 Event: IF_UP state: from DOWN to PTP"

24 2021/07/30 16:36:32.769 CEST MINOR: DEBUG #2001 Base OSPFv2
"OSPFv2: INTF
Updated the LDP Sync Timer state for I/F 2 to WAIT_FOR_ADJ"

26 2021/07/30 16:36:32.869 CEST MINOR: DEBUG #2001 Base OSPFv2
"OSPFv2: INTF
OSPF I/F 2 LDP state: new LDP_INTF_HAS_ADJ old LDP_INTF_DOWN"

```

When the LDP hello adjacency is established, the interface between PE-1 and P-2 gets an IGP cost of infinity and the LDP synchronization timer is started, as follows:

```
27 2021/07/30 16:36:32.869 CEST MINOR: DEBUG #2001 Base OSPFv2
"OSPFv2: INTF
Updated the LDP Sync Timer state for I/F 2 to TMR_ACTIVE"

```

LDP bindings are exchanged as follows, but no message indicates the end-of-lib (and if it were sent by P-2, it would be ignored by PE-1). The LDP synchronization timer is not automatically terminated when the LDP bindings are received, because the configuration does not include the end-of-lib option.

```
29 2021/07/30 16:36:32.912 CEST MINOR: DEBUG #2001 Base LDP
"LDP: Binding
Sending Label mapping label 524287 for Prefix Address Family = 1 Prefix = 192.0.2.1/32 to peer
192.0.2.2:0."

35 2021/07/30 16:36:32.916 CEST MINOR: DEBUG #2001 Base LDP
"LDP: Binding
Sending Label mapping label 524284 for Prefix Address Family = 1 Prefix = 192.0.2.3/32 to peer
192.0.2.2:0."

37 2021/07/30 16:36:32.916 CEST MINOR: DEBUG #2001 Base LDP
"LDP: Binding
Sending Label mapping label 524283 for Prefix Address Family = 1 Prefix = 192.0.2.4/32 to peer
192.0.2.2:0."

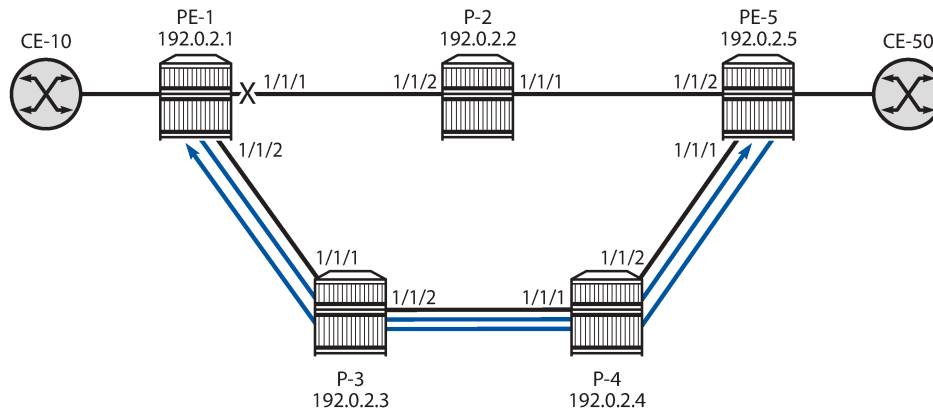
39 2021/07/30 16:36:32.916 CEST MINOR: DEBUG #2001 Base LDP
"LDP: Binding
Sending Label mapping label 524282 for Prefix Address Family = 1 Prefix = 192.0.2.5/32 to peer
192.0.2.2:0."

```



As long as the LDP synchronization timer is not terminated, traffic between CE-10 and CE-50 is redirected to the path via P-3 and P-4, as shown in [Figure 60: Rerouting via P-3 and P-4 until LDP synchronization timer terminates](#).

Figure 60: Rerouting via P-3 and P-4 until LDP synchronization timer terminates



26229

The following commands for the OSPF interfaces between PE-1 and P-2 show the Ldp Timer State = Timer Active, Ldp Sync Wait = Enabled; therefore, traffic is rerouted and the remaining time (Ldp Tm Left):

```
[/]
A:admin@PE-1# show router ospf interface "int-PE-1-P-2" detail | match Ldp
Ldp Sync      : inService          Ldp Sync Wait   : Enabled
Ldp Timer State : Timer Active      Ldp Tm Left     : 275
```

```
[/]
A:admin@P-2# show router ospf interface "int-P-2-PE-1" detail | match Ldp
Ldp Sync      : inService          Ldp Sync Wait   : Enabled
Ldp Timer State : Timer Active      Ldp Tm Left     : 263
```

The restored interface between PE-1 and P-2 will have an infinite IGP cost, so will not be used for data traffic as long as the LDP synchronization timer is active. All traffic between the CEs takes the path via P-3 and P-4, which can be verified as follows. The port statistics are cleared and 100 ICMP echo requests are sent by CE-10 to CE-50. On PE-1, port 1/1/1 is used toward P-2 and port 1/1/2 is used toward P-3. All traffic is expected to take the path toward P-3. However, there will be some IGP and LDP signaling on all interfaces, so the packet count will be greater than 100, as follows:

```
[/]
A:admin@PE-1# clear port 1/1/1 statistics
```

```
[/]
A:admin@PE-1# clear port 1/1/2 statistics
```

```
[/]
A:admin@PE-1# ping 172.16.10.2 router-instance "VPRN_10_name" count 100
PING 172.16.10.2 56 data bytes
---snip---

---- 172.16.10.2 PING Statistics ----
100 packets transmitted, 100 packets received, 0.00% packet loss
```

```
round-trip min = 3.38ms, avg = 4.40ms, max = 4.71ms, stddev = 0.249ms
```

```
[/]
A:admin@PE-1# show port 1/1/1 statistics
```

```
=====  
Port Statistics on Slot 1  
=====
```

Port Id	Ingress Packets Egress Packets	Ingress Octets Egress Octets
1/1/1	81 81	9005 9005

```
[/]
A:admin@PE-1# show port 1/1/2 statistics
```

```
=====  
Port Statistics on Slot 1  
=====
```

Port Id	Ingress Packets Egress Packets	Ingress Octets Egress Octets
1/1/2	275 275	29165 28984

The port statistics on the other nodes will also show that these packets are sent via P-3 and P-4 instead of via P-2.

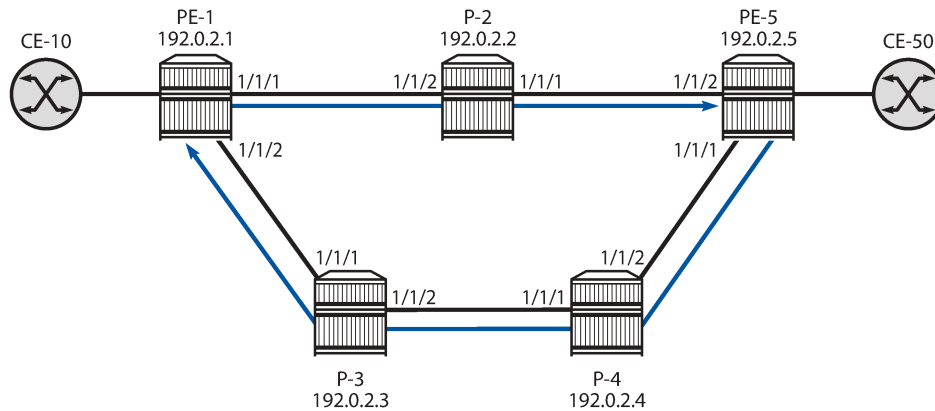
Even though the LIB was exchanged within seconds, the restored link only gets its normal IGP cost after the LDP synchronization timer has terminated. This can be done manually for a specific IGP (in this example, for OSPF on interface "int-PE-1-P-2" on PE-1) as follows:

```
[/]
A:admin@PE-1# tools perform router ospf ldp-sync-exit
Done.

[/]
A:admin@PE-1# show router ospf interface "int-PE-1-P-2" detail | match Ldp
Ldp Sync      : inService      Ldp Sync Wait   : Disabled
Ldp Timer State : Manual Exit      Ldp Tm Left     : 0
```

The LDP synchronization timer can be configured independently for each IGP on each interface. The LDP synchronization timer for OSPF on interface "int-PE-1-P-2" is terminated manually (Ldp Timer State = Manual Exit; Ldp Sync Wait = Disabled; Ldp Tm Left = 0). Traffic from CE-10 to CE-50 can use interface "int-PE-1-P-2" because that interface has its configured (default) IGP cost. However, traffic from CE-50 to CE-10 will not use interface "int-P-2-PE-1" because that interface still has an infinite IGP cost as long as the LDP synchronization timer is not terminated; therefore, traffic toward CE-10 will pass via P-3 instead. This leads to an asymmetric traffic flow: the shortest path from CE-10 to CE-50 is via P-2, while the shortest path from CE-50 to CE-10 is via P-4 and P-3, as shown in [Figure 61: Restored link with one LDP synchronization timer terminated](#).

Figure 61: Restored link with one LDP synchronization timer terminated



26230

When the second LDP synchronization timer is also terminated, the shortest path is via P-2 for all traffic between CE-10 and CE-50.

The LDP synchronization timer needs to be configured to a value that is long enough to prevent traffic being black-holed, but not too long to cause unnecessary suboptimal routing after the LIB has been exchanged and before the termination of the LDP synchronization timer. The end-of-lib option reduces the LDP synchronization time when the configured LDP synchronization timer is longer than required for the exchange of the LIB, as described in the next section.

LDP synchronization is disabled on the interfaces of PE-1, as follows:

```
# on PE-1:
[/]
A:admin@PE-1# configure {
  router "Base" {
    interface "int-PE-1-P-2" {
      delete ldp-sync-timer
    }
    interface "int-PE-1-P-3" {
      delete ldp-sync-timer
    }
  }
}
```

Similar commands to disable LDP synchronization on an interface can be configured on the other nodes.

## LDP-IGP synchronization with LDP end-of-lib

The LDP synchronization is configured with the end-of-lib option on all interfaces on all nodes; for example, for PE-1, as follows:

```
# on PE-1:
[/]
A:admin@PE-1# configure {
  router "Base" {
    interface "int-PE-1-P-2" {
      ldp-sync-timer {
        seconds 300
      }
    }
  }
}
```

```

        end-of-lib true
    }
}
interface "int-PE-1-P-3" {
    ldp-sync-timer {
        seconds 300
        end-of-lib true
    }
}
}
}
}

```

The configuration on the other nodes is similar.

A link failure is simulated by disabling and re-enabling port 1/1/1 on PE-1. Initially, the Ldp Timer State is "Wait for Ldp Adj.", as follows:

```

# on PE-1:
[/]
A:admin@PE-1# configure {
    port 1/1/1 {
        admin-state disable
    }
}

```

```

[/]
A:admin@PE-1# configure {
    port 1/1/1 {
        admin-state enable
    }
}

```

```

[/]
A:admin@PE-1# show router ospf interface "int-PE-1-P-2" detail | match Ldp
Ldp Sync      : inService      Ldp Sync Wait   : Disabled
Ldp Timer State : Wait for Ldp Adj.    Ldp Tm Left    : 0

```

After the LDP hello adjacency is established on the restored link, the LDP synchronization timer is started and PE-1 sends all LDP bindings to its peer P-2, as follows:

```

26 2021/07/30 16:49:16.069 CEST MINOR: DEBUG #2001 Base OSPFv2
"OSPFv2: INTF
OSPF I/F 2 LDP state: new LDP_INTF_HAS_ADJ old LDP_INTF_DOWN"

27 2021/07/30 16:49:16.069 CEST MINOR: DEBUG #2001 Base OSPFv2
"OSPFv2: INTF
Updated the LDP Sync Timer state for I/F 2 to TMR_ACTIVE"

29 2021/07/30 16:49:16.162 CEST MINOR: DEBUG #2001 Base LDP
"LDP: Binding
Sending Label mapping label 524287 for Prefix Address Family = 1 Prefix = 192.0.2.1/32 to peer
192.0.2.2:0."

35 2021/07/30 16:49:16.165 CEST MINOR: DEBUG #2001 Base LDP
"LDP: Binding
Sending Label mapping label 524284 for Prefix Address Family = 1 Prefix = 192.0.2.3/32 to peer
192.0.2.2:0."

37 2021/07/30 16:49:16.165 CEST MINOR: DEBUG #2001 Base LDP
"LDP: Binding

```

```
Sending Label mapping label 524283 for Prefix Address Family = 1 Prefix = 192.0.2.4/32 to peer
192.0.2.2:0."

39 2021/07/30 16:49:16.165 CEST MINOR: DEBUG #2001 Base LDP
"LDP: Binding
Sending Label mapping label 524282 for Prefix Address Family = 1 Prefix = 192.0.2.5/32 to peer
192.0.2.2:0."

43 2021/07/30 16:49:16.413 CEST MINOR: DEBUG #2001 Base OSPFv2
"OSPFv2: INTF
OSPF I/F 2 LDP state: new LDP_LBL_EXCH_DONE old LDP_INTF_HAS_ADJ"

44 2021/07/30 16:49:16.413 CEST MINOR: DEBUG #2001 Base OSPFv2
"OSPFv2: INTF
Updated the LDP Sync Timer state for I/F 2 to EXCH_DONE"
```

When a downstream node has sent its entire LIB to its upstream peer, the node sends an end-of-lib (RFC 5919) notification. When the upstream peer receives an end-of-lib notification from its downstream peer, LDP is considered to be fully operational for the link. LDP triggers the IGP to advertise the link with normal cost instead of infinity and transit traffic can be sent on the restored link. In the preceding debug messages, the LDP Sync Timer state changes to "EXCH\_DONE"; in the following show command output:

```
[/]
A:admin@PE-1# show router ospf interface "int-PE-1-P-2" detail | match Ldp
Ldp Sync      : inService          Ldp Sync Wait   : Disabled
Ldp Timer State : Label Exchg. Done  Ldp Tm Left    : 0
```

The LDP synchronization timer is terminated when the entire LIB is exchanged. In this example setup, the LDP synchronization time is reduced from 300 seconds to less than 10 seconds after enabling LDP end-of-lib.

## Conclusion

LDP-IGP synchronization (RFC 5443) allows directly connected nodes to delay the use of a restored link for transit IP packets until the LDP labels have been exchanged. RFC 5919 adds the end-of-lib option that reduces the LDP synchronization time to the minimum, so operators can configure large values for the LDP synchronization timer.

# LDP-SR Stitching for IPv4 Prefixes (IS-IS)

This chapter provides information about LDP-SR Stitching for IPv4 Prefixes (IS-IS).

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

## Applicability

This chapter was initially written for SR OS Release 14.0.R5. The MD-CLI in the current edition is based on SR OS Release 21.2.R1.

## Overview

Segment Routing (SR) allows for the construction of source-routed Label Switched Paths (LSPs) where the series of hops to be taken through the network are indicated by one or more Segment Identifiers (SIDs) assigned at the ingress PE. In the case of an MPLS data plane, these SIDs are MPLS labels learned through extensions to the OSPF/IS-IS control plane. SR provides benefits to the MPLS data plane, such as high scalability (due to lack of soft-state), traffic engineering capability, and topology-independent fast reroute.

When SR is configured in an IP/MPLS network that runs the Label Distribution Protocol (LDP), it is possible that SR and LDP will coexist, in which case preference for LDP or SR is a local matter at the LSP head end. It is equally possible that not all devices will have the capability to support SR, in which case some kind of interworking between SR and LDP is necessary to create an end-to-end LSP. Fast reroute coverage can also benefit from this SR-LDP interworking function, where SR is used to increase Loop Free Alternate (LFA) coverage using Remote or Directed LFA.

This chapter describes the configuration requirements for the interworking of LDP and SR to form a single end-to-end LSP when using IS-IS as an IGP. The chapter shows how this interworking function can be used to extend fast reroute coverage for LDP-based LSPs.

## Configuration

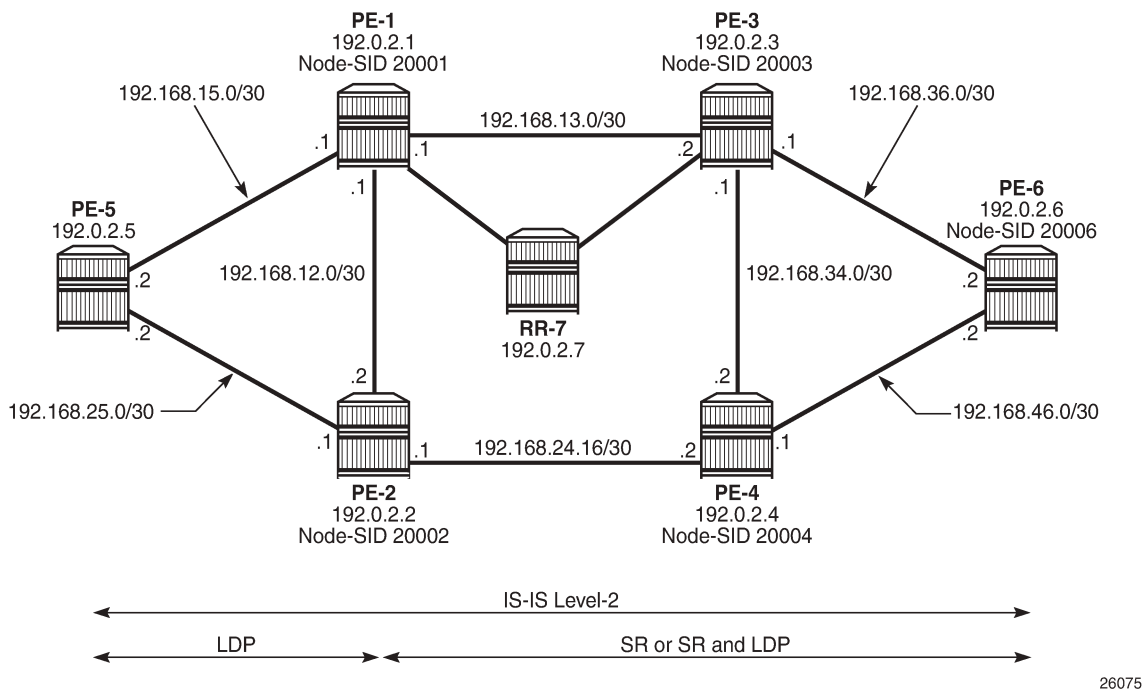
### Example topology

The topology shown in [Figure 62: Example topology](#) provides an example of SR-LDP interworking. All routers within the topology form part of Autonomous System 64496 and are IBGP clients of RR-7 for the

VPN-IPv4 address family. All routers in the topology belong to the same IS-IS Level-2 area, and all link metrics are set to 100. RR-7 does not participate in any MPLS data plane, and signals the IS-IS overload bit to avoid being used for transit traffic.

PE-5 is a router that does not support SR and, therefore, runs only LDP to its connected peers PE-1 and PE-2. PE-1, PE-2, PE-3, PE-4, and PE-6 are capable of running both SR and LDP, but are initially configured to only run SR with the associated node-SIDs shown in Figure 62: Example topology. When explicitly described, LDP will be enabled in conjunction with SR on these routers to show the difference between the two approaches, and to show how SR can be used as a fast reroute backup for SR primary LSPs.

Figure 62: Example topology



26075

The LDP configuration at PE-1 toward PE-5 is shown in the following output. The configuration at PE-2 is similar with the only exception being IP addressing.

```
# on PE-1:
configure {
  router "Base" {
    ldp {
      interface-parameters {
        interface "int-PE-1-PE-5" {
          ipv4 {
          }
        }
      }
    }
  }
}
```

PE-1, PE-2, PE-3, PE-4, and PE-6 run SR. The following output provides an example of the relevant SR configuration parameters at PE-1, with similar configurations on the remaining SR routers. For a

description of these parameters, see chapter "Segment Routing with IS-IS Control Plane" in *7750 SR and 7950 XRS Segment Routing and PCE Advanced Configuration Guide for MD CLI*.

```
# on PE-1:
configure {
  router "Base" {
    mpls-labels {
      sr-labels {
        start 20000
        end 20099
      }
    }
    isis 0 {
      admin-state enable
      advertise-router-capability as
      area-address [49.0001]
      segment-routing {
        admin-state enable
        prefix-sid-range {
          start-label 20000
          max-index 99
        }
      }
      interface "int-PE-1-PE-2" {
        interface-type point-to-point
        level-capability 2
        level 2 {
          metric 100
        }
      }
      interface "int-PE-1-PE-3" {
        interface-type point-to-point
        level-capability 2
        level 2 {
          metric 100
        }
      }
      interface "int-PE-1-PE-5" {
        interface-type point-to-point
        level-capability 2
        level 2 {
          metric 100
        }
      }
      interface "int-PE-1-RR-7" {
        interface-type point-to-point
        level-capability 2
        level 2 {
          metric 100
        }
      }
      interface "system" {
        passive true
        level-capability 2
        ipv4-node-sid {
          label 20001
        }
      }
      level 2 {
        wide-metrics-only true
      }
    }
  }
}
```



}

## SR Mapping Server

An SR Mapping Server (SR-MS) is an integral part of SR-LDP interoperability and has the responsibility for advertising prefixes to SID/label mappings on behalf of routers that do not support SR. When using IS-IS, a SID/label-binding TLV (TLV 149) containing a prefix-SID sub-TLV is used to advertise one or more SID index/labels and one or more prefixes. In the example topology, PE-4 is selected as the SR-MS and will advertise a prefix SID for the non-SR-capable router, PE-5.

The following output provides an example of the configuration required to implement SR-MS functionality. Under the **segment-routing** node, a **mapping-server** context is created that allows for origination of a SID/label-binding TLV and prefix-SID sub-TLV. The syntax begins with **node-sid-map** and is followed by an index. In SR OS, prefix-SIDs are always advertised with an index value (as opposed to an absolute label value), and the formula {start-label + SID index} is used to derive the label value. In this example, **index 5** is used and, therefore, the derived label value is {20000+5} 20005 for the PE-5 prefix 192.0.2.5/32.

An optional **prefix-sid-range** argument allows for advertisement of a contiguous range of prefixes and associated SIDs using the configured index/prefix as the beginning of the range. Non-contiguous ranges require multiple entries and are advertised as separate SID/label-binding TLVs. An additional optional **set-flags s** argument can also be used to set the S-flag, which controls the flooding scope. When set, the flooding scope is the entire IS-IS domain. When not set, the flooding scope is the IS-IS level into which the TLV was advertised.

```
# on PE-4:
configure {
  router "Base" {
    isis 0 {
      segment-routing {
        prefix-sid-range {
          start-label 20000
          max-index 99
        }
        mapping-server {
          admin-state enable
          node-sid-map 5 {
            ip-prefix 192.0.2.5/32
          }
        }
      }
    }
  }
}
```

The relevant part of the IS-IS LSP generated by PE-4, showing the SID/label-binding TLV, is shown in the following output:

```
[/]
A:admin@PE-4# show router isis database PE-4.00-00 detail

=====
Rtr Base ISIS Instance 0 Database (detail)
=====
---snip---
Displaying Level 2 database
-----
LSP ID      : PE-4.00-00                               Level      : L2
```

```

Sequence : 0x6          Checksum : 0x5102      Lifetime : 1172
Version  : 1           Pkt Type  : 20         Pkt Ver   : 1
Attributes: L1L2      Max Area  : 3         Alloc Len : 1492
SYS ID   : 1920.0000.2004  SysID Len : 6         Used Len  : 258
---snip---
TLVs :
  Area Addresses:
    Area Address : (3) 49.0001
  Supp Protocols:
    Protocols    : IPv4
  IS-Hostname   : PE-4
  Router ID     :
    Router ID    : 192.0.2.4
  Router Cap    : 192.0.2.4, D:0, S:0
    TE Node Cap : B E M P
    SR Cap: IPv4 MPLS-IPv6
    SRGB Base:20000, Range:100
    SR Alg: metric based SPF
    Node MSD Cap: BMI : 12 ERLD : 15
  SID Label Binding:
    Prefix: 192.0.2.5/32 Range:1 Weight:0 bFlgs:v4 SID:5 Algo:0 pFlgs:N
  ---snip---
Level (2) LSP Count : 1
-----
Control Info      : D = Prefix Leaked Down
                  S = Sub-TLVs Present
Attribute Flags  : N = Node Flag
                  R = Re-advertisement Flag
                  X = External Prefix Flag
                  E = Entropy Label Capability (ELC) Flag
Adj-SID Flags    : v4/v6 = IPv4 or IPv6 Address-Family
                  B = Backup Flag
                  V = Adj-SID carries a value
                  L = value/index has local significance
                  S = Set of Adjacencies
                  P = Persistently allocated
Prefix-SID Flags : R = Re-advertisement Flag
                  N = Node-SID Flag
                  nP = no penultimate hop POP
                  E = Explicit-Null Flag
                  V = Prefix-SID carries a value
                  L = value/index has local significance
Lbl-Binding Flags: v4/v6 = IPv4 or IPv6 Address-Family
                  M = Mirror Context Flag
                  S = SID/Label Binding flooding
                  D = Prefix Leaked Down
                  A = Attached Flag
SABM-flags Flags: R = RSVP-TE
                  S = SR-TE
                  F = LFA
                  X = FLEX-ALGO
FAD-flags Flags:  M = Prefix Metric
=====

```

At other routers within the SR domain, the presence of the advertised prefix can be validated as shown in the following output taken at PE-1. The SRMS field is set to Y for prefix 192.0.2.5/32, indicating that the prefix was advertised by an SR-MS. (In the case of IS-IS, the prefix-SID is a sub-TLV of the SID/label-binding TLV and the "N" (node-SID) flag is set; therefore, it can be recognized as being advertised by a mapping server.) The Y is followed by an "(S)" flag, indicating that the SRMS prefix-SID is selected to be programmed. This indication is provided in case there are multiple advertisements for the same prefix and/or node-SID from different SR mapping servers that result in some kind of conflict or inconsistency. If

there are multiple mapping servers advertising the same prefix-SID, the advertising router with the lowest system/router ID is preferred.

```
[/]
A:admin@PE-1# show router isis prefix-sids

=====
Rtr Base ISIS Instance 0 Prefix/SID Table
=====
Prefix                               SID      Lvl/Typ  SRMS   AdvRtr
                               MT      Flags
-----
192.0.2.1/32                         1        2/Int.   N      PE-1
                               0      NnP
192.0.2.2/32                         2        2/Int.   N      PE-2
                               0      NnP
192.0.2.3/32                         3        2/Int.   N      PE-3
                               0      NnP
192.0.2.4/32                         4        2/Int.   N      PE-4
                               0      NnP
192.0.2.5/32                       5       2/Int. Y(S) PE-4
                               0     NnP
192.0.2.6/32                         6        2/Int.   N      PE-6
                               0      NnP
-----
No. of Prefix/SIDs: 6 (6 unique)
-----
SRMS : Y/N = prefix SID advertised by SR Mapping Server (Y) or not (N)
        S   = SRMS prefix SID is selected to be programmed
Flags: R   = Re-advertisement
        N   = Node-SID
        nP  = no penultimate hop POP
        E   = Explicit-Null
        V   = Prefix-SID carries a value
        L   = value/index has local significance
=====
```

## SR-LDP interworking

Interworking SR and LDP essentially consists of stitching an LDP FEC and an SR node-SID route for the same prefix. In the example topology, PE-1 and PE-2 will act as the SR-LDP interworking nodes.

In the LDP-to-SR data plane direction, LDP uses an **export-tunnel-table** command under the **ldp** context to reference a policy that defines which prefixes should be redistributed from the IS-IS/SR domain into LDP. When applied, the LDP process monitors the tunnel-table until it locates a /32 SR tunnel of type sr-isis that matches a prefix defined in the export policy. LDP then programs an LDP Incoming Label Map (ILM) entry and stitches it to the SR node-SID tunnel endpoint. The LDP process also originates a FEC for the prefix and advertises that FEC to its peers.

The following output provides an example of the route policy and application of the policy at PE-1 and PE-2. In this policy, PE-1 advertises LDP FECs to PE-5 for PE-3 (192.0.2.3), PE-4 (192.0.2.4), and PE-6 (192.0.2.6), provided that PE-1 has a /32 SR tunnel of type sr-isis in the tunnel-table for those same prefixes. PE-1 also programs an LDP ILM entry for each prefix and stitches it to the appropriate SR tunnel.

```
# on PE-1 and PE-2:
configure {
    policy-options {
        prefix-list "sr-domain-prefixes" {
```

```

    prefix 192.0.2.3/32 type exact {
    }
    prefix 192.0.2.4/32 type exact {
    }
    prefix 192.0.2.6/32 type exact {
    }
}
policy-statement "SR-to-LDP-policy" {
  entry 10 {
    from {
      prefix-list ["sr-domain-prefixes"]
      protocol {
        name [isis]
      }
    }
    to {
      protocol {
        name [ldp]
      }
    }
    action {
      action-type accept
    }
  }
}
}
router "Base" {
  ldp {
    export-tunnel-table ["SR-to-LDP-policy"]
  }
}
}

```

In the SR-to-LDP data plane direction, the **export-tunnel-table ldp** command within the **segment-routing** context is the only required configuration. Unlike the LDP-to-SR data plane direction, where policy is used to control which prefixes are stitched, in the SR-to-LDP direction, no policy is explicitly referenced because the SR-MS provides a network-wide policy for the prefixes that SR needs to stitch to a corresponding LDP FEC. With the **export-tunnel-table ldp** command applied, whenever a /32 LDP tunnel destination matches a prefix for which a prefix-SID sub-TLV was received from a mapping server, the SR ILM is stitched to the corresponding LDP tunnel endpoint.

The following output shows the configuration applied at PE-1 to implement SR-to-LDP data plane interworking:

```

# on PE-1:
configure {
  router "Base" {
    isis 0 {
      segment-routing {
        admin-state enable
        export-tunnel-table ldp
      }
    }
  }
}

```

With the required configuration in the SR-LDP interworking routers (PE-1 and PE-2), it is possible to validate the correct ILM entries. In the LDP-to-SR data plane direction, the following output shows the active LDP bindings at PE-1. Each of the entries for PE-3 (192.0.2.3), PE-4 (192.0.2.4), and PE-6 (192.0.2.6) have an "(I)" flag to indicate that the prefix has an SR-ISIS next-hop. Each entry also has

an ingress label and an egress label. The ingress label represents the LDP FEC advertised for the corresponding prefix (in this case, advertised only to PE-5). The egress label represents the SR node-SID for the same prefix. Therefore, a mapping exists between LDP FEC and SR node-SID.

```
[/]
A:admin@PE-1# show router ldp bindings active prefixes ipv4

=====
LDP Bindings (IPv4 LSR ID 192.0.2.1)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
  (S) - Static (M) - Multi-homed Secondary Support
  (B) - BGP Next Hop (BU) - Alternate Next-hop for Fast Re-Route
  (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
  (C) - FEC resolved with class-based-forwarding
=====
LDP IPv4 Prefix Bindings (Active)
=====
Prefix                               Op
IngLbl                               EgrLbl
EgrNextHop                           EgrIf/LspId
-----
192.0.2.1/32                          Pop
524287                                 --
--                                     --

192.0.2.3/32(I)                       Swap
524280                                 20003
192.168.13.2                          1/1/1:100

192.0.2.4/32(I)                       Swap
524279                                 20004
192.168.12.2                          1/1/3:100

192.0.2.5/32                          Push
--                                     524287
192.168.15.2                          1/1/2:100

192.0.2.6/32(I)                       Swap
524278                                 20006
192.168.13.2                          1/1/1:100

-----
No. of IPv4 Prefix Active Bindings: 5
=====
```

In the SR-to-LDP data plane direction, the following output, taken at PE-1, shows a dump of the SR database for next-hops resolved to LDP. There is a single entry with index 5 (label value 20005) advertised by the SR-MS for the PE-5 prefix 192.0.2.5. The final line of the entry shows that an LDP FEC is the SID next-hop for SR-LDP stitching. The tunnel LSP ID is 65537. The tunnel-table verifies that this is an LDP tunnel to PE-5 (192.0.2.5).

```
[/]
A:admin@PE-1# tools dump router isis sr-database nh-type ldp detail
```

```

=====
Rtr Base ISIS Instance 0 SR Database

Legend:
label stack is ordered from bottom-most to top-most
=====
-----
SID 5
-----
Label           : 20005                Adv System Id      : 1920.0000.2005
Prefix          : 192.0.2.5
Route Level     : 2                    MT Id              : 0
Rtm Preference  : 18                   Ttm Preference     : 0
Metric          : 0                    Last Action        : AddTnl
Num Ip NextHop  : 0                    Num SR-Tnl NextHop : 1
Mtu             : 0
Mtu Prim       : 0                    Mtu Backup         : -
Exclude from LFA : 0                   LFA Type           : -
Duplicate Pending : 0                  Tunnel Active State : Reported/Ack
SR Error        : SR_ERR_OK

NHOP: IP

IsTunl GIflD/  IfId/ PgId  IsAdv Label  IsLfaX
TunlType LspId
-----
192.0.2.5      Y    2    65537 0    0    0    0
-----

No. of Entries: 1

LDP = LDP FEC is the SID NH for SR-LDP stitching
=====

```

To verify that the data plane is intact from end-to-end, a VPRN service is configured at the non-SR-capable PE-5 and the SR-capable PE-6, each with a locally configured subnet that is used to test IP connectivity. The configuration of the VPRN at PE-5 is shown in the following output. The **auto-bind-tunnel** configuration uses a resolution filter allowing only **ldp** to be used to resolve BGP next-hops for VPN-IPv4 routes. Usually, this could be configured for **resolution any**, but this configuration shows that LDP is being used. The local IP address at PE-5 is 172.31.5.1/24.

```

# on PE-5:
configure {
  service {
    vprn "VPRN1-name" {
      admin-state enable
      service-id 1
      customer "1"
      bgp-ipvpn {
        mpls {
          admin-state enable
          route-distinguisher "64496:1"
          vrf-target {
            community "target:64496:1"
          }
          auto-bind-tunnel {
            resolution filter
            resolution-filter {
              ldp true
            }
          }
        }
      }
    }
  }
}

```

```

interface "Local-Subnet" {
  ipv4 {
    primary {
      address 172.31.5.1
      prefix-length 24
    }
  }
  sap 1/2/1:1 {
  }
}
}
}
}

```

The configuration of the VPRN at PE-6 is shown in the following output. Again, the **auto-bind-tunnel** configuration uses a resolution filter, but this time it is configured for **sr-isis**. It could be set to **resolution any**, so that the tunnel-table preference would resolve an LSP with the lowest preference/metric, but the resolution filter configuration again shows that SR is being used. The **auto-bind-tunnel** context allows the transport mechanism to be a local decision at service level. The local IP address at PE-6 is 172.31.6.1/24.



**Note:**

An alternative approach would be to configure the auto-bind-tunnel context for **resolution any**, then modify the tunnel-table preference for SR using the **tunnel-table-pref** command in the **segment-routing** context.

```

# on PE-6
configure {
  service {
    vprn "VPRN1-name" {
      admin-state enable
      service-id 1
      customer "1"
      bgp-ipvprn {
        mpls {
          admin-state enable
          route-distinguisher "64496:1"
          vrf-target {
            community "target:64496:1"
          }
          auto-bind-tunnel {
            resolution filter
            resolution-filter {
              sr-isis true
            }
          }
        }
      }
    }
  }
  interface "Local-Subnet" {
    ipv4 {
      primary {
        address 172.31.6.1
        prefix-length 24
      }
    }
    sap 1/2/1:1 {
    }
  }
}
}
}
}

```

A VPRN ping between 172.31.5.1 at PE-5 and 172.31.6.1 at PE-6 verifies that the data plane is intact:

```
[/]
A:admin@PE-5# ping 172.31.6.1 source-address 172.31.5.1 router-instance "VPRN1-name"
PING 172.31.6.1 56 data bytes
64 bytes from 172.31.6.1: icmp_seq=1 ttl=64 time=3.27ms.
64 bytes from 172.31.6.1: icmp_seq=2 ttl=64 time=3.31ms.
. 64 bytes from 172.31.6.1: icmp_seq=3 ttl=64 time=3.41ms.
. 64 bytes from 172.31.6.1: icmp_seq=4 ttl=64 time=3.40ms.
. 64 bytes from 172.31.6.1: icmp_seq=5 ttl=64 time=2.70ms.

---- 172.31.6.1 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 2.70ms, avg = 3.22ms, max = 3.41ms, stddev = 0.266ms
```

## SR and LDP coexistence

The previous example demonstrates the use of SR-LDP interworking when the SR domain runs only SR. A more common scenario is that SR will coexist with LDP, because LDP is already deployed and the SR deployment will be added. In this sub-section, PE-1, PE-2, PE-3, PE-4, and PE-6 are configured to run LDP in conjunction with SR. PE-5 remains the same as the previous sub-section in that it runs only LDP to its connected peers PE-1 and PE-2.

In the SR-to-LDP data plane direction, there is no notable change when LDP coexists in the SR domain. Whenever a /32 LDP tunnel destination matches a prefix for which a prefix-SID sub-TLV was received from a mapping server, the SR ILM is stitched to the corresponding LDP tunnel endpoint.

In the LDP-to-SR data plane direction, there is a significant change. If only SR is running within the SR domain, the LDP process monitors the tunnel-table and when a /32 SR tunnel of type sr-isis is found that matches a prefix in the (**export-tunnel-table**) export policy, LDP programs an LDP ILM and stitches it to the SR node-SID tunnel endpoint. However, if an LDP FEC exists for the same /32 prefix, SR OS will resolve the LDP ILM entry to the LDP FEC. This is because LDP attempts to resolve the prefix in the route table first before looking in the tunnel-table and, therefore, prefers the LDP tunnel to the SR tunnel.

The following output is taken at PE-1 when LDP and SR coexist in the SR domain. The previous version of this output (when LDP was not running in the SR domain) showed the prefixes for PE-3, PE-4, and PE-6 as known via an SR-ISIS next-hop, and the egress labels as node-SIDs. When LDP is active in conjunction with SR, the egress labels resolve to an LDP FEC.

```
[/]
A:admin@PE-1# show router ldp bindings active prefixes ipv4

=====
LDP Bindings (IPv4 LSR ID 192.0.2.1)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
  (S) - Static (M) - Multi-homed Secondary Support
  (B) - BGP Next Hop (BU) - Alternate Next-hop for Fast Re-Route
  (I) - SR-ISIS Next Hop (O) - SR-OSPF Next Hop
  (C) - FEC resolved with class-based-forwarding
=====
```



```

LDP IPv4 Prefix Bindings (Active)
=====
Prefix                               Op
IngLbl                               EgrLbl
EgrNextHop                           EgrIf/LspId
-----
192.0.2.1/32                         Pop
524287                               --
--                                   --

192.0.2.2/32                         Push
--                                   524287
192.168.12.2                         1/1/3:100

192.0.2.2/32                         Swap
524282                               524287
192.168.12.2                         1/1/3:100

192.0.2.3/32                         Push
--                                   524283
192.168.13.2                         1/1/1:100

192.0.2.3/32                         Swap
524280                               524283
192.168.13.2                         1/1/1:100

192.0.2.4/32                         Push
--                                   524279
192.168.12.2                         1/1/3:100

192.0.2.4/32                         Swap
524279                               524279
192.168.12.2                         1/1/3:100

192.0.2.5/32                         Push
--                                   524287
192.168.15.2                         1/1/2:100

192.0.2.5/32                         Swap
524283                               524287
192.168.15.2                         1/1/2:100

192.0.2.6/32                         Push
--                                   524278
192.168.13.2                         1/1/1:100

192.0.2.6/32                         Swap
524278                               524278
192.168.13.2                         1/1/1:100

-----
No. of IPv4 Prefix Active Bindings: 11
=====
  
```

That the LDP-to-SR data path resolves to LDP FECs rather than SR tunnels may result in an asymmetric data path. Taking the previously used VPRN service between PE-5 and PE-6 as an example:

- Traffic from PE-6 to PE-5 will use SR between PE-6 and one of the SR-LDP interworking gateways at PE-1 or PE-2, after which it will use LDP.
- Traffic from PE-5 to PE-6 will use LDP between ingress and egress. The interworking function between SR and LDP has no effect.

Both directions still use an MPLS data plane. However, the MPLS control plane differs in each direction.

## LDP fast reroute using SR tunnels

With the ability to interwork LDP and SR, primary LSPs signaled using LDP can select a remote LFA SR tunnel as backup. This provides the potential to increase fast reroute coverage. As with any other backup or fast reroute mechanism, the SR backup tunnel can be installed in the forwarding database before any failure, but can only be activated when the failure of the primary path has been detected.

The ability to detect a failure quickly forms a significant part of the overall reconvergence time and may require the use of failure detection mechanisms, such as Bidirectional Forwarding Detection (BFD), the 802.3ah Ethernet in the First Mile (EFM), or just Loss of Signal (LoS). These mechanisms are beyond the scope of this chapter.

To use SR as a backup for LDP, the **fast-reroute backup-sr-tunnel** command must be configured in the **ldp** context. The **export-tunnel-table** command previously described should also be present, and should reference a policy including all of the prefixes for which backup is required. There is no requirement for an SR-MS when using SR tunnels for LDP backup, nor is there a requirement to enable SR-to-LDP interworking using the **export-tunnel-table ldp** command within the **segment-routing** context.

The following output shows the configuration applied at PE-6. When this configuration is applied, if the LFA SPF does not find an adjacent IP next-hop prefix for an LDP FEC, but can compute a remote LFA tunnel next-hop, LDP programs the LDP FEC using an LDP Next-Hop Label Forwarding Entry (NHLFE), and a backup next-hop using an LDP NHLFE pointing to the SR tunnel endpoint. The LDP packet is not tunneled over the SR tunnel, but rather the LDP label is stitched to the segment-routing label stack. This behavior is similar to the LDP-SR interworking function previously described within this chapter, but is modified such that the stitching of an LDP ILM entry to an SR tunnel only takes place if no adjacent LFA next-hop could be found for the prefix.

```
# on PE-6:
configure {
  policy-options {
    prefix-list "sr-domain-prefixes" {
      prefix 192.0.2.0/24 type longer {
      }
    }
  }
  policy-statement "SR-to-LDP-policy" {
    entry 10 {
      from {
        prefix-list ["sr-domain-prefixes"]
        protocol {
          name [isis]
        }
      }
      to {
        protocol {
          name [ldp]
        }
      }
      action {
        action-type accept
      }
    }
  }
}
router "Base" {
  isis 0 {
    loopfree-alternate {
      remote-lfa {
      }
    }
  }
}
```

```

    }
  }
  ldp {
    export-tunnel-table ["SR-to-LDP-policy"]
    fast-reroute {
      backup-sr-tunnel true
    }
  }
}

```

With the preceding configuration in place at PE-6, it is possible to verify whether a backup exists for a specific prefix, using the command shown in the following output. In this example, the backup is displayed for the PE-6 adjacent neighbor PE-3 (192.0.2.3). There are two LSPs for the prefix 192.0.2.3/32; one is known via LDP and one is known via SR-ISIS, indicated in the protocol column. The entries are defined as follows:

- The first line of the LDP entry is the primary LSP with a next-hop of 192.168.36.1 using interface 1/1/2:100 (direct to PE-3).
- The second line of the LDP entry is the backup indicated by a "(B)" flag, with a next-hop of 192.168.46.1 using interface 1/1/1:100 (via PE-4). This backup is a basic LFA, which is possible to compute due to the example topology, or more explicitly the triangular mesh between PE-6, PE-4, and PE-3. Due to this topology, if the link between PE6 and PE3 fails, PE-6 can forward packets destined for PE-3 toward PE-4. PE-4 will then forward them directly toward PE-3, not return them to PE-6 (which would create a transient micro-loop until the next SPF is run).
- The first line of the SR-ISIS entry is the primary LSP with a next-hop of 192.168.36.1 using interface 1/1/2:100 (direct to PE-3).
- The second line of the SR-ISIS entry is the backup LSP indicated by the "(B)" flag, with a next-hop of 192.168.46.1 using interface 1/1/1:100 (via PE-4). Both the primary and backup LSPs use the label 20003, representing the PE-3 node-SID. As with the LDP backup entry, the SR-ISIS backup is a basic LFA.

```

[/]
A:admin@PE-6# show router fp-tunnel-table 1 192.0.2.3/32

```

```

=====
IPv4 Tunnel Table Display
Legend:
label stack is ordered from bottom-most to top-most
B - FRR Backup
=====

```

Destination	Protocol	Tunnel-ID
Lbl		
NextHop		Intf/Tunnel
Lbl (backup)		
NextHop (backup)		
-----		
192.0.2.3/32	LDP	-
524283		
192.168.36.1		1/1/2:100
524281		
192.168.46.1(B)		1/1/1:100
192.0.2.3/32	SR-ISIS-0	524291
20003		
192.168.36.1		1/1/2:100
20003		
192.168.46.1(B)		1/1/1:100

```
-----
Total Entries : 2
-----
=====
```

To show the benefits that SR provides in increasing fast reroute coverage, the link between PE-4 and PE-3 is removed from the example topology, creating a ring topology. With this link removed, it is no longer possible for PE-6 to compute a basic LFA to PE-3 for the link between PE-6 and PE-3. If that link failed and PE-6 forwarded packets destined for PE-3 toward PE-4, PE-4 would return them to PE-6 until the next SPF was complete. Therefore, a backup tunnel is needed to a place in the network that will not loop packets back; essentially a remote LFA.

The following output at PE-6 shows the primary and backup LSPs for PE-3 (192.0.2.3) with the modified topology. Again, there are two LSPs: one known through via LDP and one known via SR-ISIS. The entries are defined as follows:

- The first line of the LDP entry is the primary LSP with a next-hop of 192.168.36.1 using interface 1/1/2:100 (direct to PE-3).
- The second line of the LDP entry is the backup indicated by a "(B)" flag with a next-hop of 192.0.2.3 (PE-3), which uses an SR tunnel. The label of "3" (implicit-null) indicates that the LDP label is not tunneled through the SR tunnel, but rather popped before the primary LDP LSP is stitched to the backup SR LSP.
- The first line of the SR-ISIS entry is the primary LSP with a next-hop of 192.168.36.1 using interface 1/1/2:100 (direct to PE-3). This LSP assigns a single label of value 20003, representing the node-SID of PE-3.
- The second line of the SR-ISIS entry is the backup indicated by a "(B)" flag with a next-hop of 192.168.46.1 using interface 1/1/1:100 (via PE-4). There are two labels assigned to this backup tunnel. The upper label has a value of 20002, which represents the node-SID of PE-2. This is the remote LFA "PQ-node". The second label has a value of 20003, which represents the node-SID of the destination, PE-3.

When this backup tunnel is operational, PE-6 encapsulates traffic destined for PE-3 to a point in the network where it will not be looped back toward the source. In the example topology, that node is PE-2. When traffic arrives at PE-2, it pops the top label (20002) and forwards traffic for PE-3 (with label 20003) on the shortest path toward the destination.

```
[/]
A:admin@PE-6# show router fp-tunnel-table 1 192.0.2.3/32

=====
IPv4 Tunnel Table Display

Legend:
label stack is ordered from bottom-most to top-most
B - FRR Backup
=====
Destination                                Protocol  Tunnel-ID
  Lbl
  NextHop
  Lbl      (backup)
  NextHop  (backup)
-----
192.0.2.3/32                                LDP      -
524283
  192.168.36.1                              1/1/2:100
  3
  192.0.2.3(B)                              SR
```

```
192.0.2.3/32          SR-ISIS-0          524291
20003
 192.168.36.1        1/1/2:100
20003/20002
 192.168.46.1(B)    1/1/1:100
-----
Total Entries : 2
-----
=====
```

## Conclusion

The SR control plane can (and likely will) coexist with other MPLS control plane clients, such as RSVP, LDP, or BGP. It is possible that these control plane clients will operate independently. However, where a mix of SR-capable and non-SR-capable routers exist within the same domain, SR-LDP interworking is necessary to form an end-to-end LSP. This chapter shows how that is possible using one or more SR mapping servers and one or more interworking routers.

SR-LDP interworking also provides an opportunity to increase fast reroute coverage in LDP-based networks. Before the introduction of SR-LDP interworking, a remote LFA could only be constructed using LDP-over-RSVP, which required the RSVP LSP to be manually configured and placed. When SR-LDP interworking is used, primary LDP LSPs can use a backup tunnel to a remote LFA signaled using SR. This requires no manual configuration, which provides the potential to greatly increase fast reroute coverage with minimal effort.

## MPLS LDP FRR using ISIS as IGP

This chapter describes Multi- Protocol Label Switching (MPLS) Label Distribution Protocol (LDP) Fast Reroute (FRR) using Intermediate System to Intermediate System (IS-IS) as the Interior Gateway Protocol (IGP).

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

### Applicability

This chapter was initially written for SR OS Release 9.0.R6, but the configuration in the current edition corresponds to MD-CLI SR OS Release 21.2.R1. There are no prerequisites for this configuration.

### Overview

LDP FRR improves convergence in case of a single link or single node failure in the network. Convergence times will be in the order of tens of milliseconds. This is important to some application services, such as voice over IP (VoIP), which are sensitive to traffic loss when running over the MPLS network.

Without FRR, link and/or node failures inside an MPLS LDP network result in traffic loss in the order of hundreds of milliseconds. The reason for that is that LDP depends on the convergence of the underlying IGP (IS-IS sending link state PDUs (LSPs) in this case). After IGP convergence, LDP itself needs to compute new primary Next Hop Label Forwarding Entries (NHLFEs) for all affected Forwarding Equivalence Classes (FECs). Finally, the different Label Forwarding Information Bases (LFIBs) are updated.

When FRR is configured on a node, the node computes primary NHLFEs for all FECs and, in addition, it will compute backup NHLFEs for all FECs. The backup NHLFE corresponds to the label received for the same FEC from a Loop-Free Alternate (LFA) next hop, see RFC 5286, *Basic Specification for IP Fast Reroute: Loop-Free Alternates*. Both primary NHLFEs and backup NHLFEs are programmed in the IOM/IMM, which makes it possible to converge very quickly.

The SR OS software has implemented Inequality 1 (link criterion) and Inequality 3 (node criterion) of RFC 5286. Similar to the Shortest Path Tree (SPT) computation that is part of standard link-state routing functionality, also the LFA next hop computation is based on the IGP metric.

The underlying LFA formulas appear in the following format:

#### Inequality 1:

$$SP(\text{backup NHR}, D) < \{SP(\text{backup NHR}, S) + SP(S, D)\}$$

### Inequality 3:

$$SP(\text{backup NHR}, D) < \{SP(\text{backup NHR}, PN) + SP(PN, D)\}$$

In these inequalities 'SP' is 'shortest IGP metric path', 'NHR' is 'next hop router', 'D' is 'destination', 'S' is 'source node or upstream node doing the actual LFA next-hop computation', and 'PN' is 'protected node'. The Inequality 3 rule is stricter than the Inequality 1 rule. See [Additional topics](#) for a practical example on these inequalities.

## Configuration

This section includes the following:

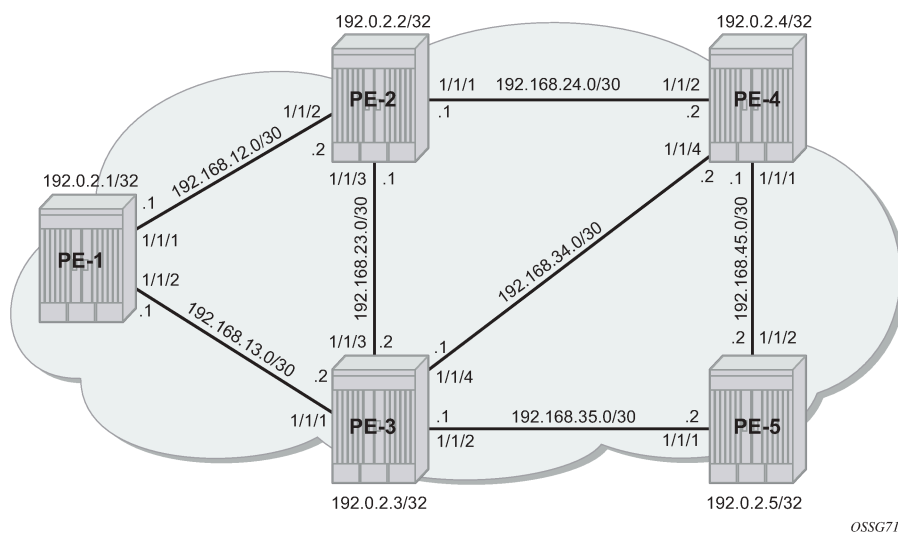
- [Configure the IP/MPLS network](#)
- [Enable LDP FRR and verify](#)
- [Enable synchronization timer](#)
- [Verify data path](#)

The subsection [Additional topics](#) includes:

- [Metric change](#)
- [IS-IS overload](#)

[Figure 63: Initial example topology](#) shows the example topology with five PEs in the same autonomous system.

*Figure 63: Initial example topology*



## Configure the IP/MPLS network

The system addresses and IP interface addresses are configured according to [Figure 63: Initial example topology](#). An interior gateway protocol (IGP) is needed to distribute routing information on all PEs. In this case, the IGP is IS-IS where each PE is acting as a level 2 router. On PE-1, the IS-IS configuration is as follows. The configuration is similar on the other PEs.

```
# on PE-1:
configure {
  router "Base" {
    isis 0 {
      admin-state enable
      level-capability 2
      area-address [49.0001]
      interface "int-PE-1-PE-2" {
        interface-type point-to-point
      }
      interface "int-PE-1-PE-3" {
        interface-type point-to-point
      }
      interface "system" {
      }
      level 2 {
        wide-metrics-only true
      }
    }
  }
}
```

IS-IS interfaces are set up as type point-to-point to improve convergence because no Designated Router/ Backup Designated Router (DR/BDR) election process is done. The **show router isis adjacency** command on PE-1 verifies that the IS-IS adjacencies are up:

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

=====
Rtr Base ISIS Instance 0 Adjacency
=====
System ID                Usage State Hold Interface          MT-ID
-----
PE-2                      L2    Up    27    int-PE-1-PE-2          0
PE-3                      L2    Up    20    int-PE-1-PE-3          0
-----
Adjacencies : 2
=====
```

The **show router route-table** command on PE-1 verifies which IP interface addresses or subnets are known on the PE:

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

=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]      Type  Proto  Age      Pref
Next Hop[Interface Name] Metric
-----
192.0.2.1/32           Local Local  00h03m34s  0
system                  0
```



```

192.0.2.2/32 Remote ISIS 00h02m59s 18
    192.168.12.2 10
192.0.2.3/32 Remote ISIS 00h02m11s 18
    192.168.13.2 10
192.0.2.4/32 Remote ISIS 00h01m34s 18
    192.168.12.2 20
192.0.2.5/32 Remote ISIS 00h01m00s 18
    192.168.13.2 20
192.168.12.0/30 Local Local 00h03m34s 0
    int-PE-1-PE-2 0
192.168.13.0/30 Local Local 00h03m34s 0
    int-PE-1-PE-3 0
192.168.23.0/30 Remote ISIS 00h02m59s 18
    192.168.12.2 20
192.168.24.0/30 Remote ISIS 00h02m59s 18
    192.168.12.2 20
192.168.34.0/30 Remote ISIS 00h02m11s 18
    192.168.13.2 20
192.168.35.0/30 Remote ISIS 00h02m11s 18
    192.168.13.2 20
192.168.45.0/30 Remote ISIS 00h01m34s 18
    192.168.12.2 30
-----
No. of Routes: 12
Flags: n = Number of times nexthop is repeated
      B = BGP backup route available
      L = LFA nexthop available
      S = Sticky ECMP requested
=====

```

The **show router fib 1** command on PE-1 shows the content of the forwarding information base (FIB):

```

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

=====
FIB Display
=====
Prefix [Flags] Protocol
NextHop
-----
192.0.2.1/32 LOCAL
    192.0.2.1 (system)
192.0.2.2/32 ISIS
    192.168.12.2 (int-PE-1-PE-2)
192.0.2.3/32 ISIS
    192.168.13.2 (int-PE-1-PE-3)
192.0.2.4/32 ISIS
    192.168.12.2 (int-PE-1-PE-2)
192.0.2.5/32 ISIS
    192.168.13.2 (int-PE-1-PE-3)
192.168.12.0/30 LOCAL
    192.168.12.0 (int-PE-1-PE-2)
192.168.13.0/30 LOCAL
    192.168.13.0 (int-PE-1-PE-3)
192.168.23.0/30 ISIS
    192.168.12.2 (int-PE-1-PE-2)
192.168.24.0/30 ISIS
    192.168.12.2 (int-PE-1-PE-2)
192.168.34.0/30 ISIS
    192.168.13.2 (int-PE-1-PE-3)
192.168.35.0/30 ISIS
    192.168.13.2 (int-PE-1-PE-3)

```

```

192.168.45.0/30                               ISIS
 192.168.12.2 (int-PE-1-PE-2)
-----
Total Entries : 12
-----
=====

```

Initially, the following default IS-IS Level 2 metric applies to all interfaces.

```

[/]
A:admin@PE-1# show router isis status | match "L2 Default Metric"
L2 Default Metric      : 10

```

The next step in the process of setting up the IP/MPLS network is setting up interface-LDP sessions on all interfaces. The LDP configuration on PE-1 is as follows:

```

# on PE-1:
configure {
  router "Base" {
    ldp {
      interface-parameters {
        interface "int-PE-1-PE-2" {
          ipv4 {
            admin-state enable
          }
        }
        interface "int-PE-1-PE-3" {
          ipv4 {
            admin-state enable
          }
        }
      }
    }
  }
}

```

There is now a full mesh of LDP label switched paths (LSPs) set up between all system interfaces of the PEs, and the tunnel table on PE-1 looks as follows:

```

[/]
A:admin@PE-1# show router tunnel-table
=====
IPv4 Tunnel Table (Router: Base)
=====
Destination      Owner      Encap TunnelId  Pref  Nexthop      Metric
  Color
-----
192.0.2.2/32     ldp        MPLS  65537    9     192.168.12.2  10
192.0.2.3/32     ldp        MPLS  65538    9     192.168.13.2  10
192.0.2.4/32     ldp        MPLS  65539    9     192.168.12.2  20
192.0.2.5/32     ldp        MPLS  65540    9     192.168.13.2  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
=====

```

The LDP LSP metric follows the IGP cost. Optionally, LSP metrics can be applied but that is beyond the scope for this chapter.

## Enable LDP FRR and verify

Because LDP FRR is using LFA next-hop computation by the IGP, as described in RFC 5286, LFA must be enabled in the IGP context, as follows:

```
# on PE-1:
configure {
  router "Base" {
    isis 0 {
      loopfree-alternate {
    }
  }
}
```

The **show router isis status** command on PE-1 verifies that LFA is enabled in IS-IS:

```
[/]
A:admin@PE-1# show router isis status | match "Loopfree"
Loopfree-Alternate      : Enabled
```

After enabling LFA inside the IGP context, FRR needs to be enabled within the LDP context, as follows:

```
# on PE-1:
configure {
  router "Base" {
    ldp {
      fast-reroute {
    }
  }
}
```

The **show router ldp status** command on PE-1 verifies that FRR is enabled in LDP:

```
[/]
A:admin@PE-1# show router ldp status | match FRR
FRR          : Enabled          Mcast Upstream FRR   : Disabled
Mcast Upst ASBR FRR: Disabled
```

This chapter describes FRR for unicast LDP. For multicast upstream FRR, see the "Multicast Label Distribution Protocol" chapter in the *7450 ESS, 7750 SR, 7950 XRS, and VSR MPLS Advanced Configuration Guide for Classic CLI*. After these two CLI commands, the software computes for each LDP FEC in the network both a primary and a backup NHLFE and uploads it to 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.

For point-to-point interfaces, when multiple LFA next hops are found for a primary next hop, the following selection criteria are used:

- It will pick the node-protect type in favor of the link-protect type.
- If there is more than one LFA next hop within the selected type, then it will pick one based on the lowest cost.
- If more than one LFA next hop with the same cost, SPF will select the first one. This is not a deterministic selection and will vary following each SPF calculation.

Several show commands are possible to display LFA information:

The **show router isis statistics** command shows the number of LFA runs on a specific node.

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

=====
Rtr Base ISIS Instance 0 Statistics
=====
---snip---

LFA Statistics
LFA Runs      : 1
  Last scheduled : 03/11/2021 08:49:32
Partial LFA Runs : 0

---snip---
```

Remote LFA (RLFA) statistics and Topology-independent LFA (TI-LFA) statistics have been removed from the preceding output, because they are beyond the scope of this chapter. RLFA and TI-LFA are used in segment routing and described in the "Segment Routing with IS-IS Control Plane" chapter in the *7450 ESS, 7750 SR, 7950 XRS, and VSR MPLS Advanced Configuration Guide for Classic CLI and Topology-Independent Loop-Free Alternate for Link Protection*.

The **show router isis lfa-coverage** command performs a mathematical calculation between the number of nodes and IPv4/IPv6 routes in the network versus present LFA next-hop protections. In the example topology (see [Figure 63: Initial example topology](#)), all IS-IS links have a default level 2 metric of 10. This results in all four nodes and all IS-IS routes learned by PE-1 being 100% LFA protected (link or node), as follows:

```
[/]
A:admin@PE-1# show router isis lfa-coverage

=====
Rtr Base ISIS Instance 0 LFA Coverage
=====
Topology          Level  Node          IPv4          IPv6
-----
IPV4 Unicast      L1     0/0(0%)      9/9(100%)    0/0(0%)
IPV6 Unicast      L1     0/0(0%)      0/0(0%)      0/0(0%)
IPV4 Multicast    L1     0/0(0%)      0/0(0%)      0/0(0%)
IPV6 Multicast    L1     0/0(0%)      0/0(0%)      0/0(0%)
IPV4 Unicast      L2     4/4(100%)    9/9(100%)    0/0(0%)
IPV6 Unicast      L2     0/0(0%)      0/0(0%)      0/0(0%)
IPV4 Multicast    L2     0/0(0%)      0/0(0%)      0/0(0%)
IPV6 Multicast    L2     0/0(0%)      0/0(0%)      0/0(0%)
=====
```

The **show router isis topology lfa detail** command shows the LFA protection type (link or node), as follows:

```
[/]
A:admin@PE-1# show router isis topology lfa detail

=====
Rtr Base ISIS Instance 0 Topology Table
=====
-----
IS-IS IP paths (MT-ID 0), Level 2
-----
Node      : PE-2.00
```

```

Nexthop   : PE-2
Interface : int-PE-1-PE-2
SNPA     : none                               Metric      : 10

LFA nh    : PE-3
LFA intf  : int-PE-1-PE-3                     LFA Metric  : 20
LFA type : linkProtection

Node      : PE-3.00
Nexthop   : PE-3
Interface : int-PE-1-PE-3
SNPA     : none                               Metric      : 10

LFA nh    : PE-2
LFA intf  : int-PE-1-PE-2                     LFA Metric  : 20
LFA type : linkProtection

Node      : PE-4.00
Nexthop   : PE-2
Interface : int-PE-1-PE-2
SNPA     : none                               Metric      : 20

LFA nh    : PE-3
LFA intf  : int-PE-1-PE-3                     LFA Metric  : 20
LFA type : nodeProtection

Node      : PE-5.00
Nexthop   : PE-3
Interface : int-PE-1-PE-3
SNPA     : none                               Metric      : 20

LFA nh    : PE-2
LFA intf  : int-PE-1-PE-2                     LFA Metric  : 30
LFA type : linkProtection

```

The **show router route-table** command adds an 'L' flag as reference that the associated prefix is having also an LFA next hop available.

```

[/]
A:admin@PE-1# show router route-table 192.0.2.4

=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]          Type  Proto  Age           Pref
  Next Hop[Interface Name]                Metric
-----
192.0.2.4/32 [L]           Remote ISIS   00h18m47s  18
  192.168.12.2                          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
=====

```

The **show router route-table alternative** command or **show router isis routes alternative** command show detailed interface address information used by the LFA calculation:

```
[/]
A:admin@PE-1# show router route-table 192.0.2.4 alternative

=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]                Type  Proto  Age      Pref
  Next Hop[Interface Name]          Metric
  Alt-NextHop                       Alt-
                                       Metric
-----
192.0.2.4/32                       Remote ISIS  00h18m47s 18
  192.168.12.2                       20
  192.168.13.2 (LFA)                 20
-----
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
=====
```

```
[/]
A:admin@PE-1# show router isis routes alternative ip-prefix-prefix-length 192.0.2.4

=====
Rtr Base ISIS Instance 0 Route Table (alternative)
=====
Prefix[Flags]                Metric  Lvl/Typ  Ver.  SysID/Hostname
  NextHop                    MT      AdminTag/SID[F]
  Alt-Nexthop                 Alt-    Alt-Type
                               Metric
-----
192.0.2.4/32                 20      2/Int.   5     PE-2
  192.168.12.2                 0       0
  192.168.13.2(L)             20      NP
-----
No. of Routes: 1 (1 path)
-----
Flags      : L = Loop-Free Alternate nexthop
Alt-Type   : LP = linkProtection, NP = nodeProtection
SID[F]       : R = Re-advertisement
              N = Node-SID
              nP = no penultimate hop POP
              E = Explicit-Null
              V = Prefix-SID carries a value
              L = value/index has local significance
=====
```

On PE-1, PE-4 (192.0.2.4/32) has a primary SPF next-hop pointing toward PE-2 (192.168.12.2) and an LFA next-hop pointing toward PE-3 (192.168.13.2).

The Inequality 3 formula on PE-1 for prefix 192.0.2.4/32 results in the following:

**Inequality 3:**

- $[SP(\text{backup NHR}, D) < \{SP(\text{backup NHR}, PN) + SP(PN, D)\}]$  or
- $[SP(\text{PE-3}, \text{PE-4}) < \{SP(\text{PE-3}, \text{PE-2}) + SP(\text{PE-2}, \text{PE-4})\}]$  or

- [10 < {10 + 10}]

This means that Inequality 3 is met. The calculated LFA next-hop for prefix 192.0.2.4/32 on PE-1 is protecting node PE-2, see [Figure 63: Initial example topology](#) for a graphical representation.

The **show router ldp bindings** command displays the Label Information Base (LIB). A BU flag is present in case the associated label is used as backup NHLFE for the prefix. As an example, a display on PE-1 for prefix PE-4 is as follows.

This is only possible because the SR OS LDP implementation is using liberal retention mode which means that every label mapping received by a peer is retained regardless of whether the LSR is the next hop for the advertised mapping.

```
[/]
A:admin@PE-1# show router ldp bindings prefixes prefix 192.0.2.4/32

=====
LDP Bindings (IPv4 LSR ID 192.0.2.1)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
=====
LDP IPv4 Prefix Bindings
=====
Prefix
Peer          FEC-Flags
IgrLbl        EgrLbl
EgrNextHop    EgrIntf/LspId
-----
192.0.2.4/32
192.0.2.2:0
524284N      524284
192.168.12.2 1/1/1

192.0.2.4/32
192.0.2.3:0
524284U      524284BU
192.168.13.2 1/1/2

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

The **show router ldp bindings active** command displays the label forwarding information base (LFIB). Also, the BU flag is present and, in addition, a reference to the label action itself: **pop** for eLER, **push** for iLER and **swap** for LSR.

```
[/]
A:admin@PE-1# show router ldp bindings active prefixes prefix 192.0.2.4/32

=====
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
-----
192.0.2.4/32                          Push
--                                     524284
192.168.12.2                           1/1/1

192.0.2.4/32                          Push
--                                     524284BU
192.168.13.2                           1/1/2

192.0.2.4/32                          Swap
524284                                 524284
192.168.12.2                           1/1/1

192.0.2.4/32                          Swap
524284                                 524284BU
192.168.13.2                           1/1/2

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

## Enable synchronization timer

Within an MPLS network using LDP, it is common practice to enable a synchronization timer between LDP and the IGP. Also, when LDP FRR is enabled, a situation can occur in which a synchronization timer between IGP and LDP will help: the revert scenario. When the interface for the previous primary next hop is restored, IGP may re-converge before LDP completed the FEC exchange with its neighbor over that interface. This may cause LDP to remove the LFA next hop from the FEC and blackhole traffic.

In order to avoid traffic being blackholed, it is recommended to first enable IGP-LDP synchronization on the interface. The time is expressed in seconds and can have a value between 1 and 1800 seconds. It is also possible to configure an end-of-LIB option to optimize the synchronization time, see the [LDP-IGP Synchronization](#) chapter. On PE-1, the following configures the LDP synchronization timer with a value of 10 seconds on the interfaces "int-PE-1-PE-2" and "int-PE-1-PE-3":

```

# on PE-1:
configure {
    router "Base" {
        interface "int-PE-1-PE-2" {
            port 1/1/1
            ldp-sync-timer {
                seconds 10
            }
        }
        ipv4 {
    
```



```
        primary {
            address 192.168.12.1
            prefix-length 30
        }
    }
}
interface "int-PE-1-PE-3" {
    port 1/1/2
    ldp-sync-timer {
        seconds 10
    }
    ipv4 {
        primary {
            address 192.168.13.1
            prefix-length 30
        }
    }
}
```

The configuration on the other nodes is similar.

When this timer is enabled, it means that when an interface is restored, the IGP will advertise this link in the network with an infinite metric. The **ldp-sync-timer** is started, LDP adjacencies are brought up together with a label exchange. After the **ldp-sync-timer** expires, the normal metric is advertised in the network again.

## Verify data path

Data path verification is performed using a Layer 2 Epipe service. Traffic generator ports are connected toward PE-1 and PE-5, and an Epipe service is created using an MPLS LDP based Service Distribution Path (SDP) on both PE-1 and PE-5. The service configuration on PE-1 is as follows:

```
# on PE-1:
configure {
    service {
        epipe "Epipe 1" {
            admin-state enable
            service-id 1
            customer "1"
            service-mtu 1450
            spoke-sdp 15:1 {
            }
            sap 1/1/3:1 {
            }
        }
        sdp 15 {
            admin-state enable
            delivery-type mpls
            ldp true
            far-end {
                ip-address 192.0.2.5
            }
        }
    }
}
```

The service configuration on PE-5 is similar.



```

=====
-----
IS-IS IP paths (MT-ID 0), Level 2
-----
---snip---

Node       : PE-5.00
Nexthop    : PE-3
Interface  : int-PE-1-PE-3
SNPA       : none                               Metric      : 20

LFA nh     : PE-2
LFA intf   : int-PE-1-PE-2                     LFA Metric   : 25
LFA type   : nodeProtection

=====
    
```

```

[/]
A:admin@PE-1# show router isis routes alternative ip-prefix-prefix-length 192.0.2.5

=====
Rtr Base ISIS Instance 0 Route Table (alternative)
=====
Prefix[Flags]          Metric    Lvl/Typ    Ver.  SysID/Hostname
NextHop                MT        AdminTag/SID[F]
Alt-Nexthop            Alt-      Alt-Type
Metric
-----
192.0.2.5/32           20        2/Int.     6     PE-3
    192.168.13.2        0         0
    192.168.12.2(L)    25        NP
-----
No. of Routes: 1 (1 path)
-----
Flags      : L = Loop-Free Alternate nexthop
Alt-Type   : LP = linkProtection, NP = nodeProtection
SID[F]     : R = Re-advertisement
            N = Node-SID
            nP = no penultimate hop POP
            E = Explicit-Null
            V = Prefix-SID carries a value
            L = value/index has local significance
=====
    
```

In normal conditions, MPLS traffic from PE-1 toward PE-5 over Epipe 1 will have two MPLS labels: an outer (transport) label given by LDP protocol, swapped on each intermediate LSR and an inner (service) label given by T-LDP, the same end-to-end. See the following show commands.

The T-LDP service label is S (524282):

```

[/]
A:admin@PE-1# show router ldp bindings services service-id 1

=====
LDP Bindings (IPv4 LSR ID 192.0.2.1)
(IPv6 LSR ID ::)
=====
Label Status:
    U - Label In Use, N - Label Not In Use, W - Label Withdrawn
    S - Status Signaled Up, D - Status Signaled Down, e - Label ELC
    WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
Service Type:
    
```

```

    E - Epipe Service, V - VPLS Service, M - Mirror Service
    A - Apipe Service, F - Fpipe Service, I - IES Service, R - VPRN service
    P - Ipipe Service, C - Cpipe Service
  FEC Flags:
    LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
    BA - ASBR Backup FEC
  =====
  LDP Service FEC 128 Bindings
  =====
  Type                VCId      SDPIId      LMTU
  Peer                SvcId     IngLbl      RMTU
                   EgrLbl
  -----
  E-Eth                1         15          1436
  192.0.2.5:0         1         524282U    1436
                   524282S
  -----
  No. of VC Labels: 1
  =====
  LDP Service FEC 129 Bindings
  =====
  SAII                AGII      IngLbl      LMTU
  TAI                Type      EgrLbl      RMTU
  Peer                SvcId     SDPIId
  -----
  No Matching Entries Found
  =====
  
```

The transport LDP label between PE-1 and PE-3 for prefix 192.0.2.5/32 is T13 (524283):

```

  [/]
  A:admin@PE-1# show router ldp bindings active prefixes prefix 192.0.2.5/32
  =====
  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
  -----
  192.0.2.5/32          Push
  --                    524283
  192.168.13.2         1/1/2
  192.0.2.5/32          Push
  
```

```

--
192.168.12.2          524283BU
                     1/1/1

192.0.2.5/32         Swap
524283               524283
192.168.13.2        1/1/2

192.0.2.5/32         Swap
524283               524283BU
192.168.12.2        1/1/1

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

```

The transport LDP label between PE-3 and PE-5 for prefix 192.0.2.5/32 is T35 (524287):

```

[/]
A:admin@PE-3# show router ldp bindings active prefixes prefix 192.0.2.5/32

=====
LDP Bindings (IPv4 LSR ID 192.0.2.3)
(IPv6 LSR ID ::)
=====
---snip---
=====
LDP IPv4 Prefix Bindings (Active)
=====
Prefix                Op
IngLbl                EgrLbl
EgrNextHop            EgrIf/LspId
-----
192.0.2.5/32          Push
--                    524287
192.168.35.2          1/1/2

192.0.2.5/32          Push
--                    524283BU
192.168.34.2          1/1/4

192.0.2.5/32          Swap
524283                524287
192.168.35.2          1/1/2

192.0.2.5/32          Swap
524283                524283BU
192.168.34.2          1/1/4

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

```

When PE-3 reboots, PE-1 performs an immediate swap to LFA next-hop for prefix 192.0.2.5/32 bypassing PE-3. The service label remains the same; only the transport labels can change on the network ports from PE-1 to PE-2, from PE-2 to PE-4, and from PE-4 to PE-5. See the following show commands.



**Note:**

The LDP FRR MPLS label stack will never contain more than two labels. This is different from RSVP-TE FRR facility mode which uses a three-label MPLS stack.

The T-LDP service label is S (524282):

```
[/]
A:admin@PE-1# show router ldp bindings services service-id 1

=====
LDP Bindings (IPv4 LSR ID 192.0.2.1)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  S - Status Signaled Up, D - Status Signaled Down, e - Label ELC
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
Service Type:
  E - Epipe Service, V - VPLS Service, M - Mirror Service
  A - Apipe Service, F - Fpipe Service, I - IES Service, R - VPRN service
  P - Ipipe Service, C - Cpipe Service
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
=====
LDP Service FEC 128 Bindings
=====
Type          VCId      SDPId      LMTU
Peer          SvcId     IngLbl     RMTU
              EgrLbl
-----
E-Eth          1         15         1436
192.0.2.5:0   1         524282U   1436
              524282S
-----
No. of VC Labels: 1
=====
---snip---
```

The transport LDP label value between PE-1 and PE-2 for prefix 192.0.2.5/32 is the same label (previously tagged as BU) as before the node failure event: T12 (524283):

```
[/]
A:admin@PE-1# show router ldp bindings active prefixes prefix 192.0.2.5/32

=====
---snip---
```

Prefix	Op
192.0.2.5/32	Push
--	524283
192.168.12.2	1/1/1
192.0.2.5/32	Swap
524283	524283
192.168.12.2	1/1/1

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

The transport LDP label between PE-2 and PE-4 for prefix 192.0.2.5/32 is T24 (524283):

```
[/]
A:admin@PE-2# show router ldp bindings active prefixes prefix 192.0.2.5/32

=====
---snip---
=====
LDP IPv4 Prefix Bindings (Active)
=====
Prefix                               Op
IngLbl                               EgrLbl
EgrNextHop                           EgrIf/LspId
-----
192.0.2.5/32                         Push
--                                   524283
192.168.24.2                         1/1/1

192.0.2.5/32                         Swap
524283                               524283
192.168.24.2                         1/1/1

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

The transport LDP label between PE-4 and PE-5 for prefix 192.0.2.5/32 is T45 (524287):

```
[/]
A:admin@PE-4# show router ldp bindings active prefixes prefix 192.0.2.5/32

=====
---snip---
=====
LDP IPv4 Prefix Bindings (Active)
=====
Prefix                               Op
IngLbl                               EgrLbl
EgrNextHop                           EgrIf/LspId
-----
192.0.2.5/32                         Push
--                                   524287
192.168.45.2                         1/1/1

192.0.2.5/32                         Swap
524283                               524287
192.168.45.2                         1/1/1

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

## Additional topics

### Metric change

On PE-4 and PE-5, the default level 2 metrics are restored, as follows:

```
[ex:/configure router "Base" isis 0 interface "int-PE-4-PE-5" level 2]
A:admin@PE-4# delete metric
```

```
[ex:/configure router "Base" isis 0 interface "int-PE-5-PE-4" level 2]
A:admin@PE-5# delete metric
```

When the IS-IS level 2 metric between PE-2 and PE-3 changes to 30, then 100% LFA coverage is no longer possible. The IS-IS level 2 metric is modified as follows:

```
[ex:/configure router "Base" isis 0 interface "int-PE-2-PE-3" level 2]
A:admin@PE-2# metric 30
```

```
[ex:/configure router "Base" isis 0 interface "int-PE-3-PE-2" level 2]
A:admin@PE-3# metric 30
```

On PE-1, Inequality 3 formula will find LFA next-hop coverages for prefix PE-4 and PE-5. Inequality formula 1 will find LFA next-hop coverages for prefix PE-4, PE-5, and the subnet between PE-4 and PE-5.

Both inequality formulas are visualized in [Figure 66: LFA computation: Inequality 3 for prefix PE-5 \(D\) on PE-1 \(S\)](#) and [Figure 65: LFA computation: Inequality 1 for prefix PE-5 \(D\) on PE-1 \(S\)](#) for prefix 192.0.2.5/32 (= PE-5) on PE-1 which serves as the source node for LFA next-hop computation.

Figure 65: LFA computation: Inequality 1 for prefix PE-5 (D) on PE-1 (S)

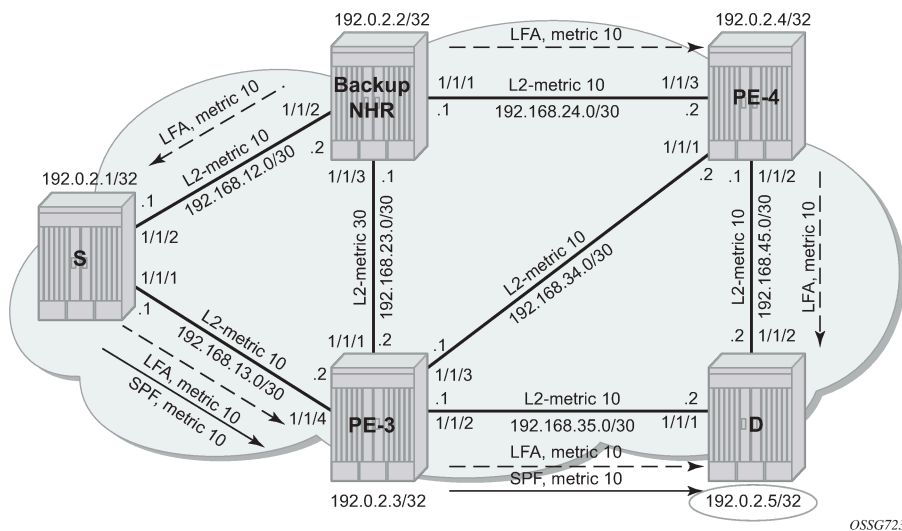
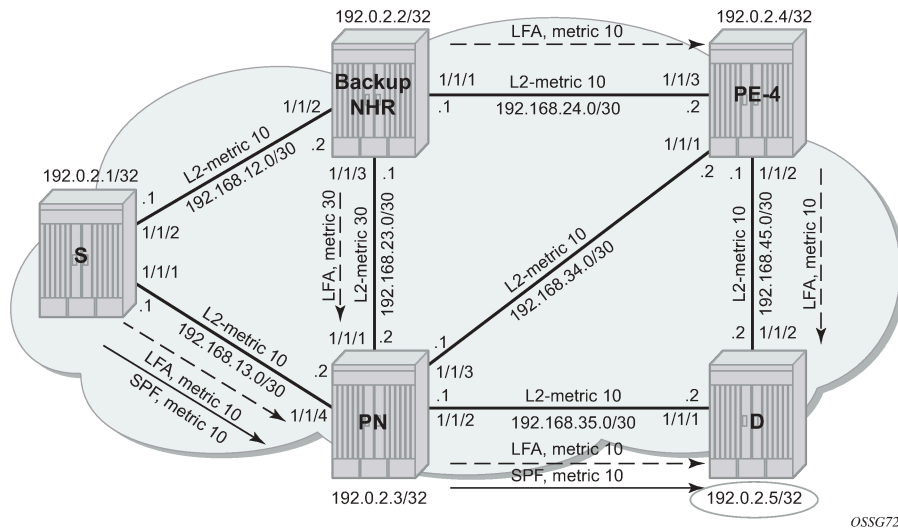




Figure 66: LFA computation: Inequality 3 for prefix PE-5 (D) on PE-1 (S)



Inequality 3 formula:

$$[SP(\text{backup NHR}, D) < \{SP(\text{backup NHR}, PN) + SP(PN, D)\}]$$

For a node LFA next-hop calculation of prefix 192.0.2.5/32 (D) on PE-1, this means that the shortest path from backup next-hop router PE-2 toward destination PE-5 must be smaller than the sum of the shortest path from backup next-hop router PE-2 toward protected node PE-3 with the shortest path from protected node PE-3 to destination PE-5.

The shortest path from backup next-hop router PE-2 toward destination PE-5 is going via PE-4, using IS-IS level 2 metric 10 for interface "int-PE-2-PE-4" and IS-IS level 2 metric 10 for interface "int-PE-4-PE-5". The shortest path from backup next-hop router (PE-2) toward protected node (PE-3) uses IS-IS level 2 metric 30 for interface "int-PE-2-PE-3". The shortest path from protected node (PE-3) to destination (PE-5) uses IS-IS level 2 metric 10 for interface "int-PE-3-PE-5". The calculation is as follows:

```
Prefix 192.0.2.5/32: SP (PE-2, PE-5) < SP (PE-2, PE-3) + SP (PE-3, PE-5)
                    10 + 10 < 30 + 10 => OK
```

Inequality 1 formula:

$$SP(\text{backup NHR}, D) < \{SP(\text{backup NHR}, S) + SP(S, D)\}$$

For a link LFA next-hop calculation of prefix 192.0.2.5/32 (D) on PE-1, this means that the shortest path from backup next-hop router PE-2 toward destination PE-5 must be smaller than the sum of the shortest path from backup next-hop router PE-2 toward source PE-1 with the shortest path from source PE-1 to destination PE-5.

The shortest path from backup next-hop router PE-2 toward destination PE-5 is going over PE-4, using IS-IS level 2 metric 10 for interface "int-PE-2-PE-4" and IS-IS level 2 metric 10 for interface "int-PE-4-PE-5". The shortest path from backup next-hop router PE-2 toward source PE-1 uses IS-IS level 2 metric 10 for interface "int-PE-2-PE-1". The shortest path from source PE-1 to destination PE-5 follows the normal SPF

calculation, going over PE-3, using IS-IS level 2 metric 10 for interface "int-PE-1-PE-3", and IS-IS level 2 metric 10 for interface "int-PE-3-PE-5".

The calculation is as follows:

```
Prefix 192.0.2.5/32 : SP(PE-2,PE-5) < SP(PE-2,PE-1) + SP(PE-1,PE-5)
                    10 + 10      <    10 + (10 + 10)          => OK
```

For completeness, all the other Inequality 1 calculations on PE-1 are as follows:

```
Prefix 192.0.2.2/32 : SP(PE-3,PE-2) < SP(PE-3,PE-1) + SP(PE-1,PE-2)          => NOK
                    30      <    10 + 10
Prefix 192.0.2.3/32 : SP(PE-2,PE-3) < SP(PE-2,PE-1) + SP(PE-1,PE-3)          => NOK
                    30      <    10 + 10
Prefix 192.0.2.4/32 : SP(PE-3,PE-4) < SP(PE-3,PE-1) + SP(PE-1,PE-4)          => OK
                    10      <    10 + (10 + 10)
Prefix 192.168.23.0/30 : SP(PE-3,D) < SP(PE-3,PE-1) + SP(PE-1,D)          => NOK
                    30      <    10 + (10 + 10)
Prefix 192.168.24.0/30 : SP(PE-3,D) < SP(PE-3,PE-1) + SP(PE-1,D)          => NOK
                    30 + 10 <    10 + (10 + 10)
Prefix 192.168.34.0/30 : SP(PE-2,D) < SP(PE-2,PE-1) + SP(PE-1,D)          => NOK
                    30 + 10 <    10 + (10 + 10)
Prefix 192.168.35.0/30 : SP(PE-2,D) < SP(PE-2,PE-1) + SP(PE-1,D)          => NOK
                    30 + 10 <    10 + (10 + 10)
Prefix 192.168.45.0/30 : SP(PE-3,D) < SP(PE-3,PE-1) + SP(PE-1,D)          => OK
                    10 + 10 <    10 + (10 + 10 + 10)
```

Considering all Inequality 1 calculations, only three of these are valid (OK).

In SR OS, the **show router isis lfa-coverage** summary command exists for LFA coverage on the router:

```
[/]
A:admin@PE-1# show router isis lfa-coverage

=====
Rtr Base ISIS Instance 0 LFA Coverage
=====
Topology      Level  Node      IPv4      IPv6
-----
IPV4 Unicast  L1    0/0(0%)   3/9(33%)  0/0(0%)
IPV6 Unicast  L1    0/0(0%)   0/0(0%)   0/0(0%)
IPV4 Multicast L1    0/0(0%)   0/0(0%)   0/0(0%)
IPV6 Multicast L1    0/0(0%)   0/0(0%)   0/0(0%)
IPV4 Unicast  L2    2/4(50%)  3/9(33%)  0/0(0%)
IPV6 Unicast  L2    0/0(0%)   0/0(0%)   0/0(0%)
IPV4 Multicast L2    0/0(0%)   0/0(0%)   0/0(0%)
IPV6 Multicast L2    0/0(0%)   0/0(0%)   0/0(0%)
=====
```

The **show router route-table alternative** command on PE-1 shows which prefixes are protected:

```
[/]
A:admin@PE-1# 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                               Local  Local  00h39m57s  0
  system
192.0.2.2/32                               Remote ISIS  00h39m21s  18
  192.168.12.2
192.0.2.3/32                               Remote ISIS  00h03m26s  18
  192.168.13.2
192.0.2.4/32                               Remote ISIS  00h37m57s  18
  192.168.12.2
  192.168.13.2 (LFA)
192.0.2.5/32                               Remote ISIS  00h03m26s  18
  192.168.13.2
  192.168.12.2 (LFA)
192.168.12.0/30                            Local  Local  00h39m57s  0
  int-PE-1-PE-2
192.168.13.0/30                            Local  Local  00h39m57s  0
  int-PE-1-PE-3
192.168.23.0/30                            Remote ISIS  00h00m48s  18
  192.168.12.2
192.168.24.0/30                            Remote ISIS  00h39m21s  18
  192.168.12.2
192.168.34.0/30                            Remote ISIS  00h03m26s  18
  192.168.13.2
192.168.35.0/30                            Remote ISIS  00h03m26s  18
  192.168.13.2
192.168.45.0/30                            Remote ISIS  00h02m29s  18
  192.168.12.2
  192.168.13.2 (LFA)
-----
No. of Routes: 12
Flags: n = Number of times nexthop is repeated
      Backup = BGP backup route
      LFA = Loop-Free Alternate nexthop
      S = Sticky ECMP requested
=====

```

The default IS-IS level 2 metrics are restored, as follows:

```
[ex:/configure router "Base" isis 0 interface "int-PE-2-PE-3" level 2]
A:admin@PE-2# delete metric
```

```
[ex:/configure router "Base" isis 0 interface "int-PE-3-PE-2" level 2]
A:admin@PE-3# delete metric
```

## IS-IS overload

As stated in RFC 3137, *OSPF Stub Router Advertisement*, sometimes it is useful and desirable for a router not to be a transit node. For those cases, it is also desirable not to have that router used as transit node during the LFA next-hop computation. Within the IS-IS protocol, this is achieved by configuring IS-IS overload. When other routers detect that IS-IS overload is configured, they will only use this router for packets destined to the overloaded router's directly connected networks and IP prefixes.

In the IS-IS overload context, the **max-metric** can be configured, as follows:

```
[ex:/configure router "Base" isis 0]
A:admin@PE-2# overload ?

overload
```

max-metric - Advertise transit links with maximum metric instead of setting overload bit

As an example, IS-IS overload is configured on PE-2, as follows:

```
*[ex:/configure router "Base" isis 0 overload]
A:admin@PE-2# max-metric false
```

With IS-IS overload on PE-2 configured **max-metric false**, the LFA coverage on PE-1 is as follows:

```
[/]
A:admin@PE-1# show router isis lfa-coverage

=====
Rtr Base ISIS Instance 0 LFA Coverage
=====
Topology          Level   Node           IPv4           IPv6
-----
IPV4 Unicast      L1      0/0(0%)        3/9(33%)      0/0(0%)
IPV6 Unicast      L1      0/0(0%)        0/0(0%)      0/0(0%)
IPV4 Multicast    L1      0/0(0%)        0/0(0%)      0/0(0%)
IPV6 Multicast    L1      0/0(0%)        0/0(0%)      0/0(0%)
IPV4 Unicast     L2     1/4(25%)      3/9(33%)    0/0(0%)
IPV6 Unicast      L2      0/0(0%)        0/0(0%)      0/0(0%)
IPV4 Multicast    L2      0/0(0%)        0/0(0%)      0/0(0%)
IPV6 Multicast    L2      0/0(0%)        0/0(0%)      0/0(0%)
=====
```

```
[/]
A:admin@PE-1# 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                                     Local  Local   00h45m01s    0
  system
192.0.2.2/32                                     Remote ISIS   00h44m26s 18
  192.168.12.2                                     10
  192.168.13.2 (LFA)                               20
192.0.2.3/32                                     Remote  ISIS    00h08m30s    18
  192.168.13.2
192.0.2.4/32                                     Remote  ISIS    00h00m29s    18
  192.168.13.2
192.0.2.5/32                                     Remote  ISIS    00h08m30s    18
  192.168.13.2
192.168.12.0/30                                  Local   Local   00h45m01s    0
  int-PE-1-PE-2
192.168.13.0/30                                  Local   Local   00h45m01s    0
  int-PE-1-PE-3
192.168.23.0/30                                  Remote ISIS   00h04m15s 18
  192.168.12.2                                     20
  192.168.13.2 (LFA)                               30
192.168.24.0/30                                  Remote ISIS   00h44m26s 18
  192.168.12.2                                     20
  192.168.13.2 (LFA)                               30
192.168.34.0/30                                  Remote  ISIS    00h08m30s    18
```

```

192.168.13.2                                20
192.168.35.0/30                            Remote ISIS 00h08m30s 18
192.168.13.2                                20
192.168.45.0/30                            Remote ISIS 00h00m29s 18
192.168.13.2                                30
-----
No. of Routes: 12
Flags: n = Number of times nexthop is repeated
      Backup = BGP backup route
      LFA = Loop-Free Alternate nexthop
      S = Sticky ECMP requested
=====

```

```

[/]
A:admin@PE-1# show router isis routes alternative

=====
Rtr Base ISIS Instance 0 Route Table (alternative)
=====
Prefix[Flags]           Metric    Lvl/Typ    Ver.  SysID/Hostname
NextHop                MT        AdminTag/SID[F]
Alt-Nexthop           Alt-     Alt-Type
                        Metric
-----
192.0.2.1/32            0         2/Int.     2     PE-1
0.0.0.0                 0         0
192.0.2.2/32           10        2/Int.     3     PE-2
192.168.12.2           0         0
192.168.13.2(L)       20        LP
192.0.2.3/32           10        2/Int.    16    PE-3
192.168.13.2           0         0
192.0.2.4/32           20        2/Int.    24    PE-3
192.168.13.2           0         0
192.0.2.5/32           20        2/Int.    16    PE-3
192.168.13.2           0         0
192.168.12.0/30       10        2/Int.     2     PE-1
0.0.0.0                 0
192.168.13.0/30       10        2/Int.    16    PE-1
0.0.0.0                 0
192.168.23.0/30       20        2/Int.    22    PE-2
192.168.12.2           0         0
192.168.13.2(L)       30        LP
192.168.24.0/30       20        2/Int.     3     PE-2
192.168.12.2           0         0
192.168.13.2(L)       30        LP
192.168.34.0/30       20        2/Int.    16    PE-3
192.168.13.2           0         0
192.168.35.0/30       20        2/Int.    16    PE-3
192.168.13.2           0         0
192.168.45.0/30       30        2/Int.    24    PE-3
192.168.13.2           0         0
-----
No. of Routes: 12 (12 paths)
-----
Flags      : L = Loop-Free Alternate nexthop
Alt-Type   : LP = linkProtection, NP = nodeProtection
SID[F]     : R = Re-advertisement
            N = Node-SID
            nP = no penultimate hop POP
            E = Explicit-Null
            V = Prefix-SID carries a value
            L = value/index has local significance
=====

```



# Multicast Label Distribution Protocol

This chapter provides information about multicast Label Distribution Protocol (mLDP).

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

## Applicability

This chapter is applicable to SR OS and was initially written for Release 13.0.R6. The CLI in the current edition corresponds to Release 23.7.R2.

In this chapter, the emphasis is on IPv4. However, multicast Label Distribution Protocol (mLDP) is also supported on IPv6 interfaces.

## Overview

Multicast Label Distribution Protocol provides extensions to LDP for the setup of point-to-multipoint (P2MP) Label Switched Paths (LSPs) and multipoint-to-multipoint (MP2MP) LSPs in MPLS networks.

The protocol is described in RFC 6388, *Label Distribution Protocol Extensions for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths*.

Multicast LSPs can be applied for IP multicast or support for multicast in BGP/MPLS Layer 3 Virtual Private Networks (L3 VPNs).

Compared to RSVP P2MP LSPs, mLDP P2MP LSPs are easier to configure and the setup direction is different. Whereas the RSVP P2MP LSPs are set up from the root node toward the leaf nodes, mLDP P2MP LSPs are set up from the leaf nodes toward the root node.

## P2MP Terminology

The following terminology is used.

*Table 3: Terminology*

Node	Description
Ingress / Root	P2MP LSPs have just one ingress (root) node. The root node receives IP multicast traffic and maps the traffic to a P2MP LSP (push). The node may perform MPLS multicast replication.

Node	Description
Egress / Leaf	P2MP LSPs have multiple egress (leaf) nodes. A leaf node removes data packets from a P2MP LSP (pop) for further processing. The node may perform IP multicast replication.
Transit	A transit Label Switching Router (LSR) can reach the root node via a directly connected upstream LSR. A transit LSR also has one or more directly connected downstream LSRs. The LSR swaps the MPLS label and may perform MPLS multicast replication.
Branch	A branch LSR is a transit LSR that has several directly connected LSRs. The LSR swaps the MPLS label and performs MPLS multicast replication.
Bud	A bud node is an egress node, but also a transit node. The node has directly connected receivers and also one or more directly connected downstream LSRs.

### Setup of mLDP P2MP LSP

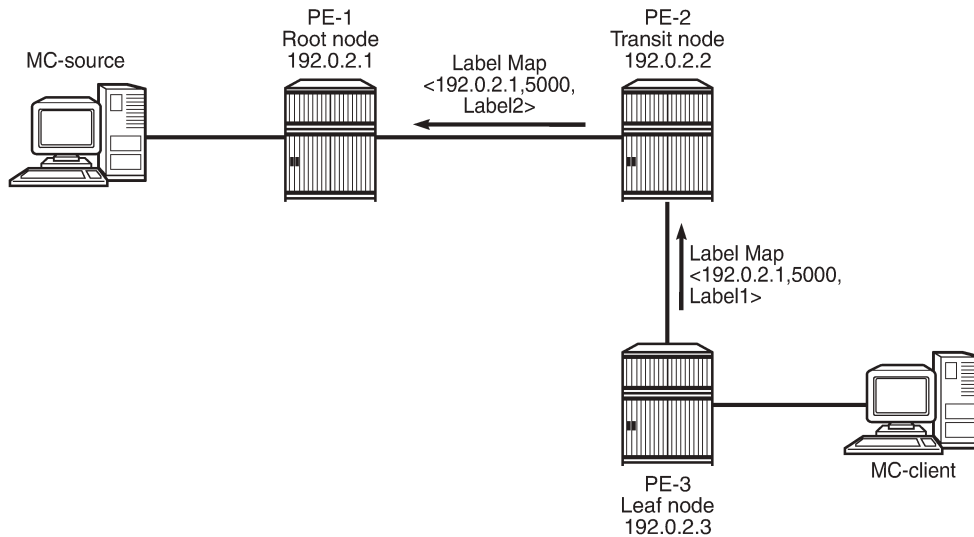
The setup of the P2MP LSP in the control plane is as follows.

1. The leaf node initiates a tree setup according to what is configured. Mandatory parameters are the IP address of the root and an opaque value. The leaf node sends an LDP label map message to its upstream hop toward the root node of the tree.
2. Each transit node receives the LDP label map message and sends another LDP label map message to its upstream hop toward the root node of the tree. Each label can be different.
3. The root node receives the LDP label map message.

The label map message contains the root node address, an opaque value, and a label. In the example in Figure 1, the root node address is 192.0.2.1 and the opaque value is 5000.



Figure 68: Setup of mLDP P2MP LSP



25513

After the LDP label map messages are sent in the control plane, the nodes program pop, swap, or push entries for the corresponding labels in the data plane.

1. The leaf node programs a pop entry for the label it sent upstream.
2. The transit node programs a swap entry for the label it sent upstream with the next-hop address and the label it received from the downstream node.
3. The root node programs a push entry and a next-hop address for the label it received from the downstream node.

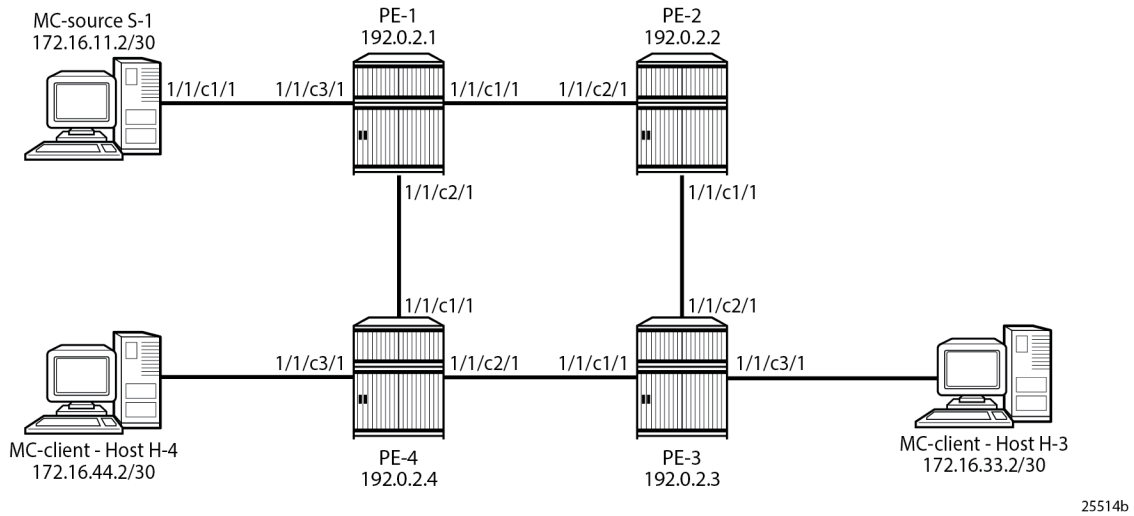
If multiple Equal Cost Multi-Path (ECMP) paths exist between two adjacent nodes, the upstream node of the multicast receiver must program all the entries in the forwarding plane. Only one entry must be active based on the ECMP hashing algorithm.

## Configuration

The example topology shown in [Figure 69: Example topology](#) is used. The multicast source S-1 is connected to root node PE-1. PE-2 or PE-4 is the transit node for traffic destined for PE-3. There are two leaf nodes: PE-3 and PE-4. Multicast client H-3 is connected to PE-3, whereas multicast client H-4 is connected to PE-4.

Under normal circumstances, PE-2 is the transit node for traffic toward PE-3 and PE-4 is an egress node. If PE-4 is the transit node for traffic toward PE-3, while it also has a directly connected receiver, PE-4 is a bud node.

Figure 69: Example topology



25514b

## Configure LDP P2MP LSP

### Initial configuration

The PEs should have the following initial configuration:

- Cards, MDAs, ports
- Router interfaces
- IGP (OSPF or IS-IS)

As an example, the router interfaces and OSPF configuration on PE-1 are as follows:

```
# on PE-1:
configure {
  router "Base" {
    interface "int-PE-1-PE-2" {
      port 1/1/c1/1
      ipv4 {
        primary {
          address 192.168.12.1
          prefix-length 30
        }
      }
    }
    interface "int-PE-1-PE-4" {
      port 1/1/c2/1
      ipv4 {
        primary {
          address 192.168.14.1
          prefix-length 30
        }
      }
    }
  }
  interface "int-PE-1-S-1" {
```

```

port 1/1/c3/1
  ipv4 {
    primary {
      address 172.16.11.1
      prefix-length 30
    }
  }
}
interface "system" {
  ipv4 {
    primary {
      address 192.0.2.1
      prefix-length 32
    }
  }
}
}
ospf 0 {
  admin-state enable
  area 0.0.0.0 {
    interface "int-PE-1-PE-2" {
      interface-type point-to-point
    }
    interface "int-PE-1-PE-4" {
      interface-type point-to-point
    }
    interface "int-PE-1-S-1" {
      interface-type point-to-point
    }
    interface "system" { }
  }
}
}
}

```

## Enabling mLDP

When LDP is enabled, mLDP is enabled by default

The following command enables mLDP on a specific interface:

```

configure {
  router "Base" {
    ldp {
      interface-parameters {
        interface "<ip-int-name>" {
          ipv4 {
            fec-type-capability {
              p2mp-ipv4 true
            }
          }
        }
      }
    }
  }
}

```

Enable LDP (including mLDP) on the router interfaces of PE-1, as follows:

```

#on PE-1:
configure {
  router "Base" {
    ldp {
      interface-parameters {
        interface "int-PE-1-PE-2" {
          ipv4 { }
        }
      }
      interface "int-PE-1-PE-4" {
        ipv4 { }
      }
    }
  }
}

```

```

    }
  }
}

```

Verify that mLDP is enabled (P2MP: Enabled), as follows:

```

[/]
A:admin@PE-1# show router ldp status

=====
LDP Status for IPv4 LSR ID 192.0.2.1
          IPv6 LSR ID ::
=====
---snip---
Admin State      : Up
IPv4 Oper State  : Up
IPv4 Up Time     : 0d 00:04:29
IPv4 Oper Down Rea*: n/a
IPv4 Oper Down Eve*: 0
IPv6 Oper State  : Down
IPv6 Down Time   : 0d 00:04:29
IPv6 Oper Down Reason: systemIpDown
IPv6 Oper Down Events: 0
---snip---

Capabilities
-----
Dynamic          : Enabled
IPv4 Prefix Fec  : Enabled
Service Fec128   : Enabled
MP MBB           : Enabled
Unrecognized Notif*: Enabled
P2MP             : Enabled
IPv6 Prefix Fec  : Enabled
Service Fec129   : Enabled
Overload         : Enabled
=====
* indicates that the corresponding row element may have been truncated.

```

Verify that mLDP is enabled on the interface "int-PE-1-PE-2" (IPv4 P2MP Fec Cap), as follows:

```

[/]
A:admin@PE-1# show router ldp interface "int-PE-1-PE-2" detail

=====
LDP Interfaces
=====

Interface "int-PE-1-PE-2"
=====

BASE
-----
Admin State      : Up
BFDF Status      : Disabled
Oper State       : Up
Load Bal Wt      : None

IPv4
-----
IPv4 Admin State : Up
Last Oper Chg    : 0d 00:01:39
Hold Time        : 15
Oper Hold Time   : 15
Keepalive Timeout : 30
Transport Addr   : System
Active Adjacencies : 1
Local LSR Type   : System
Local LSR        : None
IPv4 Pfx Fec Cap : Enabled
IPv4 P2MP Fec Cap : Enabled
32-BitLocalLsr   : Disabled
IPv6 Pfx Fec Cap : Enabled
IPv6 P2MP Fec Cap: Enabled
=====

```

```
No. of Interfaces: 1
=====
```

Disable mLDP on interface "int-PE-1-PE-2" and verify the LDP status again, as follows:

```
# on PE-1:
configure {
  router "Base" {
    ldp {
      interface-parameters {
        interface "int-PE-1-PE-2" {
          ipv4 {
            fec-type-capability {
              p2mp-ipv4 false
            }
          }
        }
      }
    }
  }
}
```

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

=====
LDP Status for IPv4 LSR ID 192.0.2.1
          IPv6 LSR ID ::
=====
---snip---
Admin State      : Up
IPv4 Oper State  : Up
IPv6 Oper State  : Down
---snip---
-----
Capabilities
-----
Dynamic          : Enabled
IPv4 Prefix Fec  : Enabled
Service Fec128   : Enabled
MP MBB           : Enabled
Unrecognized Notif*: Enabled
P2MP             : Enabled
IPv6 Prefix Fec  : Enabled
Service Fec129   : Enabled
Overload         : Enabled
-----
* indicates that the corresponding row element may have been truncated.
```

P2MP LDP is still enabled on the router, but it is disabled on interface "int-PE-1-PE-2", which can be verified as follows:

```
[/]
A:admin@PE-1# show router ldp interface "int-PE-1-PE-2" detail

=====
LDP Interfaces
=====
Interface "int-PE-1-PE-2"
=====
-----
BASE
-----
Admin State      : Up
BFDD Status      : Disabled
Oper State       : Up
Load Bal Wt      : None
-----
IPv4
-----
IPv4 Admin State : Up
IPv4 Oper State  : Up
---snip---
```

```
IPv4 Pfx Fec Cap : Enabled          IPv6 Pfx Fec Cap : Enabled
IPv4 P2MP Fec Cap : Disabled       IPv6 P2MP Fec Cap: Enabled
=====
No. of Interfaces: 1
=====
```

P2MP multicast forwarding can be disabled per LDP interface. Disabling P2MP multicast forwarding does not prevent LDP from exchanging P2MP FEC elements on that interface in the control plane. In the data plane, the forwarding plane is not programmed with the next hop on the outgoing interface that is P2MP disabled.

## Configure tunnel interface on the root and leaf nodes

Multicast LDP can be applied in different scenarios. In the following example, a tunnel interface is created on the root and leaf nodes. Other examples are Multicast Virtual Private Network (MVPN) with mLDP and dynamic PIM-mLDP mapping. In several ACG chapters on MVPN, mLDP is chosen; for example, in *Multicast VPN: Use of Wildcard Selective PMSI*.

A tunnel interface needs to be created on the root node, as follows:

```
# on PE-1:
configure {
  router "Base" {
    tunnel-interface {
      ldp-p2mp-root 5000 sender-address 192.0.2.1 { }
    }
  }
}
```

In this example, the tunnel interface gets interface index 73728, as follows:

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

=====
P2MP-RSVP P2MP-LDP Tunnel-Interfaces
=====
LSP/LDP      Type      SenderAddr      IfIndex      RootNode
-----
5000         ldp       192.0.2.1      73728       Yes
-----
Interfaces : 1
=====
```

A similar command is launched on the leaf nodes, but with the keyword **ldp-p2mp-leaf**, as follows:

```
# on PE-3 and PE-4:
configure {
  router "Base" {
    tunnel-interface {
      ldp-p2mp-leaf 5000 sender-address 192.0.2.1 { }
    }
  }
}
```

```
[/]
A:admin@PE-3# show router tunnel-interface

=====
P2MP-RSVP P2MP-LDP Tunnel-Interfaces
=====
```

LSP/LDP	Type	SenderAddr	IfIndex	RootNode
5000	ldp	192.0.2.1	73728	No
Interfaces : 1				

A P2MP LSP ping can be sent to verify the P2MP LSP. The options for P2MP LSP ping are as follows:

```
(in classic CLI context)
*A:PE-1# oam p2mp-lsp-ping ?
- p2mp-lsp-ping <lsp-name> [p2mp-instance <instance-name>
    [s2l-dest-address <ipv4-address> [... up to 5]]
    [ttl <label-ttl>]
- p2mp-lsp-ping ldp <p2mp-identifier> [vpn-recursive-fec]
    [sender-addr <ipv4-address>]
    [leaf-addr <ipv4-address> [... up to 5]]
- p2mp-lsp-ping ldp-ssm source <ip-address> group <ip-address>
    [router <router-instance>|service-name <service-name>]
    [sender-addr <ipv4-address>] [leaf-addr <ipv4-address>
    [... up to 5]]
- options common to all p2mp-lsp-ping cases: [fc <fc-name> [profile {in|out}]]
    [size <octets>] [timeout <timeout>] [detail]

<lsp-name>          : [64 chars max]
<instance-name>    : [32 chars max]
<ipv4-address>     : a.b.c.d
<in|out>           : in|out - Default: out
<fc-name>          : be|l2|af|l1|h2|ef|h1|nc - Default: be
<octets>           : [1..9786] - Default: 1
<label-ttl>        : [1..255] - Default: 255
<timeout>          : [1..120] seconds - Default: 10
<detail>           : keyword - displays detailed information
<p2mp-identifier>  : [1..4294967295]
<ldp-ssm>          : keyword - Label Distribution Protocol, Source-Specific
                    Multicast
<ip-address>       : ipv4-address - a.b.c.d
                    ipv6-address - x:x:x:x:x:x:x:x (eight 16-bit pieces)
                    x:x:x:x:x:d.d.d.d
                    x - [0..FFFF]H
                    d - [0..255]D
<router-instance> : <router-name>|<vprn-svc-id>
                    router-name - "Base" Default - Base
                    vprn-svc-id - [1..2147483647]
<service-name>    : [64 chars max]
<vpn-recursive-fec> : keyword - add a VPN Recursive FEC element to the launched packet
                    (useful for pingng a VPN BGP inter-AS Option B leaf)
```

Verify the P2MP LSP with the following ping command:

```
A:PE-1# oam p2mp-lsp-ping ldp 5000
P2MP identifier 5000: | 88 bytes MPLS payload

Total Leafs responded = 2
    round-trip min/avg/max = 1.41 / 1.75 / 2.10 ms

Responses based on return code:
EgressRtr(3)=2
```

Both leaf nodes have sent a reply. The return code 3 indicates that the replying router is an egress for the Forwarding Equivalence Class (FEC).

For a detailed output per leaf, use the following command:

```
A:PE-1# oam p2mp-lsp-ping ldp 5000 detail
P2MP identifier 5000: | 88 bytes MPLS payload

=====
Leaf Information
=====
From          RTT          Return Code
-----
192.0.2.4     =1.56ms     EgressRtr(3)
192.0.2.3     =2.47ms     EgressRtr(3)
=====

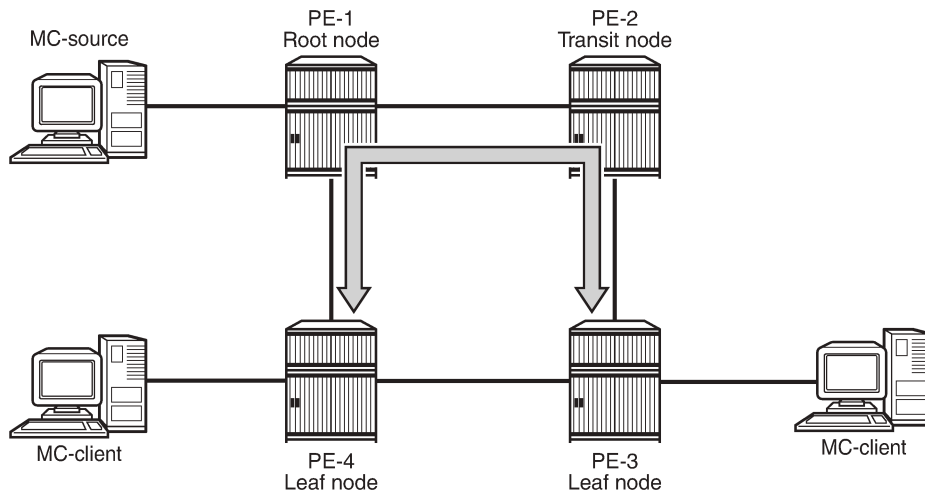
Total Leafs responded = 2
      round-trip min/avg/max = 1.56 / 2.01 / 2.47 ms

Responses based on return code:
EgressRtr(3)=2
```

### Verify LDP P2MP bindings

The example LDP P2MP LSP is shown in [Figure 70: LDP P2MP LSP](#). In this case, PE-4 is only an egress node and not a bud node.

Figure 70: LDP P2MP LSP



25515

Verify the LDP P2MP bindings on the leaf node PE-4, as follows.

The leaf node programs a pop entry for the label sent upstream.

```
[/]
A:admin@PE-4# show router ldp bindings active p2mp opaque-type generic ipv4

=====
LDP Bindings (IPv4 LSR ID 192.0.2.4)
              (IPv6 LSR ID ::)
=====
```



```

Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
=====
LDP Generic IPv4 P2MP Bindings (Active)
=====
P2MP-Id          Interface
RootAddr         Op
IngLbl           EgrLbl
EgrNH            EgrIf/LspId
-----
5000             73728
192.0.2.1       Pop
524283          --
--              --
-----
No. of Generic IPv4 P2MP Active Bindings: 1
=====

```

```

[/]
A:admin@PE-4# show router ldp bindings p2mp opaque-type generic ipv4 detail
=====
LDP Bindings (IPv4 LSR ID 192.0.2.4)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
=====
LDP Generic IPv4 P2MP Bindings
=====
P2MP Type       : 1                P2MP-Id       : 5000
Root-Addr       : 192.0.2.1
-----
Peer           : 192.0.2.1:0
Ing Lbl        : 524283U
Egr Lbl        : --
Egr Int/LspId  : --
EgrNextHop     : --
Egr. Flags     : None              Ing. Flags    : None
=====
No. of Generic IPv4 P2MP Bindings: 1
=====

```

PE-4 is only an egress node and not a transit node. There is no next hop.

Verify the LDP P2MP bindings on the leaf node PE-3, as follows:

```

[/]
A:admin@PE-3# show router ldp bindings active p2mp opaque-type generic ipv4
=====

```

```

LDP Bindings (IPv4 LSR ID 192.0.2.3)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
=====
LDP Generic IPv4 P2MP Bindings (Active)
=====
P2MP-Id          Interface
RootAddr         0p
IngLbl           EgrLbl
EgrNH            EgrIf/LspId
-----
5000              73728
192.0.2.1        Pop
524283           --
--              --
-----
No. of Generic IPv4 P2MP Active Bindings: 1
=====
  
```

Because PE-3 is an egress node, there is no next hop. The traffic toward PE-3 is sent via transit PE-2 and not via PE-4, as can be verified as follows:

```

[/]
A:admin@PE-3# show router ldp bindings p2mp opaque-type generic ipv4 detail
=====
LDP Bindings (IPv4 LSR ID 192.0.2.3)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
=====
LDP Generic IPv4 P2MP Bindings
=====
P2MP Type       : 1                P2MP-Id       : 5000
Root-Addr       : 192.0.2.1
-----
Peer            : 192.0.2.2:0
Ing Lbl         : 524283U
Egr Lbl         : --
Egr Int/LspId   : --
EgrNextHop      : --
Egr. Flags     : None              Ing. Flags    : None
=====
No. of Generic IPv4 P2MP Bindings: 1
=====
  
```

PE-2 has programmed a swap entry for the label it sent to its upstream node PE-1 with the next-hop address and the label it received from the downstream node, as follows:

```
[/]
A:admin@PE-2# show router ldp bindings active p2mp opaque-type generic ipv4

=====
LDP Bindings (IPv4 LSR ID 192.0.2.2)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
=====
LDP Generic IPv4 P2MP Bindings (Active)
=====
P2MP-Id      Interface
RootAddr     Op
IngLbl       EgrLbl
EgrNH        EgrIf/LspId
-----
5000         Unknw
192.0.2.1   Swap
524283      524283
192.168.23.2 1/1/c1/1
-----
No. of Generic IPv4 P2MP Active Bindings: 1
=====
```

```
[/]
A:admin@PE-2# show router ldp bindings p2mp opaque-type generic ipv4 detail

=====
LDP Bindings (IPv4 LSR ID 192.0.2.2)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
=====
LDP Generic IPv4 P2MP Bindings
=====
P2MP Type    : 1          P2MP-Id    : 5000
Root-Addr    : 192.0.2.1
-----
Peer        : 192.0.2.1:0
Ing Lbl     : 524283U
Egr Lbl     : --
Egr Int/LspId : --
EgrNextHop  : --
Egr. Flags  : None      Ing. Flags : None
-----
P2MP Type    : 1          P2MP-Id    : 5000
```

```

Root-Addr      : 192.0.2.1
-----
Peer           : 192.0.2.3:0
Ing Lbl        : --
Egr Lbl        : 524283
Egr Int/LspId  : 1/1/c1/1
EgrNextHop     : 192.168.23.2
Egr. Flags     : None           Ing. Flags : None
Egr If Name    : int-PE-2-PE-3
Metric         : 1             Mtu         : 8922
=====
No. of Generic IPv4 P2MP Bindings: 2
=====

```

The egress next hop is PE-3.

On the root node PE-1, there is MPLS multicast replication. One traffic stream goes via transit node PE-2 toward leaf node PE-3 and the other traffic stream goes directly toward leaf node PE-4. There are two push entries with the corresponding next-hop address, as follows:

```

[/]
A:admin@PE-1# show router ldp bindings active p2mp opaque-type generic ipv4
=====
LDP Bindings (IPv4 LSR ID 192.0.2.1)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
=====
LDP Generic IPv4 P2MP Bindings (Active)
=====
P2MP-Id      Interface
RootAddr     Op
IngLbl       EgrLbl
EgrNH        EgrIf/LspId
-----
5000         73728
192.0.2.1    Push
--          524283
192.168.12.2 1/1/c1/1

5000         73728
192.0.2.1    Push
--          524283
192.168.14.2 1/1/c2/1

-----
No. of Generic IPv4 P2MP Active Bindings: 2
=====

```

```

[/]
A:admin@PE-1# show router ldp bindings p2mp opaque-type generic ipv4 detail
=====
LDP Bindings (IPv4 LSR ID 192.0.2.1)
(IPv6 LSR ID ::)

```

```

=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
=====
LDP Generic IPv4 P2MP Bindings
=====
-----
P2MP Type      : 1                P2MP-Id       : 5000
Root-Addr      : 192.0.2.1
-----
Peer           : 192.0.2.2:0
Ing Lbl        : --
Egr Lbl        : 524283
Egr Int/LspId  : 1/1/c1/1
EgrNextHop     : 192.168.12.2
Egr. Flags     : None              Ing. Flags    : None
Egr If Name    : int-PE-1-PE-2
Metric         : 1                  Mtu           : 8922
-----
P2MP Type      : 1                P2MP-Id       : 5000
Root-Addr      : 192.0.2.1
-----
Peer           : 192.0.2.4:0
Ing Lbl        : --
Egr Lbl        : 524283
Egr Int/LspId  : 1/1/c2/1
EgrNextHop     : 192.168.14.2
Egr. Flags     : None              Ing. Flags    : None
Egr If Name    : int-PE-1-PE-4
Metric         : 1                  Mtu           : 8922
-----
No. of Generic IPv4 P2MP Bindings: 2
=====

```

## Tools command

The following tools command can be launched on any of the nodes in the P2MP LSP.

For the ingress node PE-1, where one branch goes to transit node PE-2 (192.0.2.2) and another branch to leaf node PE-4 (192.0.2.4), the output is as follows:

```

[/]
A:admin@PE-1# tools dump router ldp fec p2mp-id 5000 root 192.0.2.1
P2MP: root: 192.0.2.1, T: 1, L: 4, TunnelId: 5000
  Create Time      : 09/06/23 21:43:41.278 (elapsed: 0d 00:11:16)
  Last Mod. Time   : 09/06/23 21:43:51.708 (elapsed: 0d 00:11:06)
  FEC Flags        : Push Mttm
  FEC typedFlags   : none
  TunlIfId        : 73728      (OperState : up)
  mttmTunnAttr     : None
  LSP ID           : 65540      LSP ID Acct.  : 4
  useMcastRteTbl  : No         mttmRd        : 0x0
  isIngressMttm   : Yes       HasLeaf      : No
  isIngrItermdte  : No         CanProgIngress: Yes
  InPhopFrr       : No         InAsbrFrr     : No
  isStitchedUprr  : No         isLwrMbbCplte: No

```

```

stitchRteType : Rtm          detctBkpAsbrLp: No
bkpLoop       : No          bkpAsbrLoop  : No
sysIpAddr     : 0.0.0.0
RslvdPhop(p) : 0.0.0.0:0 (seqNum 0 isBgp No isTtm No)
RslvdPhop(b) : 0.0.0.0:0 (seqNum 0 isBgp No isTtm No)
pri Upstream  : None
mbb Upstream  : None
bkp Upstream  : None
AdvInLabel(p) : 0
AdvInLabel(b) : 0
RslvdAsbr(p) : 0.0.0.0 (seqNum 0 recursive: No vpn: No)
RslvdAsbr(b) : 0.0.0.0 (seqNum 0 recursive: No vpn: No)
Num Resolved  Nhops : 2
Num MBB Req.  Nhops : 0
Num Programmed Nhops : 2
Num MBB Req. Lwr Fecs : 0
Programmed Nhop[01] : 192.0.2.2:0, OutLabel 524283
Programmed Nhop[02] : 192.0.2.4:0, OutLabel 524283
Metric        : 1          Mtu          : 8922

Num of Peers : 2

FEC Peer: 192.0.2.2:0
Peer Flags: MPush (0x800)
ModTime   : 09/06/23 21:43:50.310 (elapsed.: 0d 00:11:07)

->Num Egress Labels:
-> (Label: 524283   Status: UsePush)
Flow Label Tx: no, Rx: no

<-Num Ingress Labels:
None

<Resolved as Next Hop>
Next Hop Info :
metric: 1 mtu: 8922
[01]: Next Hop: 192.168.12.2   Interface: 2 Inner Label: 0

FEC Peer: 192.0.2.4:0
Peer Flags: MPush (0x800)
ModTime   : 09/06/23 21:43:51.702 (elapsed.: 0d 00:11:06)

->Num Egress Labels:
-> (Label: 524283   Status: UsePush)
Flow Label Tx: no, Rx: no

<-Num Ingress Labels:
None

<Resolved as Next Hop>
Next Hop Info :
metric: 1 mtu: 8922
[01]: Next Hop: 192.168.14.2   Interface: 3 Inner Label: 0

```

The labels that are pushed at PE-1 are 524283 for traffic to PE-2 and 524283 for traffic to PE-4.

On transit node PE-2, the incoming label 524283 is swapped to outgoing label 524283 toward PE-3, as follows:

```

[/]
A:admin@PE-2# tools dump router ldp fec p2mp-id 5000 root 192.0.2.1
P2MP: root: 192.0.2.1, T: 1, L: 4, TunnelId: 5000

```

```
Create Time      : 09/06/23 21:43:50.304 (elapsed: 0d 00:11:09)
Last Mod. Time  : 09/06/23 21:43:50.304 (elapsed: 0d 00:11:09)
FEC Flags       : Swap
FEC typedFlags  : none
TunIfId        : 0          (OperState : dn)
mtmTunnAttr     : None
LSP ID         : 0          LSP ID Acct. : 0
useMcastRteTbl: No         mtmRd       : 0x0
isIngressMtm   : No         HasLeaf     : No
isIngrItermdte: No         CanProgIngr: No
InPhopFrr      : No         InAsbrFrr   : No
isStitchedUp   : No         isLwrMbbCplte : No
stitchRteType  : None       detctBkpAsbrLp: No
bkpLoop        : No         bkpAsbrLoop  : No
sysIpAddr      : 0.0.0.0
RslvdPhop(p)   : 192.0.2.1:0 (seqNum 2 isBgp No isTtm No)
RslvdPhop(b)   : 0.0.0.0:0 (seqNum 0 isBgp No isTtm No)
pri Upstream   : 192.0.2.1:0, AdvLabel 524283
mbb Upstream   : None
bkp Upstream   : None
AdvInLabel(p) : 524283
AdvInLabel(b)  : 0
RslvdAsbr(p)   : 0.0.0.0 (seqNum 0 recursive: No vpn: No)
RslvdAsbr(b)   : 0.0.0.0 (seqNum 0 recursive: No vpn: No)
PrgInLabel(b)  : 1
Num Resolved   Nhops : 1
Num MBB Req.   Nhops : 0
Num Programmed Nhops : 1
Num MBB Req. Lwr Fecs : 0
Programmed Nhop[01] : 192.0.2.3:0, OutLabel 524283
Metric         : 1          Mtu          : 8922

Num of Peers   : 2

FEC Peer: 192.0.2.1:0
Peer Flags: none (0x0)
ModTime      : 09/06/23 21:43:50.309 (elapsed.: 0d 00:11:09)

->Num Egress Labels:
None

<-Num Ingress Labels:
<- (Label: 524283 Status: UseSwap)
Rej Status: OK
Flow Label Tx: no, Rx: no
Flow Label Tx Sent: no, Rx Sent: no

<Resolved as CUR Upstream>

FEC Peer: 192.0.2.3:0
Peer Flags: MSwap (0x1000)
ModTime      : 09/06/23 21:43:50.301 (elapsed.: 0d 00:11:09)

->Num Egress Labels:
-> (Label: 524283 Status: UseSwap)
Flow Label Tx: no, Rx: no

<-Num Ingress Labels:
None

<Resolved as Next Hop>
Next Hop Info :
metric: 1 mtu: 8922
[01]: Next Hop: 192.168.23.2 Interface: 3 Inner Label: 0
```

On leaf node PE-3, the incoming label from PE-2 (524283) is popped. There is no next hop.

```
[/]
A:admin@PE-3# tools dump router ldp fec p2mp-id 5000 root 192.0.2.1
P2MP: root: 192.0.2.1, T: 1, L: 4, TunnelId: 5000
  Create Time   : 09/06/23 21:43:50.305 (elapsed: 0d 00:11:11)
  Last Mod. Time: 09/06/23 21:43:50.305 (elapsed: 0d 00:11:11)
  FEC Flags     : Pop Mtm
  FEC typedFlags: none
  TunlIfId     : 73728      (OperState : up)
  mttmTunnAttr : None
  LSP ID       : 0          LSP ID Acct. : 0
  useMcastRteTbl: No       mttmRd      : 0x0
  isIngressMtm : No       HasLeaf    : Yes
  isIngrItermdte: No     CanProgIngress: No
  InPhopFrr    : No       InAsbrFrr   : No
  isStitchedUp : No       isLwrMbbCplte: No
  stitchRteType: None     detctBkpAsbrLp: No
  bkpLoop      : No       bkpAsbrLoop : No
  sysIpAddr    : 0.0.0.0
  RslvdPhop(p) : 192.0.2.2:0 (seqNum 2 isBgp No isTtm No)
  RslvdPhop(b) : 0.0.0.0:0 (seqNum 0 isBgp No isTtm No)
  pri Upstream : 192.0.2.2:0, AdvLabel 524283
  mbb Upstream : None
  bkp Upstream : None
  AdvInLabel(p) : 524283
  AdvInLabel(b) : 0
  RslvdAsbr(p) : 0.0.0.0 (seqNum 0 recursive: No vpn: No)
  RslvdAsbr(b) : 0.0.0.0 (seqNum 0 recursive: No vpn: No)
  PrgInLabel(b) : 1
  Num Resolved Nhops : 1
  Num MBB Req. Nhops : 0
  Num Programmed Nhops : 1
  Num MBB Req. Lwr Fecs : 0
  Programmed Nhop[01] : 0.0.0.0:0, OutLabel 0 (Leaf)
  Metric      : 0          Mtu      : 0

  Num of Peers : 1

  FEC Peer: 192.0.2.2:0
  Peer Flags: none (0x0)
  ModTime   : 09/06/23 21:43:50.311 (elapsed.: 0d 00:11:11)

  ->Num Egress Labels:
  None

  <-Num Ingress Labels:
  <- (Label: 524283 Status: UsePop)
  Rej Status: OK
  Flow Label Tx: no, Rx: no
  Flow Label Tx Sent: no, Rx Sent: no

  <Resolved as CUR Upstream>
```

The output for leaf node PE-4 is similar.



## Debug commands

Debugging was enabled on the nodes when LDP was configured. To distinguish which messages are being logged for a debug command, the debug configuration is different for the nodes, as follows:

```
# on PE-2:
debug {
  router "Base" {
    ldp {
      peer 192.0.2.1 {
        event {
          bindings { }
        }
      }
    }
  }
}

# on PE-3::
debug {
  router "Base" {
    ldp {
      peer 192.0.2.2 {
        packet {
          label {
            detail true
          }
        }
      }
    }
  }
}

# on PE-4:
debug {
  router "Base" {
    ldp {
      peer 192.0.2.1 {
        packet {
          init {
            detail true
          }
        }
      }
    }
  }
}
```

The following LDP messages are logged. The first two messages correspond to the label mapping messages to establish LDP bindings. The following message is sent from transit node PE-2 to root node PE-1.

```
# on PE-2:
debug {
  router "Base" {
    ldp {
      peer 192.0.2.1 {
        event {
          bindings { }
        }
      }
    }
  }
}
```

```
# on PE-2:
7 2023/09/06 23:43:50.308 CEST MINOR: DEBUG #2001 Base LDP
"LDP: Binding
Sending Label mapping label 524283 for P2MP: root = 192.0.2.1, T: 1, L: 4, TunnelId: 5000
to peer 192.0.2.1:0."
```

The following LDP message is sent by the leaf node PE-3 to the transit node PE-2.

```
# on PE-3::
debug {
  router "Base" {
    ldp {
```

```
peer 192.0.2.2 {
  packet {
    label {
      detail true
    }
  }
}
```

```
# on PE-3:
7 2023/09/06 23:43:50.307 CEST MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Label Mapping packet (msgId 30) to 192.0.2.2:0
Protocol version = 1
Label 524283 advertised for the following FECs
P2MP: root = 192.0.2.1, T: 1, L: 4, TunnelId: 5000
"
```

The following message shows the negotiation of capabilities when LDP bindings are initialized.

```
# on PE-4:
debug {
  router "Base" {
    ldp {
      peer 192.0.2.1 {
        packet {
          init {
            detail true
          }
        }
      }
    }
  }
}
```

```
# on PE-4:
1 2023/09/06 23:41:22.613 CEST MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Initialization packet (msgId 2) to 192.0.2.1:0
Protocol version = 1
Keepalive Timeout = 30 Label Advertisement = downstreamUnsolicited
Loop Detection = Off PathVector Limit = 0 Max Pdu = 4096
P2MP Capability = yes
MP MBB Capability = yes
Overload Capability = yes
Dynamic Capability = yes
Unrecognized Notification Capability = yes
"
```

## Configure multicast LDP and verify traffic

### Configure PIM and IGMP on the root and leaf nodes

PIM needs to be enabled on the root node on the interface toward the multicast source S-1, as follows:

```
# on PE-1:
configure {
  router "Base" {
    pim {
      interface "int-PE-1-S-1" { }
    }
  }
}
```

On the leaf nodes, PIM needs to be enabled (no shutdown), but no interfaces need to be assigned.

The IGMP configuration for root node PE-1 is needed to forward the incoming traffic for multicast group 232.1.1.1 from source 172.16.11.2 to the tunnel-interface. If IGMP is not configured, the incoming traffic on the interface toward the multicast source is dropped, because no outgoing interface is defined.

```
# on PE-1:
configure {
  router "Base" {
    igmp {
      tunnel-interface {
        ldp-p2mp-root 5000 sender-address 192.0.2.1 {
          static {
            group 232.1.1.1 {
              source 172.16.11.2 { }
            }
          }
        }
      }
    }
  }
}
```

The IGMP configuration for leaf node PE-3 is as follows:

```
# on PE-3:
configure {
  router "Base" {
    igmp {
      interface "int-PE-3-H-3" {
        static {
          group 232.1.1.1 {
            source 172.16.11.2 { }
          }
        }
      }
    }
  }
}
```

The incoming traffic from the tunnel interface is forwarded to the outgoing interface toward the receiving multicast host H-3.

The IGMP configuration for leaf node PE-4 is similar.

At this point, the IGMP/PIM configuration on the root node is complete. This can be verified, as follows:

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

=====
Legend:  A = Active   S = Standby
=====
PIM Groups ipv4
=====
Group Address          Type          Spt Bit  Inc Intf      No.Oifs
  Source Address      RP
-----
232.1.1.1              (S,G)                int-PE-1-S-1   1
  172.16.11.2
-----
Groups : 1
=====
```

```
[/]
A:admin@PE-1# show router pim group detail

=====
PIM Source Group ipv4
=====
Group Address      : 232.1.1.1
Source Address     : 172.16.11.2
RP Address         : 0
Advrt Router       : 192.0.2.1
```

```

Flags          :                               Type          : (S,G)
Mode           : sparse
MRIB Next Hop  : 172.16.11.2
MRIB Src Flags : direct
Keepalive Timer : Not Running
Up Time        : 0d 00:03:32      Resolved By    : rtable-u

Up JP State    : Joined           Up JP Expiry   : 0d 00:00:00
Up JP Rpt      : Not Joined StarG Up JP Rpt Override : 0d 00:00:00

Register State : No Info
Reg From Anycast RP: No

Rpf Neighbor   : 172.16.11.2
Incoming Intf : int-PE-1-S-1
Outgoing Intf List : mpls-if-73728

Curr Fwding Rate : 9751.560 kbps
Forwarded Packets : 17930           Discarded Packets : 0
Forwarded Octets  : 26572260      RPF Mismatches    : 0
Spt threshold     : 0 kbps         ECMP opt threshold : 7
Admin bandwidth   : 1 kbps

-----
Groups : 1
=====

```

The incoming interface is the interface facing the multicast source S-1. The outgoing interface is a reference to the tunnel interface. The name for the outgoing interface (mpls-if-73728) contains the tunnel interface index 73728 as in previous CLI output. The multicast source S-1 is already sending traffic, but the receivers cannot receive it yet.

The configuration on the leaf nodes is still incomplete. A multicast policy needs to be configured and applied first.

## Configure and apply multicast policy on leaf nodes

The leaf nodes need to get multicast traffic off the LDP P2MP LSP. Therefore, a multicast policy needs to be created and applied, as follows:

```

# on PE-3 and PE-4:
configure {
    multicast-management {
        multicast-info-policy "p2mp-pol" {
            bundle "bundle1" {
                primary-tunnel-interface {
                    ldp-p2mp 5000
                    sender 192.0.2.1
                }
                channel start 232.1.1.1 end 232.1.1.1 { }
            }
        }
    }
}

configure {
    router "Base" {
        multicast-info-policy "p2mp-pol"
    }
}

```

## Verify multicast traffic on leaf nodes

Verify the multicast traffic, as follows:

```
[/]
A:admin@PE-3# show router pim group

=====
Legend:  A = Active  S = Standby
=====
PIM Groups ipv4
=====
Group Address          Type          Spt Bit  Inc Intf  No.0ifs
  Source Address      RP           State    Inc Intf(S)
-----
232.1.1.1              (S,G)                mpls-if-73728  1
  172.16.11.2
-----
Groups : 1
=====
```

The multicast source S-1 sends a multicast stream with group address 232.1.1.1. The multicast traffic is received by the leaf nodes, which can be verified as follows:

```
[/]
A:admin@PE-3# show router pim group detail

=====
PIM Source Group ipv4
=====
Group Address      : 232.1.1.1
Source Address     : 172.16.11.2
RP Address         : 0
Advt Router        :
Flags              :
Mode               : sparse
MRIB Next Hop      :
MRIB Src Flags     : remote
Keepalive Timer    : Not Running
Up Time            : 0d 00:03:26
Resolved By        : unresolved

Up JP State        : Joined
Up JP Rpt          : Not Joined StarG
Up JP Expiry       : 0d 00:00:05
Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       :
Incoming Intf    : mpls-if-73728
Outgoing Intf List : int-PE-3-H-3

Curr Fwding Rate   : 9751.560 kbps
Forwarded Packets  : 19246
Forwarded Octets   : 28522572
Spt threshold      : 0 kbps
Admin bandwidth    : 1 kbps
Discarded Packets  : 0
RPF Mismatches     : 0
ECMP opt threshold : 7
-----
Groups : 1
=====
```

The incoming interface is from the tunnel interface, whereas the outgoing interface is toward the receiving multicast host H-3.

## mLDP fast upstream switchover

mLDP fast upstream switchover allows a downstream node of an mLDP FEC to perform a fast switchover and source the traffic from another upstream node. This switchover is necessary when IGP and LDP are converging after a failure of the upstream LSR, which is the primary next hop of the root LSR for the P2MP FEC. There is traffic duplication toward the node that has the upstream alternate backup (in this case to PE-3), but only one stream is accepted. The multicast stream is sent to the primary next hop as well as to the loopfree alternate backup. As long as there is no failure, the primary next hop accepts the traffic and forwards it. The backup rejects the traffic. When a failure occurs and the primary LDP session goes down, the backup starts accepting packets.

mLDP fast upstream switchover provides an upstream Fast Reroute (FRR) node-protection capability for the mLDP FEC packets. This multicast upstream FRR node protection is at the expense of traffic duplication from two different upstream nodes into the node that performs the fast upstream switchover. This feature is described in *draft-pdutta-mpls-mldp-up-redundancy*.

Multicast upstream FRR can be configured for mLDP, as follows:

```
# on PE-1:
configure {
  router "Base" {
    ospf 0 {
      loopfree-alternate { }
    }
    ldp {
      mcast-upstream-frr true
    }
  }
}
```

This configuration can be repeated on some or all of the nodes. In this example, it is configured on all nodes. FRR for unicast can be configured in combination with this, but that is not required. FRR for unicast can be enabled as follows:

```
# on all nodes:
configure {
  routing-options {
    ip-fast-reroute true
  }
  router "Base" {
    ldp {
      fast-reroute { }
    }
  }
}
```

In this example, it is assumed that unicast IP and unicast LDP prefixes do not need to be protected. Therefore, unicast FRR remains disabled.

FRR can be verified as disabled for unicast (FRR) and enabled for multicast (Mcast Upstream FRR), as follows:

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

```

=====
LDP Status for IPv4 LSR ID 192.0.2.1
                IPv6 LSR ID ::
=====
---snip---
Admin State      : Up
IPv4 Oper State  : Up
IPv6 Oper State  : Down
---snip---
FRR            : Disabled           Mcast Upstream FRR : Enabled
Mcast Upst ASBR FRR: Disabled
MP MBB Time      : 3
---snip---
-----
Capabilities
-----
Dynamic          : Enabled           P2MP              : Enabled
IPv4 Prefix Fec  : Enabled           IPv6 Prefix Fec   : Enabled
Service Fec128   : Enabled           Service Fec129    : Enabled
MP MBB           : Enabled           Overload          : Enabled
Unrecognized Notif*: Enabled
=====
* indicates that the corresponding row element may have been truncated.

```

Of the three nodes in the example topology that have upstream nodes, only PE-3 has an upstream alternate for FRR. PE-4 becomes a transit node for traffic destined for PE-3 (but PE-3 drops it, until the primary LDP session fails). PE-3 sends a label mapping message to PE-4 for label 524282, as in the following trace message.

```

# on PE-3:
8 2023/09/07 00:00:27.538 CEST MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Label Mapping packet (msgId 140) to 192.0.2.4:0
Protocol version = 1
Label 524282 advertised for the following FECs
P2MP: root = 192.0.2.1, T: 1, L: 4, TunnelId: 5000
MP Status MBB = REQ
"

```

PE-4 has an additional LDP P2MP binding where the label is swapped, as follows:

```

[/]
A:admin@PE-4# show router ldp bindings active p2mp opaque-type generic ipv4
=====
LDP Bindings (IPv4 LSR ID 192.0.2.4)
                (IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
=====
LDP Generic IPv4 P2MP Bindings (Active)
=====
P2MP-Id          Interface
RootAddr         Op
IngLbl           EgrLbl
EgrNH            EgrIf/LspId
-----

```

```

5000                               73728
192.0.2.1                           Pop
524283                               --
--                                   --

5000                               73728
192.0.2.1                           Swap
524283                               524282
192.168.34.1                         1/1/c2/1

-----
No. of Generic IPv4 P2MP Active Bindings: 2
=====
    
```

PE-3 has an additional entry for the FRR backup that is available (BU - Alternate for Fast Re-Route). PE-3 gets duplicated traffic, but rejects all traffic from PE-4 and only accept traffic from PE-2 as long as there is no failover.

```

[/]
A:admin@PE-3# show router ldp bindings active p2mp opaque-type generic ipv4

=====
LDP Bindings (IPv4 LSR ID 192.0.2.3)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
=====
LDP Generic IPv4 P2MP Bindings (Active)
=====
P2MP-Id                               Interface
RootAddr                             Op
IngLbl                                EgrLbl
EgrNH                                  EgrIf/LspId
-----
5000                                   73728
192.0.2.1                             Pop
524283                                 --
--                                     --

5000                                   73728
192.0.2.1                             Pop
524282BU                              --
--                                     --

-----
No. of Generic IPv4 P2MP Active Bindings: 2
=====
    
```

Because LoopFree Alternate (LFA) and ECMP are mutually exclusive, LFA is only useful when ECMP is disabled. When both are enabled, ECMP has preference.

mLDP fast upstream switchover relies on the fast detection of loss of the LDP session to the upstream peer to which the primary Ingress Label Map (ILM) label had been advertised. As a result, Nokia recommends to perform the following:



1. Enable Bidirectional Forwarding Detection (BFD) on all LDP interfaces to upstream LSR nodes. When BFD detects the loss of the last adjacency to the upstream LSR, BFD brings down the LDP session immediately. The backup ILM is activated.
2. If there is a concurrent T-LDP adjacency to the same LSR node, enable BFD on the T-LDP peer as well as on the interface.
3. Enable the **ldp-sync-timer** option on all interfaces to the upstream LSR nodes.

If the LDP session for the primary ILM to the upstream LSR goes down for any other reason than a failure of the interface or of the upstream LSR, routing and LDP go out of sync. The backup ILM remains activated until the Interior Gateway Protocol (IGP) seeks the next Shortest Path First (SPF). By enabling the **ldp-sync-timer**, this process is accelerated because the advertised link metric gets the maximum value as soon as the LDP session goes down. This triggers the IGP to calculate an SPF route. See chapter [LDP-IGP Synchronization](#).

The FRR configuration can be removed, as follows:

```
# on all nodes:
configure {
  router "Base" {
    ldp {
      delete mcast-upstream-frr
    }
    ospf 0 {
      delete loopfree-alternate
    }
  }
}
```

## Multipoint make-before-break (MP MBB)

Multipoint MBB is performed when the best path to the root changes, but the existing path can still be used, such as when a link comes up or when the routing metric changes. The goal of MBB is to establish a new P2MP LSP before the old P2MP is removed, so as to avoid traffic loss.

Leaf or transit nodes must allocate a new label and program the ILM with a duplicate set of existing Next-Hop Label Forwarding Entries (NHLFEs) toward the upstream nodes. This may lead to traffic duplication for a short period of time.

Multipoint MBB is enabled by default, as follows:

```
[/]
A:admin@PE-3# show router ldp status

=====
LDP Status for IPv4 LSR ID 192.0.2.3
          IPv6 LSR ID ::
=====
---snip---
Admin State      : Up
IPv4 Oper State  : Up          IPv6 Oper State      : Down
---snip---
MP MBB Time      : 3
---snip---
-----
Capabilities
-----
Dynamic          : Enabled          P2MP                 : Enabled
IPv4 Prefix Fec  : Enabled          IPv6 Prefix Fec      : Enabled
```

```

Service Fec128      : Enabled          Service Fec129      : Enabled
MP MBB             : Enabled        OverLoad            : Enabled
Unrecognized Notif*: Enabled
=====
* indicates that the corresponding row element may have been truncated.
    
```

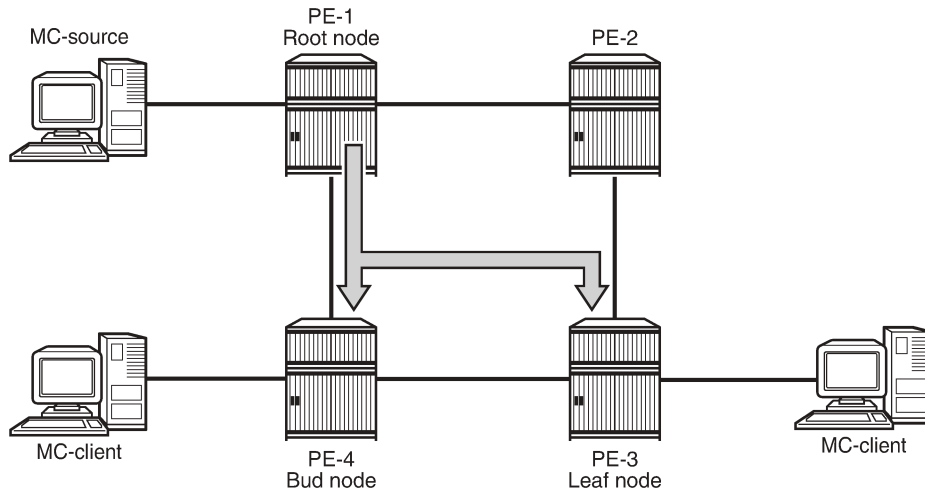
When the metric is increased on the interface (int-PE-3-PE-2) toward the active upstream node, PE-3 sends out an OSPF link status update. Traffic still arrives at PE-3 using the original P2MP LSP. The MBB P2MP LSP is set up.

PE-3 sends a label mapping message toward PE-4, including an MP status TLV carrying an MBB status code indicating that MBB procedures apply to the LSP. PE-4 sends an LDP notification toward PE-3, including an MP status TLV indicating that PE-4 has a state for the existing P2MP LSP.

PE-3 sends an LDP withdrawal message to PE-2. PE-2 replies with an LDP release message.

The multicast traffic arrives at PE-3 using the new LDP P2MP LSP. This way, PE-4 becomes a bud node, and PE-2 is not used for transit anymore; see [Figure 71: New LDP P2MP LSP after metric change](#).

Figure 71: New LDP P2MP LSP after metric change



25516

Originally, leaf node PE-3 preferred the route via PE-2 toward root node PE-1, as follows:

```

[/]
A:admin@PE-3# show router route-table 192.0.2.1
=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]                Type  Proto  Age           Pref
  Next Hop[Interface Name]                Metric
-----
192.0.2.1/32                       Remote OSPF   00h23m14s  10
    192.168.23.1                               2
-----
No. of Routes: 1
---snip---
=====
    
```

Consequently, the label map messages were originally sent to PE-2, not to PE-4. PE-2 is the transit node for traffic destined for PE-3, as follows:

```
[/]
A:admin@PE-2# show router ldp bindings active p2mp opaque-type generic ipv4

=====
LDP Bindings (IPv4 LSR ID 192.0.2.2)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
=====
LDP Generic IPv4 P2MP Bindings (Active)
=====
P2MP-Id          Interface
RootAddr         Op
IngLbl           EgrLbl
EgrNH            EgrIf/LspId
-----
5000              Unknw
192.0.2.1        Swap
524283           524283
192.168.23.2   1/1/c1/1
-----
No. of Generic IPv4 P2MP Active Bindings: 1
=====
```

The metric is changed on the interface between PE-3 and PE-2, as follows:

```
# on PE-3:
configure {
  router "Base" {
    ospf 0 {
      area 0.0.0.0 {
        interface "int-PE-3-PE-2" {
          metric 1000
        }
      }
    }
  }
}
```

The preferred route from leaf node PE-3 to root node PE-1 is now via PE-4, as follows:

```
[/]
A:admin@PE-3# show router route-table 192.0.2.1

=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]          Type   Proto   Age           Pref
Next Hop[Interface Name]   Metric
-----
192.0.2.1/32                Remote OSPF    00h00m23s    10
192.168.34.2                2
-----
No. of Routes: 1
---snip---
```

The leaf node PE-3 prefers to set up a path from PE-4 instead of from PE-2. PE-3 sends label mapping messages to PE-4. The old P2MP LSP is used until the new P2MP LSP is set up. There is no traffic interruption.

PE-3 sends a label withdrawal message to PE-2 and PE-2 is no longer a transit node, as follows:

```
[/]
A:admin@PE-2# show router ldp bindings active p2mp opaque-type generic ipv4

=====
LDP Bindings (IPv4 LSR ID 192.0.2.2)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
=====
LDP Generic IPv4 P2MP Bindings (Active)
=====
P2MP-Id          Interface
RootAddr         Op
IngLbl           EgrLbl
EgrNH            EgrIf/LspId
-----
No Matching Entries Found
=====
```

PE-4 is the transit node for traffic to PE-3, and also has a local multicast client H-4, so it is a bud node, as follows:

```
[/]
A:admin@PE-4# show router ldp bindings active p2mp opaque-type generic ipv4

=====
LDP Bindings (IPv4 LSR ID 192.0.2.4)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
=====
LDP Generic IPv4 P2MP Bindings (Active)
=====
P2MP-Id          Interface
RootAddr         Op
IngLbl           EgrLbl
EgrNH            EgrIf/LspId
-----
5000             73728
192.0.2.1        Pop
524283           --
--              --
```

```

5000                               73728
192.0.2.1                          Swap
524283                             524282
192.168.34.1                       1/1/c2/1

-----
No. of Generic IPv4 P2MP Active Bindings: 2
=====

```

There is no traffic multiplication at the root node PE-1. All traffic goes to PE-4, as follows:

```

[/]
A:admin@PE-1# show router ldp bindings active p2mp opaque-type generic ipv4

=====
LDP Bindings (IPv4 LSR ID 192.0.2.1)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
=====
LDP Generic IPv4 P2MP Bindings (Active)
=====
P2MP-Id      Interface
RootAddr     Op
IngLbl       EgrLbl
EgrNH        EgrIf/LspId
-----
5000         73728
192.0.2.1    Push
--          524283
192.168.14.2 1/1/c2/1

-----
No. of Generic IPv4 P2MP Active Bindings: 1
=====

```

The switchover to this new P2MP LSP occurred without traffic loss.

The following debugging was enabled before the metric change:

```

# on PE-3:
debug {
  router "Base" {
    ldp {
      peer 192.0.2.2 {
        packet {
          label {
            detail true
          }
          init {
            detail true
          }
        }
      }
      peer 192.0.2.4 {
        packet {

```

```

        label {
            detail true
        }
        init {
            detail true
        }
    }
}

```

The first trace message shows that label 524282 is advertised to PE-4. MBB is requested, as follows:

```

# on PE-3:
12 2023/09/07 00:02:53.483 CEST MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Label Mapping packet (msgId 157) to 192.0.2.4:0
Protocol version = 1
Label 524282 advertised for the following FECs
P2MP: root = 192.0.2.1, T: 1, L: 4, TunnelId: 5000
MP Status MBB = REQ
"

```

The next message is a notification from PE-4 confirming that there is no fatal error and MBB can be applied, as follows:

```

# on PE-3:
13 2023/09/07 00:02:53.485 CEST MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Recv Notification packet (msgId 157) from 192.0.2.4:0
Protocol version = 1
Status Code = MPStatus (0x00000040) Non-fatal
Causing message Id = 0
Causing message type = NULL
P2MP: root = 192.0.2.1, T: 1, L: 4, TunnelId: 5000
MP Status MBB = ACK
"

```

The following message is a label withdraw message for label 524283 sent to PE-2:

```

# on PE-3:
14 2023/09/07 00:02:53.485 CEST MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Label Withdraw packet (msgId 158) to 192.0.2.2:0
Protocol version = 1
Label 524283 withdrawn for the following FECs
P2MP: root = 192.0.2.1, T: 1, L: 4, TunnelId: 5000
"

```

The last message is a label release message for label 524283 received from PE-2, as follows. This message is only sent after the new P2MP LSP is set up.

```

# on PE-3:
15 2023/09/07 00:02:53.486 CEST MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Recv Label Release packet (msgId 157) from 192.0.2.2:0
Protocol version = 1
Label 524283 released for the following FECs
P2MP: root = 192.0.2.1, T: 1, L: 4, TunnelId: 5000
"

```

## Conclusion

Multicast LDP provides extensions to the LDP protocol for the setup of P2MP and MP2MP LSPs in MPLS networks. mLDP is simple to configure compared to RSVP. FRR and MBB are supported for mLDP.

# Path MTU Discovery

This chapter provides information about Path MTU discovery.

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

## Applicability

This chapter was initially written for SR OS Release 14.0.R7, but the CLI in the current edition corresponds to SR OS Release 21.2.R1.

## Overview

The Maximum Transmission Unit (MTU) is the largest packet size (in bytes) that a network can transmit. IP datagrams larger than the MTU are fragmented into smaller packets before being sent. [Table 4: MTU types](#) describes the MTU types that are supported in SR OS at both port and service level.

*Table 4: MTU types*

MTU type	Description
Port MTU	Maximum frame size on a physical wire
Service MTU	Maximum end-to-end frame size sent from the customer across an L2 VPN service
SDP path MTU	Maximum frame size of encapsulated packets sent over the SDP between service endpoints in IP/MPLS VPN
VC MTU	Maximum IP payload size that can be carried inside the tunnel. The VC MTU is derived from the service MTU and negotiated by T-LDP.
LSP path MTU	MTU value negotiated by RSVP path/resv messages
OSPF MTU	Maximum size of the OSPF packet
IP MTU	Used in L3 VPN services (IES or VPRN). Maximum IP packet size that L3 VPN customers can send across the provider network.



[Table 5: MTU values for Ethernet frames](#) lists the values for the MTU types for Ethernet frames. In SR OS, the MTU value never includes the Frame Check Sequence (FCS).

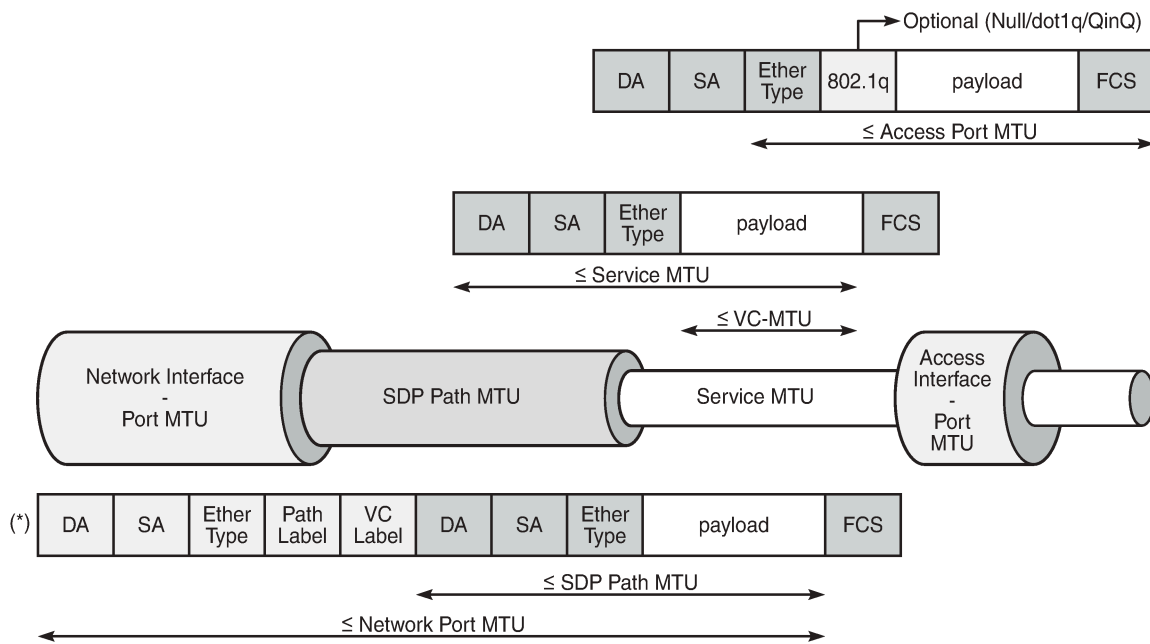
*Table 5: MTU values for Ethernet frames*

MTU type	Value
Access Port MTU	Configurable in the port context. Value should be greater than or equal to the sum of the service MTU and the port encapsulation overhead (0 for null, 4 for dot1q, 8 for QinQ).
Network Port MTU	Configurable in the port context. Value should be greater than or equal to the sum of the SDP path MTU, the MPLS labels (transport, service, hash (entropy), and OAM labels), and an Ethernet header (possibly with a VLAN tag for dot1q).
Service MTU	Configurable in the service context. Maximum payload (IP + Ethernet) that the service offers to the client. Only used in L2 services.
SDP path MTU	By default, not configured. Derived from the network port MTU. Should be, at a minimum, the value of the service MTU. Should be, at a maximum, the result of {network port MTU - 2 labels - Ethernet header}. However, the service will become operationally up when the SDP path MTU is higher. The SDP path MTU need not match on both sides of the SDP.
VC MTU	Not configurable. Derived from the service MTU and negotiated by T-LDP. The VC MTU value must match the other side. VC MTU = service MTU - 14 bytes (Ethernet header).
LSP path MTU	Derived from the port MTU of the network port
OSPF MTU	MTU negotiated by OSPF and derived from the port MTU or administratively set
IP MTU	Configurable in the L3 routing interfaces

The values of the first five MTU types listed in Table 2 are important in getting L2 services to an operational state of up. For L3 services, the IP MTU is used instead of the service MTU.

[Figure 72: L2 services MTUs for Ethernet frames](#) shows the MTUs used for Ethernet frames in an L2 service, such as an Epipe or a VPLS service.

Figure 72: L2 services MTUs for Ethernet frames



(\*) Optionally additional 802.1q header (4 bytes) for dot1q

26371

The VC MTU contains the IP payload. The service MTU contains IP payload and Ethernet header. The SDP path MTU must be greater than or equal to the service MTU. Typically, the VLAN tags are stripped at the service ingress, unless VLAN range SAPs are defined and one VLAN tag is preserved. The physical port MTU on an Ethernet access interface needs to be set to at least 1514 for null encapsulation (1500 + 14 (Ethernet header)), at least 1518 for dot1q (1500 + 14 + 4 (dot1q)), and at least 1522 for QinQ (1500 + 14 + 4 + 4).

**Figure 73: Minimum network port MTU for Ethernet frames in MPLS encapsulation** shows the minimum physical MTU on network interfaces for a router that needs to support services offering a 1514 byte service payload over MPLS for Ethernet.

Figure 73: Minimum network port MTU for Ethernet frames in MPLS encapsulation

Overhead	Ethernet
Service Payload	1514
MPLS tag used as service ID	4
MPLS tag used for egress LSP	4
Optionally, more MPLS tags	(n*4)
Ethernet Header	14
Total	1536 (+ n*4)

Maximum 12 MPLS labels.  
Optionally 1 VLAN tag for dot1q.

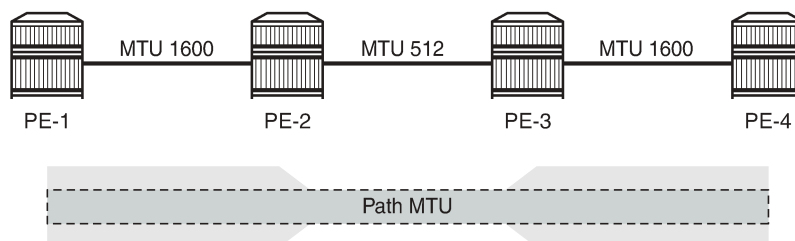
26372

The network port MTU must be at least the maximum service MTU to be supported plus the largest encapsulation type used. The SDP path MTU is at least equal to the service MTU, which is at a minimum 1514 for a service running on a typical Ethernet access interface. This is also valid when the access interface is dot1q or QinQ, because the VLAN tags are stripped at ingress and replaced by the appropriate VLAN tag at egress, unless VLAN range SAPs are defined, in which case one VLAN tag is preserved. The VC tag (service ID) adds a 4 byte service label, the MPLS path adds—at least—one 4 byte transport label, and the Ethernet header adds 14 bytes, for a total of at least 1536 for Ethernet encapsulation. For MPLS, the maximum label stack depth is 12.

The default behavior in SR OS is that the network port MTU is set to its maximum per MDA type, if the network port MTU is not explicitly configured. By default, the SDP path MTU is derived from the network port MTU. For example, when the network port is set to 1600, the SDP path MTU = 1600 (network port MTU) - 4 (MPLS service label) - 4 (MPLS path label) - 14 (Ethernet label) = 1578. However, the SDP path MTU is only accurate when the end-to-end path is considered and the lowest network port MTU in the path is taken.

Figure 74: Path MTU shows that the path MTU is determined by the lowest MTU along the path that the service needs to transit. When IP hosts transmit IP datagrams to each other, the path MTU is the largest size for which no fragmentation is required along the path.

Figure 74: Path MTU



26373

## Path MTU discovery (PMTUD)

PMTUD is a technique for dynamically discovering the MTU size on the network path between two IP hosts, to maximize packet efficiency and avoid packet fragmentation. PMTUD is standardized in RFC 1191 and for IPv6 in RFC 1981.

PMTUD can be enabled in LDP and BGP in the following contexts:

```
[/]
A:admin@Dut-A# tree flat detail | match path-mtu-discovery
configure groups group <string> router <string> bgp group <string> path-mtu-discovery <boolean>
configure groups group <string> router <string> bgp neighbor <string | ipv4-address-with-zone |
  ipv4-address | ipv6-address-linklocal-with-zone | ipv6-address | ipv6-address-with-zone> path-
  mtu-discovery <boolean>
configure groups group <string> router <string> bgp path-mtu-discovery <boolean>
configure groups group <string> router <string> ldp tcp-session-parameters peer. For BGP, -
  transport <string | ipv4-address | ipv6-address> path-mtu-discovery <boolean>
configure groups group <string> service vprn <string> bgp group <string> path-mtu-discovery
  <boolean>
configure groups group <string> service vprn <string> bgp neighbor <string | ipv4-address-with-
  zone | ipv4-address | ipv6-address-linklocal-with-zone | ipv6-address | ipv6-address-with-zone>
  path-mtu-discovery <boolean>
configure groups group <string> service vprn <string> bgp path-mtu-discovery <boolean>
configure router <string> bgp group <string> path-mtu-discovery <boolean>
configure router <string> bgp neighbor <ipv4-address-with-zone | ipv4-address | ipv6-address-
  linklocal-with-zone | ipv6-address | ipv6-address-with-zone> path-mtu-discovery <boolean>
configure router <string> bgp path-mtu-discovery <boolean>
configure router <string> ldp tcp-session-parameters peer-transport <ipv4-address | ipv6-
  address> path-mtu-discovery <boolean>
configure service vprn <string> bgp group <string> path-mtu-discovery <boolean>
configure service vprn <string> bgp neighbor <ipv4-address-with-zone | ipv4-address | ipv6-
  address-linklocal-with-zone | ipv6-address | ipv6-address-with-zone> path-mtu-discovery
  <boolean>
configure service vprn <string> bgp path-mtu-discovery <boolean>
state system connections statistics tcp path-mtu-discovery-backoff <number>
```

PMTUD can be enabled in BGP at different levels: global, per group, or per neighbor. PMTUD can be enabled in BGP in the base router or in a VPRN. For LDP, PMTUD is enabled per peer.

PMTUD works by setting the Don't Fragment (DF) option bit in the IP header of outgoing packets. The source assumes initially that the path MTU is the MTU of its egress interface. Any device along the path with an MTU smaller than the IPv4 packet will drop the packet and notify the source by sending back an Internet Control Message Protocol (ICMP) "Fragmentation Needed" (type 3, code 4) error message containing its MTU. IPv6 packets larger than the MTU will also be dropped in which case an ICMPv6 error message "Packet Too Big" (type 2, code 0) containing its MTU will be sent back. The source can then reduce its path MTU to this received MTU. The process repeats until the MTU is small enough to traverse the entire path without fragmentation.

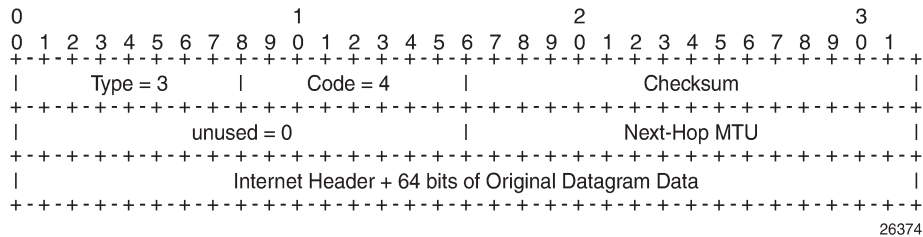
If the path MTU changes to a lower value after the connection is set up, the first larger packet will cause an ICMP error message and the new, lower path MTU will be determined.

PMTUD is used to determine the most efficient packet size for protocols or applications that may send large packets or large data transfers, including BGP updates, LDP, IGP, FTP/TFTP/SCP transfers. With PMTUD enabled, each connection can start with the maximum MTU—based on egress MTU—then allow remote and/or transit routers to lower the effective MTU for the session if the current MTU is too large for one of their next hops. The path MTU is handled and tracked on a per session/connection basis.

All routers along the path must be able to send ICMP error messages of type 3 ("Destination Unreachable") and code 4 ("Fragmentation Needed").

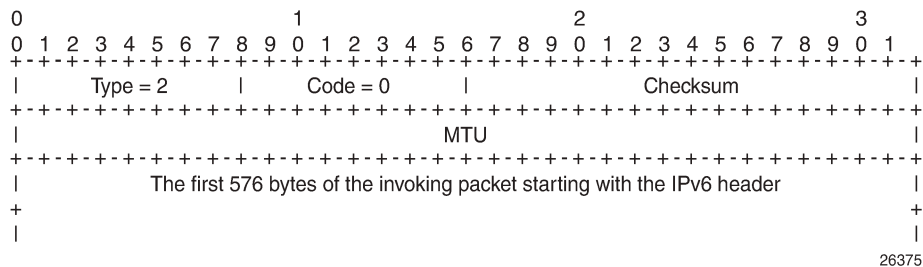
[Figure 75: ICMP "Destination Unreachable" Message - Fragmentation Needed](#) shows the format of such an ICMP message. The next hop MTU is the MTU of the egress interface to the destination of the packet on the router that dropped the packet. The MTU is a count of the octets of the IP header and IP data, without lower-level headers.

*Figure 75: ICMP "Destination Unreachable" Message - Fragmentation Needed*



The mechanism for IPv6 is similar, but the format of the ICMPv6 message is different. For IPv6, the router will send an ICMPv6 error message of type 2 ("Packet Too Big") and code 0, as shown in [Figure 76: ICMPv6 "Packet Too Big" message](#). The MTU field is populated with the MTU of the egress interface to the destination of the packet on the router that dropped the packet. The MTU is a count of the octets of the IP header and IP data, but no lower-level headers.

*Figure 76: ICMPv6 "Packet Too Big" message*



When PMTUD is enabled, the IP MTU is initially set to the egress MTU size, based on the source IP interface for that session. When a node along the path is unable to forward a packet due to a smaller MTU, the node drops the packet and sends back an ICMP error message with the MTU of the egress interface. The node that receives the ICMP error message will adjust its MTU accordingly. The IP header and the following bytes of the original IP datagram should be used to determine which connection caused the error.

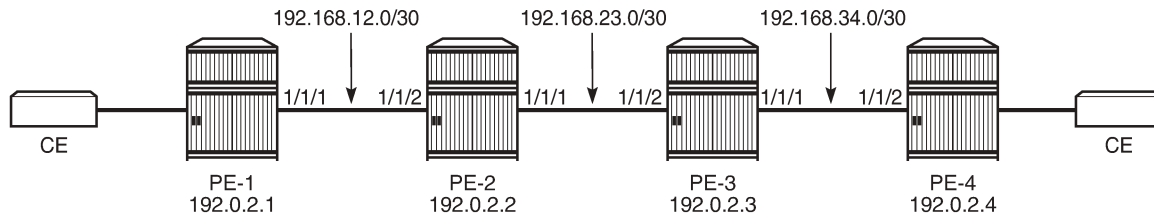
## Configuration

The following examples are configured:

- PMTUD in LDP for an IPv4 peer
- PMTUD in LDP for an IPv6 peer
- PMTUD in BGP for an IPv4 peer
- PMTUD in BGP for an IPv6 peer

**Figure 77: Example topology** shows the example topology with four PE nodes in autonomous system 64496. The interfaces have IPv4 and IPv6 addresses, but in this figure, only the IPv4 addresses are shown.

*Figure 77: Example topology*



26376

The initial configuration on the nodes includes:

- Cards, MDAs, ports
- Router interfaces with IPv4 and IPv6 address
- IS-IS as IGP on all interfaces between the PEs (alternatively, OSPF can be used)
- LDP enabled on all interfaces between the PEs for IPv4 and IPv6

The initial configuration on PE-1 is as follows:

```
# on PE-1:
configure {
  router "Base" {
    interface "int-PE-1-PE-2" {
      port 1/1/1
      ipv4 {
        primary {
          address 192.168.12.1
          prefix-length 30
        }
      }
      ipv6 {
        address 2001:db8:12:: {
          prefix-length 127
        }
      }
    }
    interface "system" {
      ipv4 {
        primary {
          address 192.0.2.1
          prefix-length 32
        }
      }
      ipv6 {
        address 2001:db8::1 {
          prefix-length 128
        }
      }
    }
  }
  isis 0 {
    admin-state enable
    ipv6-routing native
    area-address [49.0001]
    interface "int-PE-1-PE-2" {
```

```

    interface-type point-to-point
  }
  interface "system" {
  }
}
ldp {
  interface-parameters {
    interface "int-PE-1-PE-2" {
      ipv4 {
      }
      ipv6 {
      }
    }
  }
}
}
}
}
}

```

The configuration is similar on the other PEs.

In the example, the default service MTU is used (= 1514 bytes), the access port MTU is 1518 (dot1q encapsulation), and a network port MTU (1600) is set, high enough to support the service MTU (1514):

```

[/]
A:admin@PE-1# show port
=====
Ports on Slot 1
=====
Port      Admin Link Port   Cfg  Oper  LAG/  Port  Port  Port  C/QS/S/XFP/
Id        State  State State  MTU  MTU   Bndl  Mode  Encp  Type  MDIMDX
-----
1/1/1     Up     Yes  Up     1600 1600  - netw null xgige 10GBASE-LR *
1/1/2     Down  No   Down   1578 1578  - netw null xgige 10GBASE-LR *
1/1/3     Down  No   Down   1578 1578  - netw null xgige 10GBASE-LR *
1/1/4     Down  No   Down   1578 1578  - netw null xgige 10GBASE-LR *
1/1/5     Down  No   Down   1578 1578  - netw null xgige 10GBASE-LR *
1/1/6     Down  No   Down   1578 1578  - netw null xgige 10GBASE-LR *
1/1/7     Down  No   Down   1578 1578  - netw null xgige 10GBASE-LR *
1/1/8     Down  No   Down   1578 1578  - netw null xgige 10GBASE-LR *
1/1/9     Down  No   Down   1578 1578  - netw null xgige 10GBASE-LR *
1/1/10    Down  No   Down   1578 1578  - netw null xgige 10GBASE-LR *
1/2/1     Up     Yes  Up     1518 1518  - accs dotq xgige 10GBASE-LR *
1/2/2     Up     Yes  Up     1518 1518  - accs dotq xgige 10GBASE-LR *
1/2/3     Down  No   Down   1578 1578  - netw null xgige 10GBASE-LR *
1/2/4     Down  No   Down   1578 1578  - netw null xgige 10GBASE-LR *
1/2/5     Down  No   Down   1578 1578  - netw null xgige 10GBASE-LR *
1/2/6     Down  No   Down   1578 1578  - netw null xgige 10GBASE-LR *
1/2/7     Down  No   Down   1578 1578  - netw null xgige 10GBASE-LR *
1/2/8     Down  No   Down   1578 1578  - netw null xgige 10GBASE-LR *
1/2/9     Down  No   Down   1578 1578  - netw null xgige 10GBASE-LR *
1/2/10    Down  No   Down   1578 1578  - netw null xgige 10GBASE-LR *
=====
Ports on Slot A
=====
Port      Admin Link Port   Cfg  Oper  LAG/  Port  Port  Port  C/QS/S/XFP/
Id        State  State State  MTU  MTU   Bndl  Mode  Encp  Type  MDIMDX
-----
A/1       Up     Yes  Up     1514 1514  - netw null faste MDI
A/3       Down  No   Down   1514 1514  - netw null faste
A/4       Down  No   Down   1514 1514  - netw null faste

```

```

=====
Ports on Slot B
=====
Port      Admin Link Port   Cfg  Oper  LAG/  Port Port Port   C/QS/S/XFP/
Id        State      State MTU   MTU   Bndl  Mode Encp Type  MDIMDX
-----
B/1       Up      No   Ghost  1514 1514   - netw null faste
B/3       Down   No   Ghost  1514 1514   - netw null faste
B/4       Down   No   Ghost  1514 1514   - netw null faste
=====

```

The network port MTU on the link between PE-2 and PE-3 is configured to 512 for IPv4. For IPv6, this network port MTU on the link between PE-2 and PE-3 is reconfigured with a value of 1300.

The service MTU is 1514, the SAP MTU is 1518 (dot1q encapsulation on access port), and the SDP MTU is 1578 (= 1600 (network port MTU) - 14 (Ethernet) - 8 (2 MPLS labels: service label and transport label)), as shown for an Epipe service on PE-1. The configuration for SDP 14 is shown in section [SDP path MTU for IPv4](#); for Epipe\_100\_name, in section [PMTUD for LDP IPv4](#). This SDP MTU does not consider the lowest network port MTU in the path, but only the local network MTU.

```

[/]
A:admin@PE-1# show service id 100 base

=====
Service Basic Information
=====
Service Id       : 100                Vpn Id           : 0
Service Type     : Epipe
MACSec enabled   : no
Name             : Epipe_100_name
Description      : (Not Specified)
Customer Id      : 1                Creation Origin   : manual
Last Status Change: 08/05/2021 17:00:44
Last Mgmt Change : 08/05/2021 16:52:43
Test Service     : No
Admin State      : Up                Oper State        : Up
MTU             : 1514
Vc Switching     : False
SAP Count        : 1                SDP Bind Count    : 1
Per Svc Hashing  : Disabled
Vxlan Src Tep Ip : N/A
Force QTag Fwd   : Disabled
Oper Group       : <none>

-----
Service Access & Destination Points
-----
Identifier                Type      AdmMTU  OprMTU  Adm  Opr
-----
sap:1/2/1:100             q-tag    1518    1518    Up   Up
sdp:14:100 S(192.0.2.4)   Spok     0        1578    Up   Up
=====

```

### SDP path MTU for IPv4

The network port MTU is configured to 512 on the interfaces between PE-2 and PE-3, as follows:

```

# on PE-2:
configure {

```



```

port 1/1/1 {
  ethernet {
    mtu 512
  }
}

```

```

# on PE-3:
configure {
  port 1/1/2 {
    ethernet {
      mtu 512
    }
  }
}

```

On PE-1, SDP 14 is configured toward PE-4, as follows:

```

# on PE-1:
configure {
  service {
    sdp 14 {
      admin-state enable
      delivery-type mpls
      ldp true
      far-end {
        ip-address 192.0.2.4
      }
    }
  }
}

```

The configuration is similar on PE-4, but with a far end of 192.0.2.1 instead.

The SDP path MTU is derived from the lowest network port MTU in the path: 512 - 14 (Ethernet header) - 4 (MPLS service label) - 4 (MPLS path label) = 490. This can be verified on PE-1 for the end-to-end path with the following OAM command (in classic CLI) that sends packets with an incrementing size: from 400 to 500 bytes in steps of 10 bytes. The packet with size 490 bytes gets a response, whereas the packet with size 500 gets a timeout.

```
A:PE-1# oam sdp-mtu 14 size-inc 400 500 step 10
```

Size	Sent	Response
400	.	Success
410	.	Success
420	.	Success
430	.	Success
440	.	Success
450	.	Success
460	.	Success
470	.	Success
480	.	Success
490	.	Success
500	...	Request Timeout

**Maximum Response Size: 490**

The next step is to repeat the OAM command to send packets with incrementing size from 490 to 500 in steps of 1:

```
A:PE-1# oam sdp-mtu 14 size-inc 490 500 step 1
Size   Sent   Response
-----
490    .       Success
491    ...     Request Timeout

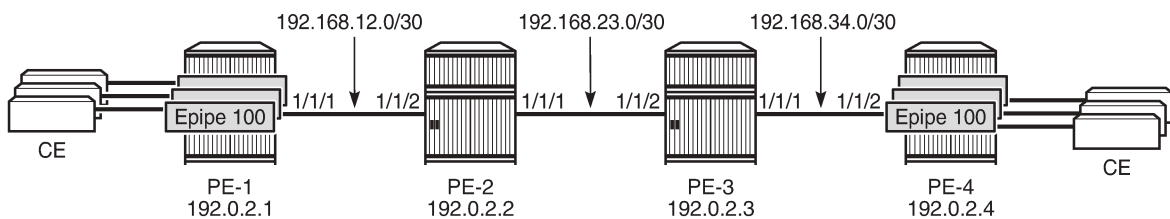
Maximum Response Size: 490
```

The SDP path MTU is 490 bytes.

## PMTUD for LDP IPv4

Figure 78: Multiple Epipes Using LDP SDPs shows that multiple Epipe services are configured on PE-1 and PE-4.

Figure 78: Multiple Epipes Using LDP SDPs



26377

The following multiple Epipes are configured on PE-1:

```
# on PE-1:
configure {
  service {
    epipe "Epipe_100_name" {
      admin-state enable
      service-id 100
      customer "1"
      spoke-sdp 14:100 {
      }
      sap 1/2/1:100 {
      }
    }
    epipe "Epipe_101_name" {
      admin-state enable
      service-id 101
      customer "1"
      spoke-sdp 14:101 {
      }
      sap 1/2/1:101 {
      }
    }
    ---snip--- for 102 through 108
    epipe "Epipe_109_name" {
      admin-state enable
      service-id 109
      customer "1"
    }
  }
}
```

```

        spoke-sdp 14:109 {
        }
        sap 1/2/1:109 {
        }
    }
}

```

The following configuration enables PMTUD for LDP IPv4 peer 192.0.2.4 on PE-1. The configuration is similar on PE-4.

```

# on PE-1:
configure {
    router "Base" {
        ldp {
            tcp-session-parameters {
                peer-transport 192.0.2.4 {
                    path-mtu-discovery true
                }
            }
        }
    }
}

```

```
[/]A:admin@PE-1# show router ldp tcp-session-parameters ipv4
```

```

=====
LDP IPv4 TCP Session Parameters
=====
-----
Peer Transport: 192.0.2.4
-----
Authentication Key : Disabled           Path MTU Discovery : Enabled
Auth key chain    :                     Min-TTL           : 0
=====
No. of IPv4 Peers: 1
=====

```

When LDP is disabled and re-enabled on PE-1, all label mappings are signaled again.

```

# on PE-1:
configure {
    router "Base" {
        ldp {
            admin-state disable
            commit
            admin-state enable
            commit
        }
    }
}

```

The size of the LDP label mapping messages may exceed the MTU between PE-2 and PE-3. The DF bit is set, so the packet is discarded at the egress of PE-2 to PE-3. PE-2 sends an ICMP error message of type 3 and code 4 to PE-1. The following ICMP error message is received on PE-1 when debugging is enabled for ICMP (in classic CLI):

```

# on PE-1:
debug
router

```

```

ip
  icmp
exit all

[/]
A:admin@PE-1# show log log-id 2

=====
Event Log 2 log-name 2
=====
Description : (Not Specified)
Memory Log contents [size=100  next event=3  (not wrapped)]

2 2021/08/05 17:17:08.614 CEST MINOR: DEBUG #2001 Base PIP
"PIP: ICMP
instance 1 (Base), interface index 2 (int-PE-1-PE-2),
ICMP ingressing on int-PE-1-PE-2:
  192.168.23.1 -> 192.0.2.1
  type: Destination Unreachable (3)  code: Fragmentation Needed and Don't Fragment was Set (4)
"
---snip---
  
```

On the egress interface "int-PE-2-PE-3" on PE-2, the network MTU is 512, the IP MTU is 498 (= 512 - 14 (Ethernet header)), and the TCP Maximum Segment Size (OperMss) is 458 (= 498 - 20 (IP header) - 20 (TCP header)), as shown on PE-1:

```

[/]
A:admin@PE-1# show system connections port 646

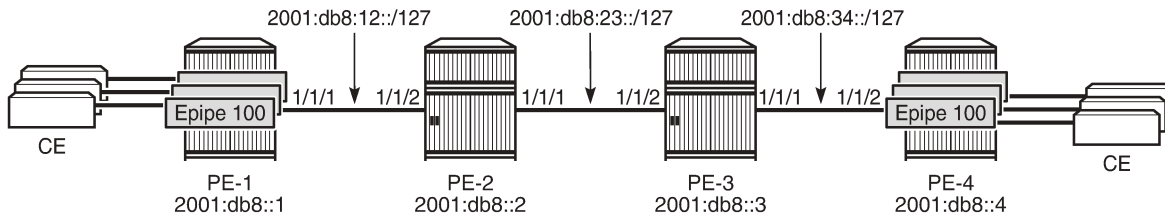
=====
Connections
=====
Prot   RecvQ   TxmtQ   Local Address           State
      RcvdMss OperMss  Remote Address          vRtrID
-----
TCP    0        0  192.0.2.1.646          LISTEN
      0      1024  0.0.0.0.0              1
TCP    0        0  192.0.2.1.646          ESTABLISH
      1024   1024  192.0.2.2.50025        1
TCP    0        0  192.0.2.1.646          ESTABLISH
      1538   458  192.0.2.4.50933      1
---snip--- for IPv6 addresses
-----
No. of Connections: 5
=====
  
```

TCP port 646 is used for LDP messages. The LDP TCP session with PE-2 keeps the (default) TCP OperMss value of 1024, whereas the LDP TCP session with PE-4 has a reduced TCP OperMss of 458 octets. PE-1 adapts the TCP OperMss size to 458 and retransmits the LDP mapping messages to PE-4. With TCP OperMss set to 458, no fragmentation is required along the path.

### PMTUD for LDP IPv6

Multiple Epipes are configured between PE-1 and PE-4. [Figure 79: Multiple Epipes between PE-1 and PE-4 - IPv6](#) shows the IPv6 addresses used.

Figure 79: Multiple Epipes between PE-1 and PE-4 - IPv6



26378

The service configuration is the same as the preceding service configuration, but the far end of the SDP is an IPv6 address instead, as follows:

```
# on PE-1:
configure {
  service {
    sdp 146 {
      admin-state enable
      delivery-type mpls
      ldp true
      far-end {
        ip-address 2001:db8::4
      }
    }
  }
}
```

With the configured network MTU of 512 on the link between PE-2 and PE-3, SDP 146 (for IPv6) is operationally down, whereas SDP 14 (for IPv4) is up, as follows:

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

=====
Services: Service Destination Points
=====
SdpId  AdmMTU  OprMTU  Far End          Adm  Opr      Del  LSP  Sig
-----
14     0       1578   192.0.2.4       Up   Up       MPLS L    TLDP
146    0       1578   2001:db8::4     Up   Down     MPLS L    TLDP

-----
Number of SDPs : 2
-----
Legend: R = RSVP, L = LDP, B = BGP, M = MPLS-TP, n/a = Not Applicable
        I = SR-ISIS, 0 = SR-OSPF, T = SR-TE, F = FPE
=====
```

RFC 2460 IPv6 specification states that links with a configurable MTU should have an MTU of at least 1280 octets; preferably 1500 or greater to accommodate possible tunneling encapsulations without the need for fragmentation.

In this example, the network MTU on the link between PE-2 and PE-3 is configured with a value of 1300 and SDP 146 will then be operationally up.

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

```

port 1/1/1 {
    ethernet {
        mtu 1300
    }
}

```

```

# on PE-3:
configure {
    port 1/1/2 {
        ethernet {
            mtu 1300
        }
    }
}

```

The SDP path MTU for SDP 146 is 1278 (= 1300 - 14 - 4 - 4). This can be verified on PE-1 with the following OAM command (in classic CLI):

```

A:PE-1# oam sdp-mtu 146 size-inc 1270 1280 step 1
Size      Sent      Response
-----
1270      .         Success
1271      .         Success
1272      .         Success
1273      .         Success
1274      .         Success
1275      .         Success
1276      .         Success
1277      .         Success
1278      .         Success
1279      ...      Request Timeout

Maximum Response Size: 1278

```

PMTUD is enabled for LDP IPv6 peer 2001:db8::4 on PE-1, as follows:

```

# on PE-1:
configure {
    router "Base" {
        ldp {
            tcp-session-parameters {
                peer-transport 2001:db8::4 {
                    path-mtu-discovery true
                }
            }
        }
    }
}

```

```

[/]
A:admin@PE-1# show router ldp tcp-session-parameters ipv6
=====
LDP IPv6 TCP Session Parameters
=====
Peer Transport: 2001:db8::4
-----
Authentication Key : Disabled           Path MTU Discovery : Enabled
Auth key chain      :                   Min-TTL             : 0

```

```
=====
No. of IPv6 Peers: 1
=====
```

With an SDP path MTU of 1280 octets, it is extremely unlikely that LDP packets will exceed this size. An example of an ICMPv6 message that is sent when the packet is too big is shown for BGP in section [PMTUD for BGP IPv6](#).

The TCP OperMss for the IPv6 LDP connection between PE-1 and PE-4 is the default value of 1024 bytes. When the SDP path MTU is big enough for TCP segments with segments of 1024 bytes, the TCP OperMss is set to 1024, unless **tcp-mss** is configured manually on the IPv6 interfaces. This TCP OperMss value may change after an ICMPv6 "Packet Too Big" message is received on PE-1.

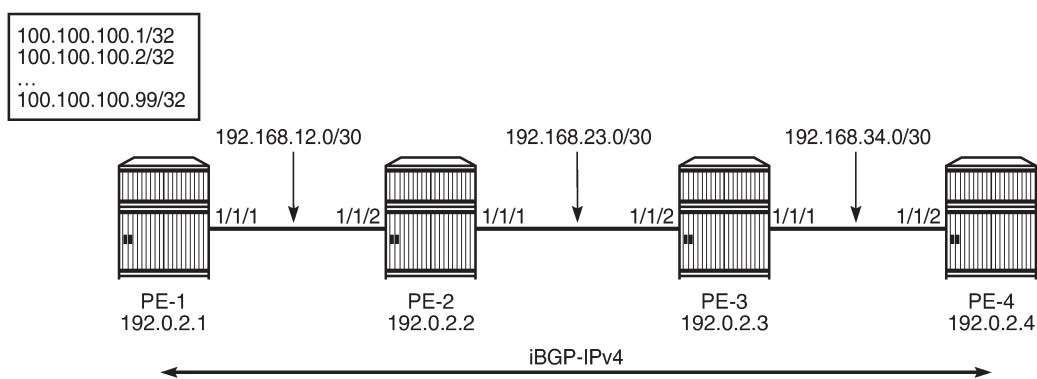
```
[/]
A:admin@PE-1# show system connections address 2001:db8::4 port 646

=====
Connections
=====
Prot  RecvQ  TxmtQ  Local Address  State
      RcvdMss OperMss Remote Address  vRtrID
-----
TCP    0        0 2001:db8::1.646  ESTABLISH
      1518   1518 2001:db8::4.50960
-----
No. of Connections: 1
=====
```

## PMTUD for BGP IPv4

[Figure 80: BGP-IPv4](#) shows that a BGP session is established between PE-1 and PE-4 for address family IPv4. Static routes on PE-1 are exported as BGP routes to PE-4.

Figure 80: BGP-IPv4



26379

The network port MTU on the link between PE-2 and PE-3 is set to 512 again:

```
# on PE-2:
configure {
  port 1/1/1 {
    ethernet {
```

```

        mtu 512
    }
}

```

```

# on PE-3:
configure {
    port 1/1/2 {
        ethernet {
            mtu 512
        }
    }
}

```

BGP is configured for address family IPv4 on PE-1, as follows:

```

# on PE-1:
configure {
    policy-options {
        policy-statement "export_static_policy" {
            entry 10 {
                from {
                    protocol {
                        name [static]
                    }
                }
                action {
                    action-type accept
                }
            }
        }
    }
    router "Base" {
        autonomous-system 64496
        bgp {
            ebgp-default-reject-policy {
                import false
                export false
            }
            group "iBGP_IPv4_name" {
                peer-as 64496
            }
            neighbor "192.0.2.4" {
                group "iBGP_IPv4_name"
                path-mtu-discovery true
                export {
                    policy ["export_static_policy"]
                }
            }
        }
    }
}

```

The export policy exports static routes as BGP routes to neighbor 192.0.2.4. PMTUD can be enabled in the global **bgp** context, per **group**, or per **neighbor**. In this example, PMTUD is enabled for neighbor 192.0.2.4. The configuration on PE-4 is similar, but with a neighbor 192.0.2.1 and without any export policy.

Also, a range of static routes is configured on PE-1 to ensure that the size of the BGP update messages will be larger than the SDP path MTU, as follows:

```

# on PE-1:

```



```
configure {
  router "Base" {
    static-routes {
      route 100.100.100.1/32 route-type unicast {
        blackhole {
          admin-state enable
        }
      }
      route 100.100.100.2/32 route-type unicast {
        blackhole {
          admin-state enable
        }
      }
      ---snip--- for 3 through 98
      route 100.100.100.99/32 route-type unicast {
        blackhole {
          admin-state enable
        }
      }
    }
  }
}
```

Debugging is enabled for ICMP (in classic CLI), as follows:

```
# on PE-1:
clear log 2
no debug
debug
  router
  ip
    icmp
  exit all
```

BGP is disabled and re-enabled to ensure that all BGP routes are re-advertised to PE-4. The BGP route update messages exceed the MTU on the egress port of PE-2 to PE-3, and PE-2 should have to fragment them to be able to forward them on the egress interface toward PE-3, but the DF bit is set. Therefore, PE-2 discards the packet and sends an ICMP error message to PE-1 of type 3 ("Destination Unreachable") and code 4 ("Fragmentation Needed and Don't Fragment was Set"). PE-1 receives the following ICMP error message:

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

=====
Event Log 2 log-name 2
=====
Description : (Not Specified)
Memory Log contents [size=100  next event=6  (not wrapped)]

5 2021/08/05 17:31:46.740 CEST MINOR: DEBUG #2001 Base PIP
"PIP: ICMP
instance 1 (Base), interface index 2 (int-PE-1-PE-2),
ICMP ingressing on int-PE-1-PE-2:
 192.168.23.1 -> 192.0.2.1
  type: Destination Unreachable (3)  code: Fragmentation Needed and Don't Fragment was Set (4)
"
---snip---
```

The following output shows that the TCP OperMss for the BGP connection between PE-1 and PE-4 is 458. TCP destination port 179 is used for BGP traffic.

```
[/]
A:admin@PE-1# show system connections port 179

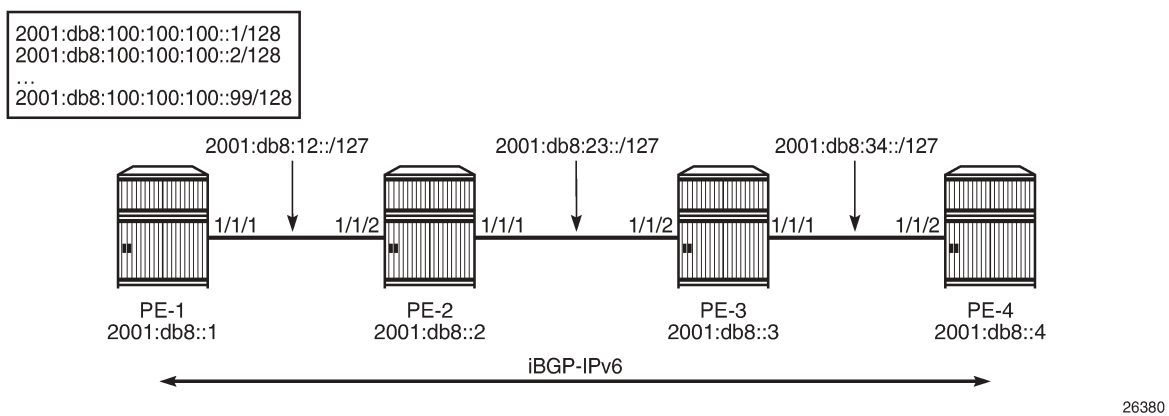
=====
Connections
=====
Prot  RecvQ  TxmtQ  Local Address          State
      RcvdMss OperMss Remote Address          vRtrID
-----
TCP   0       0  0.0.0.0.179           LISTEN
      0       1024  0.0.0.0.0             1
TCP   0       0  192.0.2.1.50595       ESTABLISH
      1538    458  192.0.2.4.179        1
TCP   0       0  ::.179                LISTEN
      0       1024  :::0                  1
-----
No. of Connections: 3
=====
```

The TCP OperMss is calculated as follows:  $512 - 14 - 20 - 20 = 458$ , where 512 is the lowest network port MTU in the path, 14 bytes are used for the Ethernet header, 20 bytes for the IPv4 header, and 20 bytes for the TCP header.

### PMTUD for BGP IPv6

Figure 81: BGP-IPv6 shows that a BGP session is established between PE-1 and PE-4 for address family IPv6. PE-1 exports a range of IPv6 routes to PE-4.

Figure 81: BGP-IPv6



The network port MTU on the link between PE-2 and PE-3 is set to 1300 again:

```
# on PE-2:
configure {
  port 1/1/1 {
    ethernet {
      mtu 1300
    }
  }
}
```

```
    }  
  }  
  
# on PE-3:  
configure {  
  port 1/1/2 {  
    ethernet {  
      mtu 1300  
    }  
  }  
}
```

The BGP configuration is similar for IPv6 to the configuration for IPv4, only the BGP address family and the neighbor addresses are different. The export policy is identical. PMTUD is enabled in the BGP group "iBGP\_IPv6\_group". The static routes have now IPv6 addresses.

```
# on PE-1:  
configure {  
  policy-options {  
    policy-statement "export_static_policy" {  
      entry 10 {  
        from {  
          protocol {  
            name [static]  
          }  
        }  
        action {  
          action-type accept  
        }  
      }  
    }  
  }  
}  
router "Base" {  
  autonomous-system 64496  
  bgp {  
    group "iBGP_IPv6_name" {  
      path-mtu-discovery true  
      peer-as 64496  
      family {  
        ipv6 true  
      }  
    }  
    neighbor "2001:db8::4" {  
      group "iBGP_IPv6_name"  
      export {  
        policy ["export_static_policy"]  
      }  
    }  
  }  
  commit  
  info  
}  
static-routes {  
  route 2001:db8:100:100:100::1/128 route-type unicast {  
    blackhole {  
      admin-state enable  
    }  
  }  
  route 2001:db8:100:100:100::2/128 route-type unicast {  
    blackhole {  
      admin-state enable  
    }  
  }  
}
```

```

---snip--- for 3 through 80
route 2001:db8:100:100:100::81/128 route-type unicast {
    blackhole {
        admin-state enable
    }
}
## The trick is to have BGP packets that will not be fragmented at the source PE-1!
## When you have 99 routes, the BGP part contains 1700 bytes and will be fragmented by PE-1.
}
}
}

```

The configuration on PE-4 resembles this configuration, but with a different neighbor address. When the group "iBGP\_IPv6\_group" is disabled and re-enabled, PE-1 advertises all the IPv6 routes to its peer 2001:db8::4. PE-2 cannot forward the large BGP messages and discards them. PE-2 sends an ICMPv6 error message to PE-1 indicating that the packet is too big (type 2, code 0). PE-1 receives the following ICMPv6 error message:

```

# on PE-1 (debugging is enabled in classic CLI):
clear log 2
no debug
debug
  router "Base"
  ip
    icmp6
  exit all

```

```

[/]A:admin@PE-1# show log log-id 2

=====
Event Log 2 log-name 2
=====
Description : (Not Specified)
Memory Log contents [size=100  next event=10  (not wrapped)]

---snip---
5 2021/08/05 17:37:33.740 CEST MINOR: DEBUG #2001 Base TIP
"TIP: ICMP6_PKT
ICMP6 ingressing on int-PE-1-PE-2 (Base):
  2001:db8:23:: -> 2001:db8::1
  Type: Packet Too Big (2)
  Code: No Code (0)
  MTU : 1286
"
---snip---

```

The MTU is 1286 and includes the IP header and the IP data, but not the Ethernet header. The calculation is as follows: 1300 - 14 = 1286, where 1300 is the lowest network port MTU in the path and 14 bytes are used for the Ethernet header.

On PE-1, the TCP OperMss for BGP traffic with destination address 2001:db8::4 is 1226, as follows:

```

[/]
A:admin@PE-1# show system connections port 179

=====
Connections
=====
Prot  RecvQ  TxmtQ  Local Address                               State

```

	RcvdMss	OperMss	Remote Address	vRtrID
TCP	0	0	0.0.0.0.179	LISTEN
	0	1024	0.0.0.0.0	1
TCP	0	0	192.0.2.1.50595	ESTABLISH
	1538	458	192.0.2.4.179	1
TCP	0	0	:::179	LISTEN
	0	1024	:::0	1
<b>TCP</b>	<b>0</b>	<b>0</b>	<b>2001:db8::1.50650</b>	<b>ESTABLISH</b>
	<b>1518</b>	<b>1226</b>	<b>2001:db8::4.179</b>	<b>1</b>
-----				
No. of Connections: 4				
=====				

The TCP OperMss is calculated as follows:  $1300 - 14 - 40 - 20 = 1226$ , where 1300 is the lowest network port MTU in the path, 14 bytes are used for the Ethernet header, 40 bytes for the IPv6 header, and 20 bytes for the TCP header. This TCP OperMss value is larger than the default value of 1024, so the ICMPv6 "Packet Too Big" message can result in a larger TCP OperMss value.

## Conclusion

PMTUD is a technique to determine the MTU size on the network path between two IP hosts, to maximize packet efficiency and avoid packet fragmentation. PMTUD can be enabled for LDP and BGP connections.

# Remote Loop-Free Alternate Node Protection

This chapter describes the Remote Loop-Free Alternate Node Protection.

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

## Applicability

This chapter was initially written for SR OS Release 16.0.R6, but the MD-CLI in the current edition corresponds to SR OS Release 21.2.R1. Remote Loop-Free Alternate (R-LFA) node protection is supported for IS-IS and OSPF in SR OS Release 16.0.R4 and later. There are no prerequisites for this configuration.

## Overview

The Loop-Free Alternates (LFAs) computed following the Remote LFA (R-LFA) specifications in RFC 7490 only guarantee point-to-point link protection by using a repair tunnel. The repair tunnel is a Segment Routed (SR) shortest path between the computing router S and the PQ-node, to ensure that the primary protected link SE is avoided. However, the R-LFA link protection algorithm does not guarantee that the repair path toward the PQ node will avoid the primary next hop router E, and that the traffic emerging from the repair tunnel at the PQ node toward the destination router will avoid the primary next hop router E.

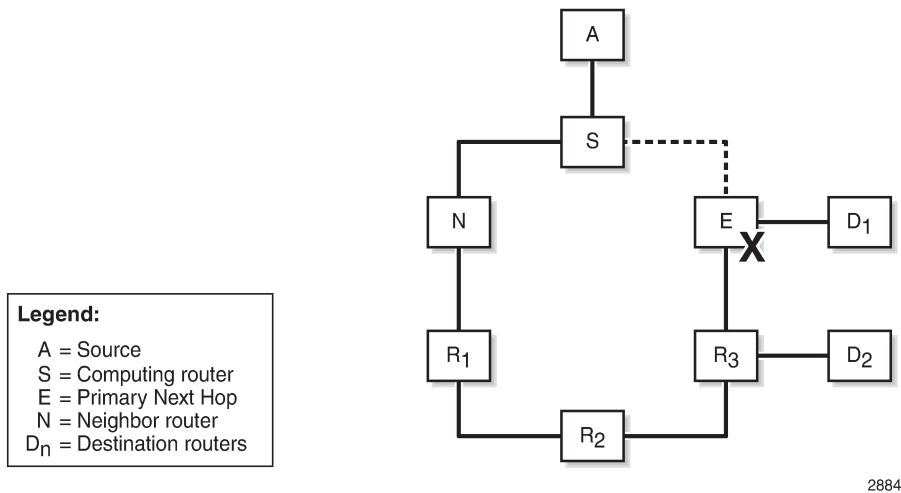
In the remainder of this chapter, SR refers to "Segment Routing", unless specified otherwise. Product and release references, such as 7750 SR and SR OS, continue to refer to "Service Router".

## Inequalities for remote LFA node protection

RFC 8102, *Remote LFA node protection*, defines the specifications to protect a path from a source A to a destination D1 or D2, when the primary next hop router E of a computing router S fails; see [Figure 82: LFA node protection - topology & denominations](#). The R-LFA alternate path through a given PQ node to a given destination comprises two path segments:

- path segment from the computing router S to the PQ node (R-LFA alternate next hop)
- path segment from the PQ node to the destination D1 or D2

Figure 82: LFA node protection - topology & denominations



28843

To ensure that an R-LFA alternate path next hop for a given destination provides node protection, none of the path segments may be affected in the event of a failure of the primary next-hop node E. The following four-step algorithm is used to satisfy this requirement:

1. Calculate the node protection extended P-space of router S with respect to the protected node E.
2. Calculate the link protection Q-space of router E with respect to the protected link SE.  
Based on the results of step 1 and 2, a list of one or more candidate PQ-routers is compiled.
3. For each candidate PQ-router, perform an additional forward Shortest Path First (SPF) run to ensure that the path from the PQ-router to the destination router does not traverse the protected router E.
4. If more than one candidate PQ-router satisfies the condition from step 3, router S chooses the PQ-router based on criteria that are specified later in this chapter.

The *node protection extended P-space* is the set of routers  $Y_i$  that are reachable from the direct neighbor(s) N of S without traversing protected router E. This excludes the direct neighbors for which there is at least one ECMP path from direct neighbor traversing router E. For a router  $Y_i$  to be member of a node protection P-space, the following inequality must be true:

$$\text{cost}(N, Y_i) < \text{cost}(N, E) + \text{cost}(E, Y_i)$$

The *link protection Q-space* is the set of routers that can reach E without traversing the protected link SE, as defined in RFC 7490. This excludes equal cost path routes that traverse the SE link. For a router  $Y_i$  to be member of the link protection Q-space, the following inequality must be true:

$$\text{cost}(Y_i, E) < \text{cost}(Y_i, S) + \text{cost}(S, E)$$

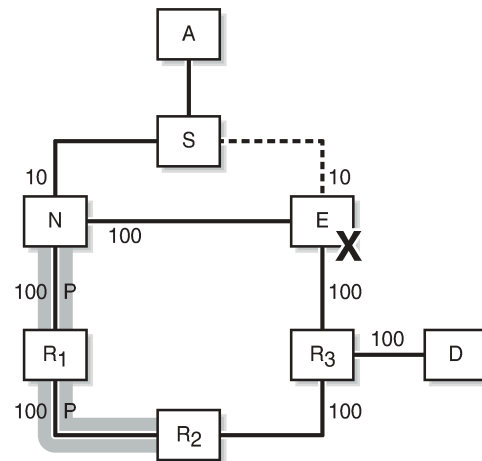
If, with respect to router E, a router  $Y_i$  is present in the node protection extended P-space and present in the link protection Q-space, it is a candidate PQ node.

Figure 83: Node protecting extended P-space shows the example topology, with metrics and the calculations in table format to determine the node protection extended P-space of router S with respect to the protected node E. Only routers N, R1, and R2 meet the inequality, and therefore belong to the node protecting extended P-space.

Figure 83: Node protecting extended P-space

Router (Yi)	cost(N,Yi)	cost(N,E)	cost(E,Yi)	Inequality met?
N	0	20	20	Yes (0<20+20)
R <sub>1</sub>	100	20	120	Yes (100<20+120)
R <sub>2</sub>	200	20	200	Yes (200<20+200)
R <sub>3</sub>	120	20	100	No (120<20+100)

**Legend:**  
 A = Source  
 S = Computing router  
 E = Primary Next Hop  
 N = Neighbor router  
 D = Destination routers

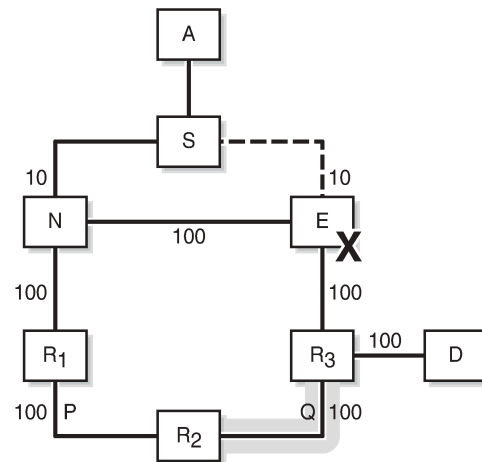


28844

Figure 84: Link protecting Q-space shows the example topology, with metrics and the calculations in table format to determine the link protecting Q-space of router E with respect to the protected link SE. Only routers R2 and R3 meet the inequality, and therefore belong to the link protecting Q-space.

Figure 84: Link protecting Q-space

**Legend:**  
 A = Source  
 S = Computing router  
 E = Primary Next Hop  
 N = Neighbor router  
 D = Destination routers

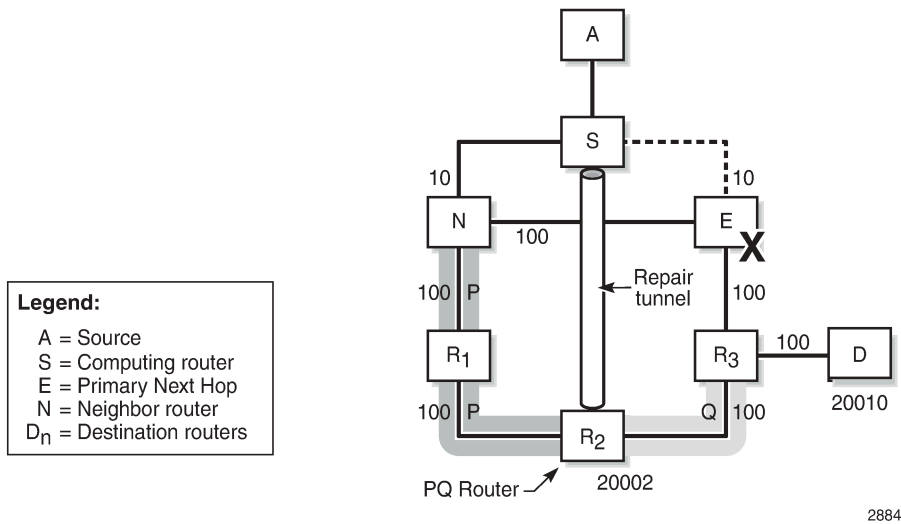


28845

Candidate PQ routers are routers that belong to the extended P-space and the Q-space. In this example, only R2 is a candidate PQ node; see [Figure 85: One candidate PQ-router – repair tunnel](#).



Figure 85: One candidate PQ-router – repair tunnel



An additional forward SPF run is required to check that the shortest path from the candidate PQ node R2 toward destination D *does not* traverse protected node E. Therefore, the following inequality must be met:

$$\text{cost}(PQ_i, D) < \text{cost}(PQ_i, E) + \text{cost}(E, D)$$

Applied to this topology, the R2-R3-D path does not go via E; therefore, R2 is a valid R-LFA node protection PQ node. The previous inequality evaluates to true, as follows:

$$200 < 200 + 200 \text{ (True)}$$

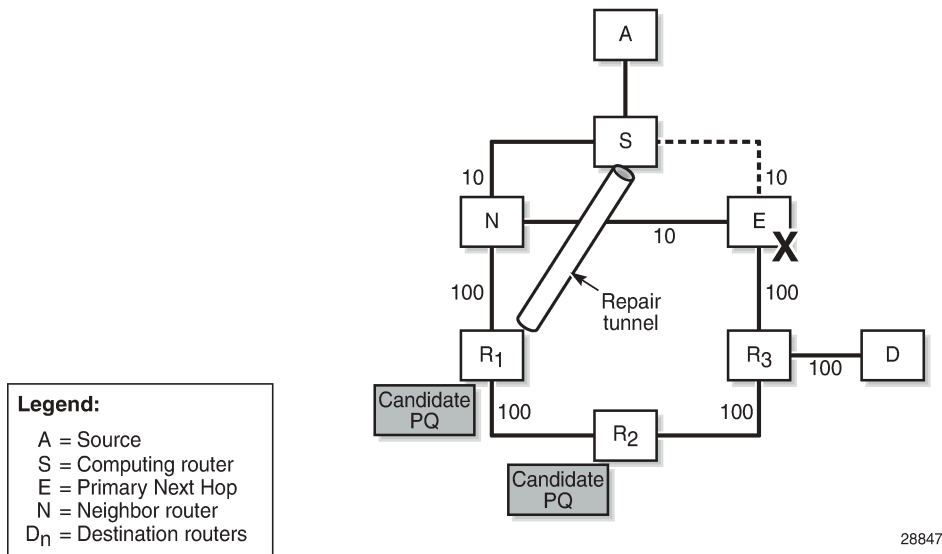
Figure 85: One candidate PQ-router – repair tunnel shows that S constructs a repair tunnel to PQ router R2. To reach destination D from S using the repair tunnel, S pushes a (20002, 20010) label stack, where 20002 and 20010 represent the node SIDs for R2 and D, respectively, while additionally setting the next-hop address to router N. The label 20002 is the first and top label swapped at N and R1, and popped at R2, while 20010 is the second label pushed at R2, swapped at R3, and ultimately popped at D.

In case multiple candidate PQ routers are available, the computing node S selects a PQ router based on the following criteria:

1. lowest IGP path cost from S
2. if multiple PQ routers satisfy (1), S selects the PQ router reachable via the neighbor with the lowest system-ID or router-ID for IS-IS and OSPF, respectively
3. if multiple PQ routers satisfy (1) and (2), S selects the PQ router with the lowest system-ID or router-ID for IS-IS and OSPF, respectively

Figure 86: Two candidate PQ routers – repair tunnel shows an example, with reduced metric between N and E, where R1 and R2 are the candidate PQ routers for protecting router E. In this example, R1 is chosen as the PQ router, because R1 is closer to S than R2. Router S will create an R-LFA repair tunnel for prefixes downstream of R3. To reach those prefixes, the R1 node SID and the D node SID are pushed, with N as the next hop. Prefixes downstream of N, R1, and R2 are unaffected by a failure of E, so they keep using N as their primary next-hop.

Figure 86: Two candidate PQ routers – repair tunnel



## LFA and remote LFA interaction

The LFA and remote LFA CLI commands are applied in the OSPF and IS-IS router contexts. The configuration in the IS-IS context is as follows:

```

configure {
  router "Base" {
    isis 0 {
      loopfree-alternate {
        remote-lfa {
          node-protect
        }
      }
    }
  }
}

```

Regular LFA is enabled through the **loopfree-alternate** command. Additionally, the **remote-lfa** and **remote-lfa { node-protect }** command can be configured. In other words, by enabling remote LFA, regular LFA is also enabled.



### Note:

A remote LFA repair tunnel is only calculated and created if no regular LFA backup next-hop exists. If this is a concern, Topology Independent LFA (TI-LFA) should be enabled; see the [Topology-Independent Loop-Free Alternate for Link Protection](#) chapter.

The LFA SPF algorithms are run using the following sequence:

1. A regular LFA is computed for each router and prefix, to provide a backup next-hop per prefix.
2. TI-LFA is computed for all routers and prefixes regardless of the outcome of step 1, and the TI-LFA computed next-hops override the regular LFA next-hops, if TI-LFA is enabled.
3. Remote LFA SPF is only run for the prefixes that are not protected after steps 1 and 2.

As a result, remote LFA next-hops, whether link or node protecting, are only computed and installed when no regular LFA next-hops are available for a given next-hop failure, assuming that TI-LFA is not configured. When the **remote-lfa { node-protect }** command is enabled, the router will prefer a node protect over a link-protect repair tunnel for a given prefix if both are found in the Remote LFA SPF computations.

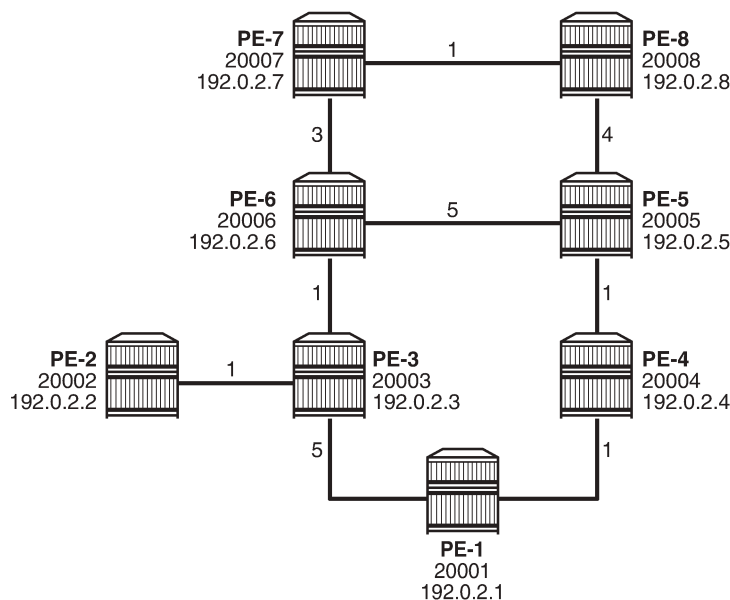
## Configuration

Three steps demonstrate the relationship between regular LFA and remote LFA, based on the example topology shown in [Figure 87: Example topology](#). The traffic flow is going from PE-7 to PE-2, and a failure of PE-6 is simulated so that PE-7 is the computing router S, PE-2 is the destination router D, PE-6 is the failing primary next hop E, and PE-8 is the primary backup neighbor N.

The following scenarios are described:

- enable regular LFA on PE-7 — node or link protection cannot be provided for prefixes downstream to PE-6
- enable remote LFA link protection on PE-7 — define the repair tunnel
- enable remote LFA node protection on PE-7 — define the repair tunnel

*Figure 87: Example topology*



28848

The configuration includes the following:

- Cards, MDAs, ports
- Single stack router interfaces (IPv4 only)
- IS-IS as IGP on the router interfaces. The metrics shown in [Figure 87: Example topology](#) are used.
- Segment routing (SR-ISIS) with node SIDs 2000x

The system addresses and the node SIDs for all routers are also shown in [Figure 87: Example topology](#).

## Regular LFA

The Segment Routing Global Block (SRGB) is defined consistently across all nodes in the network, as follows:

```
# on all nodes:
configure {
  router "Base" {
    mpls-labels {
      sr-labels {
        start 20000
        end 20099
      }
    }
  }
}
```

The IS-IS configuration on PE-7 is as follows, and has regular LFA enabled:

```
# on PE-7:
configure {
  router "Base" {
    mpls-labels {
      sr-labels {
        start 20000
        end 20099
      }
    }
  }
  isis 0 {
    admin-state enable
    level-capability 2
    area-address 49.0001.0000
    traffic-engineering true
    advertise-router-capability area
    segment-routing {
      admin-state enable
      prefix-sid-range {
        global
      }
    }
  }
  loopfree-alternate {
  }
  interface "system" {
    admin-state enable
    ipv4-node-sid {
      index 7
    }
  }
  interface "int-PE-7-PE-6" {
    admin-state enable
    interface-type point-to-point
    level 2 {
      metric 3
    }
  }
  interface "int-PE-7-PE-8" {
    admin-state enable
    interface-type point-to-point
    level 2 {
      metric 1
    }
  }
}
```

```
}

```

PE-7 calculates the *regular LFA node protection* for prefixes downstream of PE-6. The shortest path from the primary backup neighbor PE-8 to router PE-2 must be less than the shortest path from the backup neighbor PE-8 node via PE-6, so the inequality becomes:

$$\text{cost}(\text{PE-8,PE-2}) < \text{cost}(\text{PE-8,PE-6}) + \text{cost}(\text{PE-6,PE-2})$$

$$(1 + 3 + 1 + 1) < (1 + 3) + (1 + 1) \text{ (False)}$$

PE-7 calculates the *regular LFA link protection* for the PE-6-PE-7 link for prefixes downstream of PE-6. The shortest path from the primary backup neighbor PE-8 to router PE-2, must be less than the shortest path from the backup neighbor PE-8 via PE-7, so the inequality becomes:

$$\text{cost}(\text{PE-8,PE-2}) < \text{cost}(\text{PE-8,PE-7}) + \text{cost}(\text{PE-7,PE-2})$$

$$(1 + 3 + 1 + 1) < 1 + (3 + 1 + 1) \text{ (False)}$$

Because both inequalities are false, PE-7 cannot provide regular LFA PE-6 node protection or regular LFA PE-6-PE-7 link protection.

## Remote LFA with link protection

On PE-7, LFA is reconfigured so that *remote LFA with link protection* applies, as follows:

```
# on PE-7:
configure {
  router "Base" {
    isis 0 {
      loopfree-alternate {
        remote-lfa {
        }
      }
    }
  }
}
```

A repair tunnel will be established, avoiding and protecting the PE-6-PE-7 link, where the endpoint of the repair tunnel is situated on a PQ router.

[Figure 88: Link protection extended P-space calculation](#) provides the calculations in table format, along with a graphical representation, to determine the link protecting extended P-space of router PE-7 with respect to the protected PE-6-PE-7 link. Routers PE-1, PE-4, and PE-5 meet the inequality, and therefore belong to the link protecting extended P-space, meaning that they can be reached from backup neighbor PE-8 using an SPF path excluding the PE-6-PE-7 link.

Figure 88: Link protection extended P-space calculation

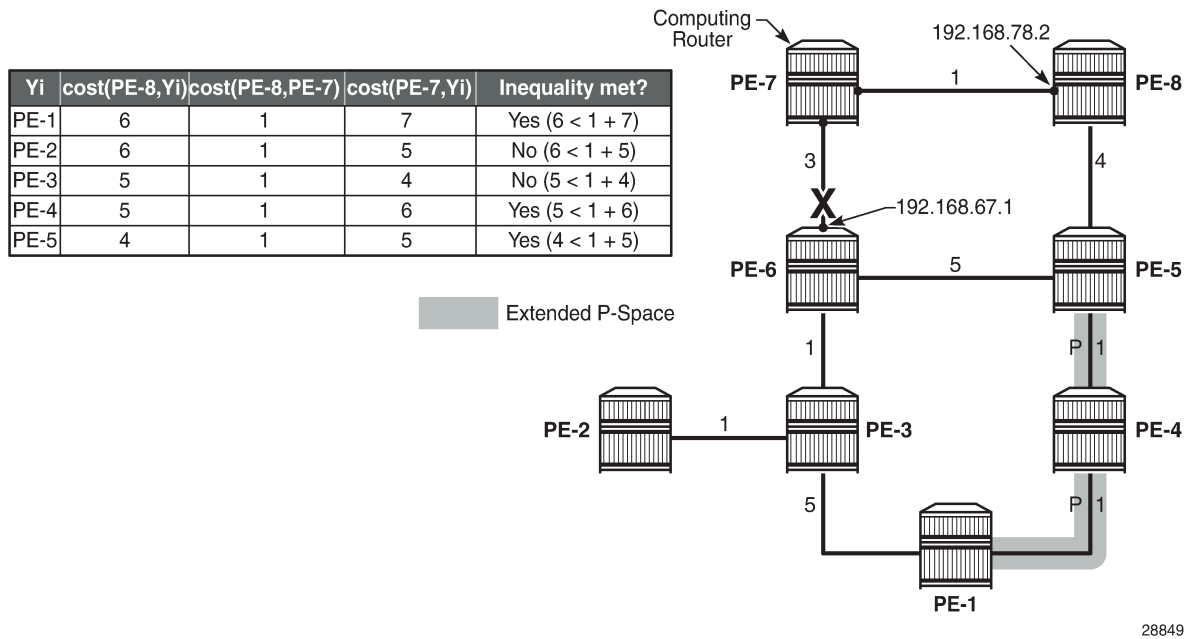
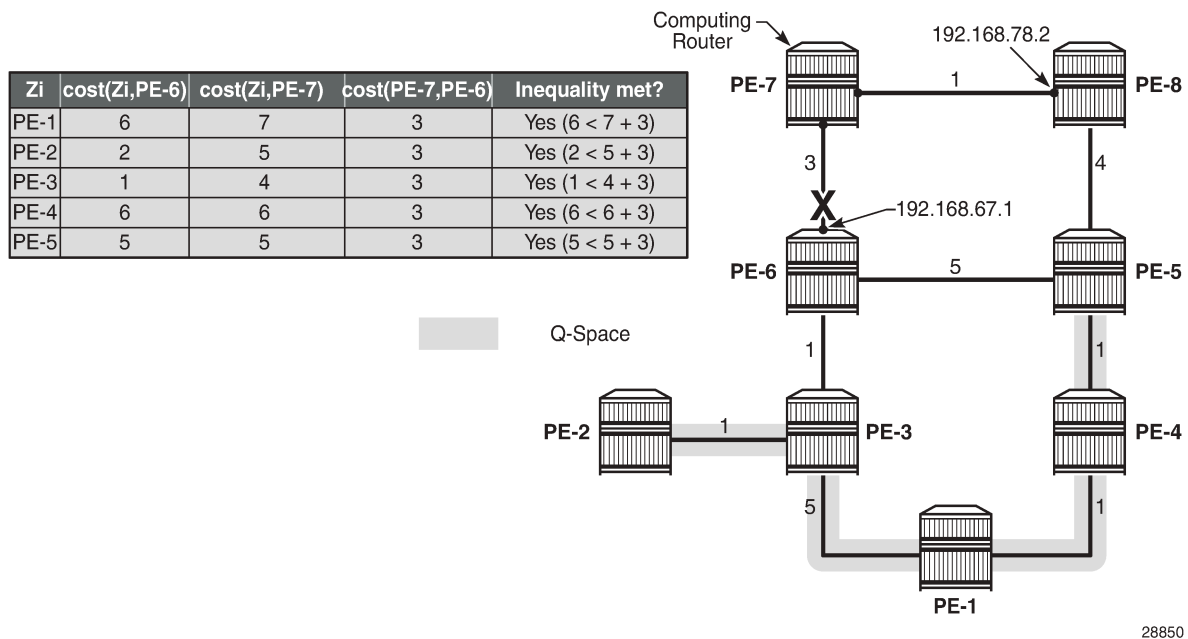


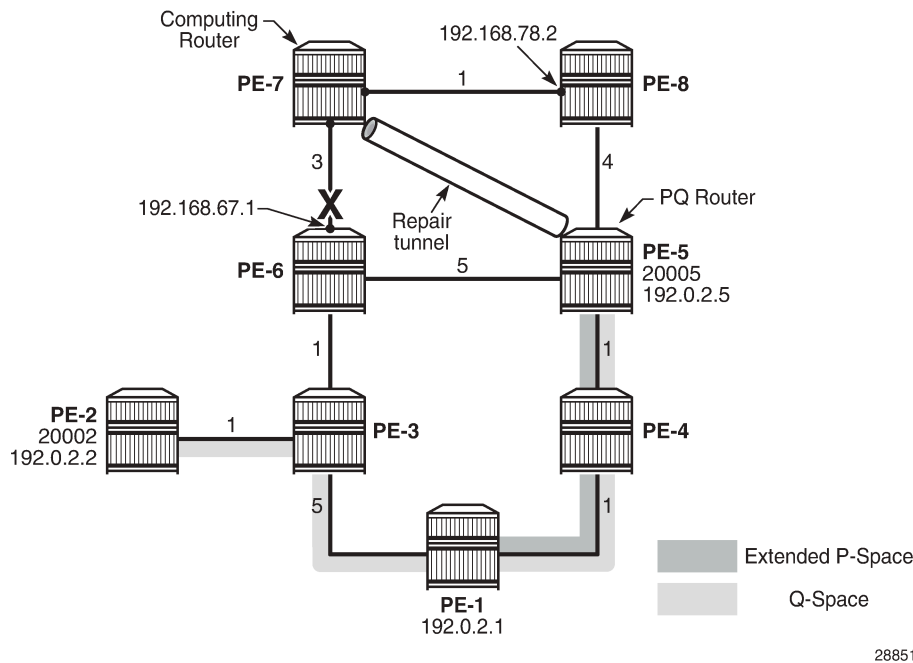
Figure 89: Link protecting Q-space calculation provides the calculations in table format, along with a graphical representation, to determine the link protecting Q-space of router PE-6 with respect to protected PE-7-PE-6 link. Routers PE-1, PE-2, PE-3, PE-4, and PE-5 meet the inequality, and therefore belong to the link protecting Q-space.

Figure 89: Link protecting Q-space calculation



**Figure 90: Repair tunnel** shows that PE-1, PE-4, and PE-5 are the candidate PQ routers. PE-5 is chosen as the repair tunnel endpoint because of the lowest path cost toward computing node PE-7 (IGP cost from PE-7 to PE-5 = 5). The closest PQ router is chosen to maximize the opportunity for load sharing traffic between the repair tunnel endpoint and the destination router.

Figure 90: Repair tunnel



On the computing node PE-7, the tunnel table for PE-2 destination (192.0.2.2) on IOM 1 shows that 192.168.67.1 is the next hop for the primary path, and that 192.168.78.2 is the next hop for the backup path, as follows. In the normal situation, the PE-7 to PE-2 traffic is routed along the PE-7-PE-6-PE-3-PE-2 path. In case of a PE-7-PE-6 link failure, the traffic on PE-7 node is pushed out with labels 20002 and 20005 to PE-8 (192.168.78.2). The top label is 20005, representing the node SID for PE-5, and 20002 is the label representing the node SID for PE-2.



**Note:**

Traffic destined for PE-2 and arriving at PE-5 with label 20005 will take the shortest path to PE-2 and therefore will traverse node PE-6.

```
[/]
A:admin@PE-7# show router fp-tunnel-table 1 192.0.2.2/32

=====
IPv4 Tunnel Table Display

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

```

NextHop (backup)
-----
192.0.2.2/32                               SR-ISIS-0           524292
20002
  192.168.67.1                             1/1/2
20002/20005
  192.168.78.2(B)                         1/1/1
-----
Total Entries : 1
=====

```

Similar information can be obtained with a **tools dump** command, as follows:

```

[/]
A:admin@PE-7# tools dump router segment-routing tunnel in-label 20002
=====
Legend: (B) - Backup Next-hop for Fast Re-Route
        (D) - Duplicate
label stack is ordered from top-most to bottom-most
=====
----+
Prefix
Sid-Type      Fwd-Type      In-Label  Prot-Inst(algoId)      Out-Label(s) Interface
              Next Hop(s)
              /Tunnel-ID |
-----+-----
192.0.2.2
Node          Orig/Transit   20002     ISIS-0
              192.168.67.1
              (B)192.168.78.2
              20002     int-PE-7-PE-6
              20005     int-PE-7-PE-8
              20002
-----+-----
No. of Entries: 1
-----+

```

Another tools command indicates the used LFA type through flags, as follows. Only RLFA link protection applies, and not node protection.

```

[/]
A:admin@PE-7# tools dump router isis sr-database prefix 192.0.2.2 sid 2
=====
Rtr Base ISIS Instance 0 SR Database
=====
SID  Label Prefix      Last-act Lev MT RtmPref TtmPref Metric IpNh SrNh
  Mtu  MtuPrim MtuBk  D xL LT Act AdvSystemId  SrErr
-----+-----
2    20002 192.0.2.2  LfaNhops 2  0  18    11    5    1    1
  1556 1564 1564  0  0  R  +R  1920.0000.2002 SR_ERR_OK
-----+-----
No. of Entries: 1
-----+
Lev = route level
IpNh = number of IP next-hops

```



```
SrNh = number of SR-tunnel next-hops
D = duplicate pending
xL = exclude from LFA
LT = LFA type (L:LFA, R:RLFA, T:TILFA, n:nodeProtection)
Act = tunnel active state (R:reported, F:failed, +:SR-ack)
=====
```

## Remote LFA with node protection

On PE-7, LFA is reconfigured so that remote LFA with node protection applies, as follows:

```
# on PE-7:
configure {
  router "Base" {
    isis 0 {
      loopfree-alternate {
        remote-lfa {
          node-protect
        }
      }
    }
  }
}
```

The general node protecting inequality from the [Overview](#) section must be used for defining the node protecting extended P-space. Using the topology from [Figure 87: Example topology](#), the inequality becomes:

$$\text{cost}(N, Y_i) < \text{cost}(N, E) + \text{cost}(E, Y_i)$$

$$\text{cost}(\text{PE-8}, Y_i) < \text{cost}(\text{PE-8}, \text{PE-6}) + \text{cost}(\text{PE-6}, Y_i)$$

[Figure 91: Node protecting extended P-space calculation](#) provides the calculations in table format, along with a graphical representation, to determine the node protecting extended P-space of router PE-7 with respect to protected PE-6 node. Routers PE-1, PE-4, and PE-5 meet the inequality, and therefore belong to the node protecting extended P-space, meaning that they can be reached from backup neighbor PE-8 through an SPF path not passing through node PE-6.



Figure 92: Link protecting Q-space calculation

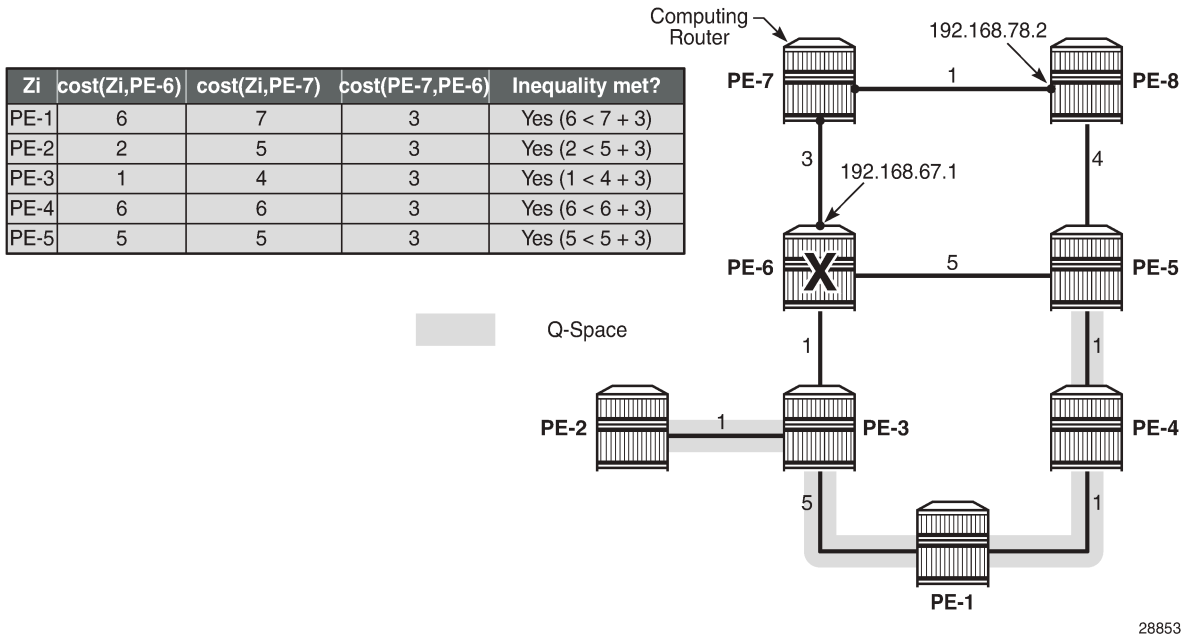
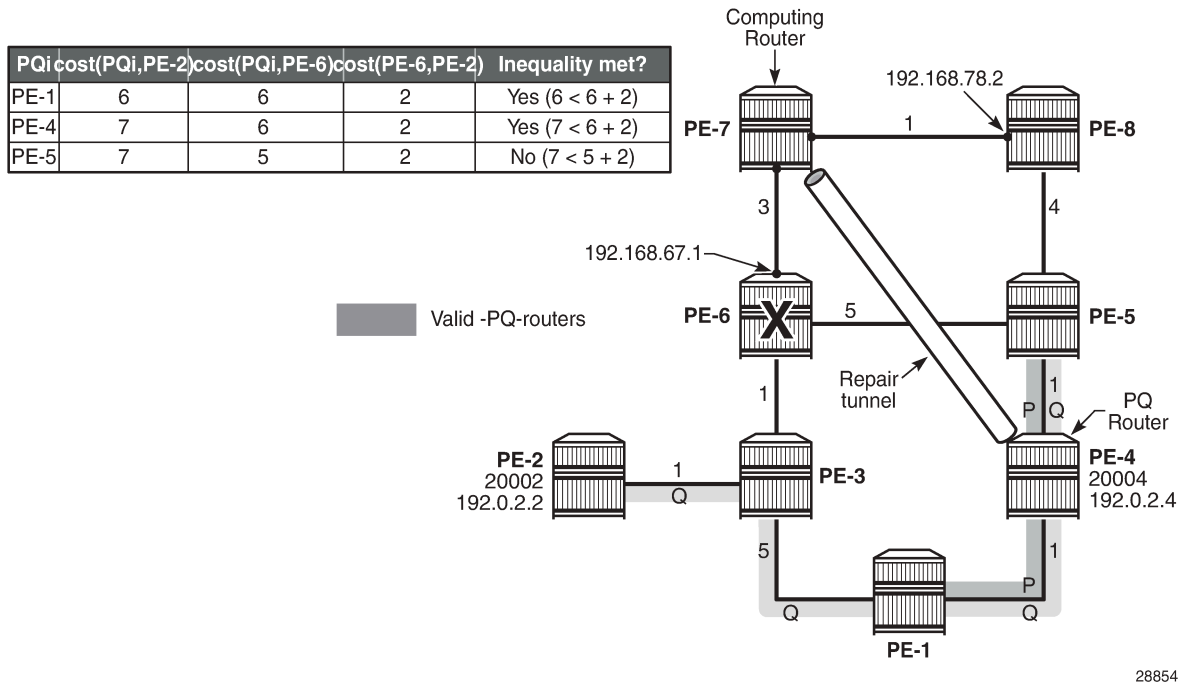


Figure 93: Validating candidate PQ routers - repair tunnel calculation shows that PE-1, PE-4, and PE-5 are the candidate PQ routers for protecting router PE-6. An additional forward SPF run is required for every candidate PQ router, to ensure that the shortest path from that candidate PQ router to destination PE-2 does not traverse the protected router PE-6. The general formula from the Overview section becomes:

$$\text{cost}(PQ_i, D) < \text{cost}(PQ_i, E) + \text{cost}(E, D)$$

$$\text{cost}(PQ_i, PE-2) < \text{cost}(PQ_i, PE-6) + \text{cost}(PE-6, PE-2)$$

Figure 93: Validating candidate PQ routers - repair tunnel calculation



After validating all three candidate PQ routers, only routers PE-1 and PE-4 are valid for terminating a repair tunnel. The tie-breaker for defining the repair tunnel termination is the lowest IGP path cost from the computing node PE-7 point of view. The cost from PE-7 to PE-4 is lower than the cost from PE-7 to PE-1 (6 < 7), so PE-4 becomes the PQ router.

The tunnel table for destination 192.0.2.2 on IOM 1 shows that 192.168.67.1 is the next hop for the primary path, and that 192.168.78.2 is the next hop for the backup path, as follows. In the normal situation, the PE-7 to PE-2 traffic is routed along the PE-7-PE-6-PE-3-PE-2 path. In case of a PE-6 node failure, the traffic from PE-7 is pushed out to PE-8 (192.168.78.2), with two labels. The label 20004 represents the node SID for PQ node PE-4 and is used as the top (first) label, while 20002 represents the node SID for PE-2 and is used as the second label.

```
[/]
A:admin@PE-7# show router fp-tunnel-table 1 192.0.2.2/32

=====
IPv4 Tunnel Table Display

Legend:
label stack is ordered from bottom-most to top-most
B - FRR Backup
=====
Destination                                Protocol    Tunnel-ID
Lbl                                          Intf/Tunnel
NextHop                                     (backup)
Lbl                                          (backup)
NextHop                                     (backup)
-----
192.0.2.2/32                               SR-ISIS-0  524292
20002                                       1/1/2
192.168.67.1
```

```

20002/20004
192.168.78.2(B) 1/1/1
-----
Total Entries : 1
-----
=====

```

Similar information can be obtained with a **tools dump** command, as follows:

```

[/]
A:admin@PE-7# tools dump router segment-routing tunnel in-label 20002
=====
====
Legend: (B) - Backup Next-hop for Fast Re-Route
        (D) - Duplicate

Label stack is ordered from top-most to bottom-most

=====
====
-----+
Prefix
|
Sid-Type      Fwd-Type      In-Label  Prot-Inst(algoId)
|
|              Next Hop(s)
|
-----+
192.0.2.2
Node          Orig/Transit  20002      ISIS-0
              192.168.67.1
              (B)192.168.78.2
              20002      int-PE-7-PE-6
              20004      int-PE-7-PE-8
              20002
-----+
No. of Entries: 1
-----+

```

Another tools command indicates the used LFA type through flags, as follows. RLFA and node protection applies.

```

[/]
A:admin@PE-7# tools dump router isis sr-database prefix 192.0.2.2 sid 2
=====
Rtr Base ISIS Instance 0 SR Database
=====
SID  Label Prefix      Last-act Lev MT RtmPref TtmPref Metric IpNh SrNh
Mtu  MtuPrim MtuBk  D xL LT Act AdvSystemId  SrErr
-----+
2    20002 192.0.2.2      LfaNhops 2  0 18    11    5    1  1
1556 1564  1564  0  0  Rn +R 1920.0000.2002 SR_ERR_OK
-----+
No. of Entries: 1
-----+
Lev = route level
IpNh = number of IP next-hops
SrNh = number of SR-tunnel next-hops

```

```
D = duplicate pending
xL = exclude from LFA
LT = LFA type (L:LFA, R:RLFA, T:TILFA, n:nodeProtection)
Act = tunnel active state (R:reported, F:failed, +:SR-ack)
=====
```

The LFA coverage is as follows:

```
[/]
A:admin@PE-7# show router isis sr-lfa-coverage

=====
Rtr Base ISIS Instance 0 SR LFA Coverage
=====
MT-ID  SidType      Level Proto LFA      RLFA      TILFA      Coverage
-----
0      node-sid     L2   ipv4  3(42%)  4(57%)   0(0%)      7/7(100%)
0      adj-sid      L2   ipv4  0(0%)   2(100%)  0(0%)      2/2(100%)
=====
```

## Conclusion

Remote LFA Node Protection provides operators the means to create resilient networks, with precalculated backup paths and with improved coverage.

# RSVP Point-to-Point LSPs

This chapter provides information about point-to-point label switched paths (LSPs) established using resource reservation protocol (RSVP) with or without traffic engineering (TE).

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

## Applicability

This chapter was initially written for SR OS Release 7.0.R5, but the MD-CLI in the current edition is based on SR OS Release 21.2.R1. There are no prerequisites or conditions on the hardware for this configuration.

## Overview

Due to the connectionless nature of the network layer protocol IP, packets travel through the network on a hop-by-hop basis with routing decisions made at each node. As a result, hyperaggregation of data on certain links may occur and it may impact the provider's ability to provide guaranteed service levels across the network end-to-end. To address these shortcomings, Multi-Protocol Label Switching (MPLS) was developed. MPLS provides the capability to establish connection-oriented paths, called Label Switched Paths (LSPs), over a connectionless (IP) network.

The LSP offers a mechanism to engineer network traffic independently from the underlying network routing protocol (mostly IP) to improve the network resiliency and recovery options and to permit delivery of new services that are not readily supported by conventional IP routing techniques, such as Layer 2 IP Virtual Private Networks (VPNs). These benefits are essential for today's communication network explaining the wide deployment base of the MPLS technology.

RFC 3031, *Multiprotocol Label Switching Architecture*, specifies the MPLS architecture whereas this document describes the configuration and troubleshooting of RSVP point-to-point LSPs on SR OS. Besides RSVP P2P LSPs, there are also [Static Point-to-Point LSPs](#), [LDP Point-to-Point LSPs](#), and Segment Routing (SR) LSPs (SR-ISIS, SR-OSPF, and SR-TE). For SR-ISIS, see chapter "Segment Routing with IS-IS Control Plane" in *7750 SR and 7950 XRS Segment Routing and PCE Advanced Configuration Guide for MD CLI*.

## Packet forwarding

As a packet of a connectionless network layer protocol travels from one router to the next, each router in the network makes an independent forwarding decision by performing the following basic tasks: first

analyzing the packet header, then referencing the local routing table to find the longest match based on the destination address in the IP header, and finally sending out the packet on the selected interface. In other words, the first function partitions the entire set of possible packets into a set of Forwarding Equivalence Classes (FECs). All packets associated to a particular FEC will be forwarded along the same logical path to the same destination. The second function maps each FEC to a next hop destination router. Each router along the data path performs these actions.

In MPLS, the assignment of a packet to a particular FEC is done just once, as the packet enters the network. In turn, the FEC is mapped to an LSP, which is pre-signaled prior to any data flowing. An MPLS label, representing the FEC to which the packet is assigned, is attached to the packet (push operation) and once labeled, the packet is forwarded to the next hop router along that LSP path.

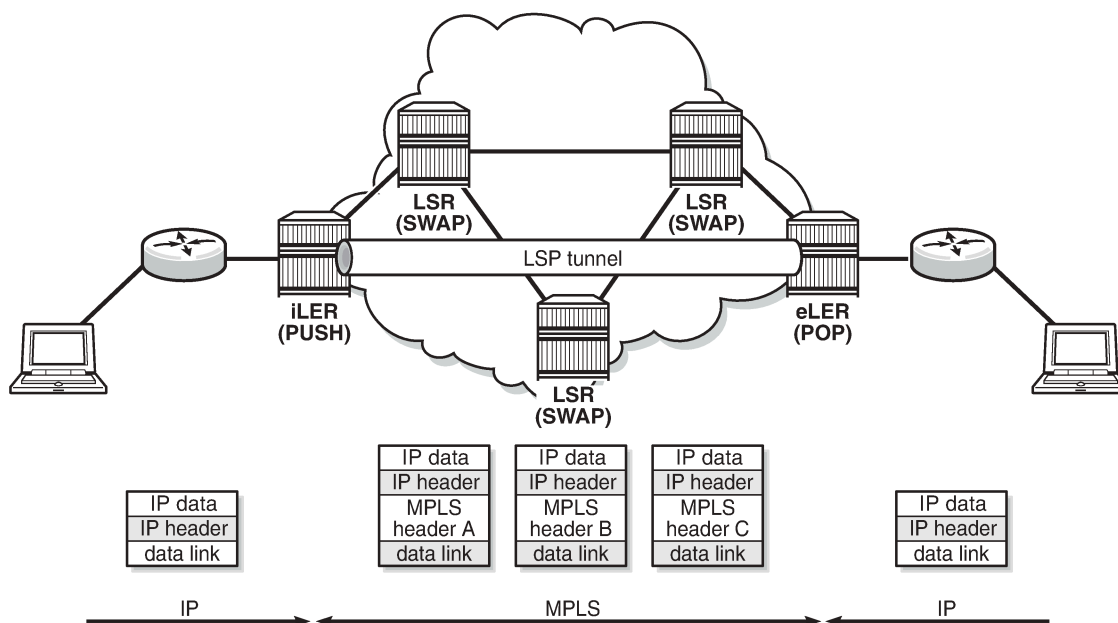
At subsequent hops, there is no further analysis of the packet network layer header. Instead, the label is used as an index into a table which specifies the next hop and a new label. The old label is replaced with the new label (swap operation), and the packet is forwarded to its next hop.

At the MPLS network egress, the label is removed from the packet (pop operation). If this router is the destination (based on the remaining packet), the packet is handed to the receiving application, such as a Virtual Private LAN Service (VPLS). If this router is not the final destination of the packet, the packet will be sent into a new MPLS tunnel or forwarded by conventional IP forwarding toward the Layer 3 destination.

## Terminology

Figure 94: Generic MPLS network, MPLS label operations shows a general network topology clarifying the MPLS-related terms.

Figure 94: Generic MPLS network, MPLS label operations



25762

A Label Edge Router (LER) is a device at the edge of an MPLS network, with at least one interface outside the MPLS domain. A router is usually defined as an LER based on its position relative to a particular LSP.



The MPLS router at the head-end of an LSP is called the ingress Label Edge Router (iLER). The MPLS router at the tail-end of an LSP is called the egress Label Edge Router (eLER).

The iLER receives unlabeled packets from outside the MPLS domain, then applies MPLS labels to the packets, and forwards the labeled packets into the MPLS domain. The eLER receives labeled packets from the MPLS domain, then removes the labels, and forwards unlabeled packets outside the MPLS domain. The eLER can signal an implicit-null label (numeric value 3). This informs the previous hop to send MPLS packets without an outer label and is known as Penultimate Hop Popping (PHP).

A Label Switching Router (LSR) is a device internal to an MPLS network, with all interfaces inside the MPLS domain. These devices switch labeled packets inside the MPLS domain. In the core of the network, LSRs ignore the packet network layer (IP) header and simply forward the packet using the MPLS label swapping mechanism.

A single LSP is unidirectional. In common practice, because the bidirectional nature of most traffic flows is implied, the term LSP often is used to define the pair of LSPs that enable the bidirectional flow. For ease of terminology and discussion however, the LSP in this chapter is referred to as a single entity.

## LSP establishment

Prior to packet forwarding, the LSP must be established. In order to do so, labels need to be distributed for the path. Labels are usually distributed by a downstream router in the upstream direction (relative to the data flow). There are a number of ways used for label distribution: static, LDP, and RSVP. For static P2P LSPs, see chapter [Static Point-to-Point LSPs](#); for LDP P2P LSPs, see chapter [LDP Point-to-Point LSPs](#).

RSVP-TE (RFC 3209, *RSVP-TE: Extensions to RSVP for LSP Tunnels*) can be used to signal LSPs across the network. RSVP-TE is used for traffic engineering when the ingress router creates an LSP with specific constraints beyond the best route chosen by the IGP. RSVP-TE identifies the specific path desired for the LSP and may include resource requirements for the path.

The most important benefit of the label swapping mechanism RSVP-TE is its ability to map any type of user traffic to an LSP that has been specifically engineered to satisfy user traffic requirements. Customized LSPs may be created based on hop count, bandwidth requirements, administrative groups, or Shared Risk Link Groups (SRLGs). They can even be routed through a strict path with specific network links or nodes, as specified by the ingress node. This offers service providers precise control over the flow of traffic in their networks and results in a network that operates more efficiently and provides more predictable and scalable services. For information about SRLG, see chapter [Shared Risk Link Groups for RSVP-Based LSPs](#).

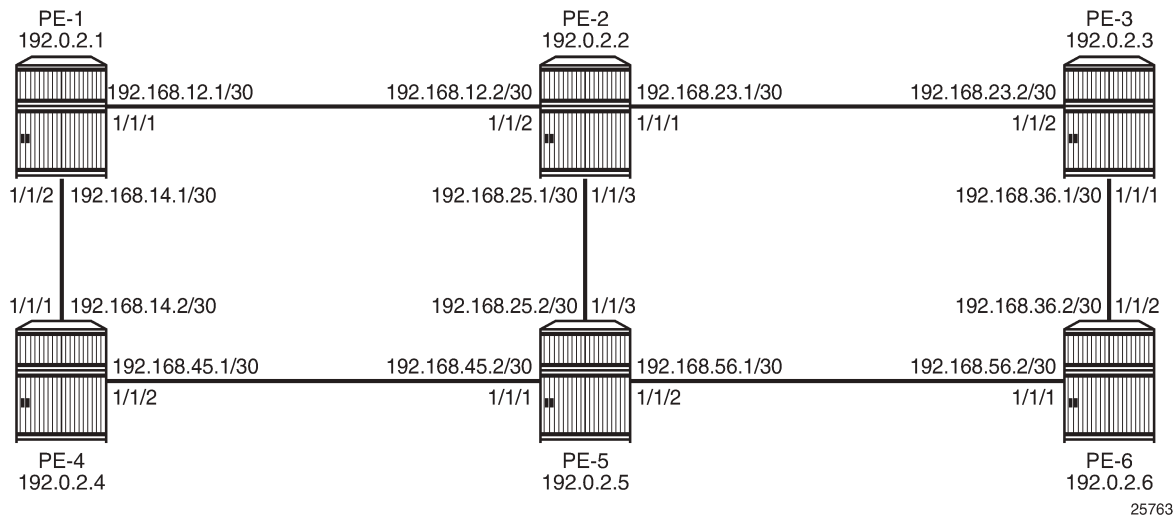
Fast reroute (FRR) allows to signal backup paths before a failure takes place. This allows traffic to flow almost continuously, without waiting for routing protocol convergence. Two different FRR methods exist for RSVP-TE LSPs: one-to-one and facility.

- FRR one-to-one defines detour tunnels toward the eLER for a particular LSP only. The advantage is that the detour tunnel is the best path to the eLER that avoids the node or link at the point of failure. The drawback is that when different LSPs would need the same detour, a dedicated RSVP-TE detour LSP needs to be signaled for each LSP.
- FRR facility defines local repair tunnels avoiding one particular node (the next hop in the data path) or one particular link (the next link in the data path), ignoring the eLER. These bypass tunnels originate in a point of local repair (PLR) and terminate in a merge point (MP) on the LSP. Bypass tunnels are shared between LSPs.

## Example topology

Figure 95: MPLS example topology shows the example topology consisting of six SR OS nodes located in a single autonomous system.

Figure 95: MPLS example topology



## Configuration

In this chapter, RSVP LSPs are configured manually, but they can also be configured automatically using LSP templates; see chapter [Automatic Creation of RSVP-TE LSPs](#).

As a general prerequisite for the configuration of MPLS LSPs, a correctly working Interior Gateway Protocol (IGP) is required. Open Shortest Path First (OSPF) or Intermediate System to Intermediate System (IS-IS) can be used as IGP.

RSVP-TE, an extension of the original RSVP protocol, has two major benefits adding to the basic MPLS functionality. The first benefit is traffic engineering, which allows the ingress router to create an LSP with specific constraints beyond the best route chosen by the IGP. The second benefit is improved network resiliency when a link or node fails in the network, using FRR and secondary paths. FRR is also supported for LDP, see chapter [MPLS LDP FRR using ISIS as IGP](#).

In this chapter, several RSVP-TE LSPs are configured:

- A simple LSP with a primary path that has strict hops and no specific TE constraints
- A simple LSP with a dynamic path without any configured hops is created. Initially, there are no constraints and the actual path is calculated based on the IGP best route.
- An LSP configured with constrained shortest path first (CSPF) that will use the TE metric, even though the IGP metric can also be used
- An LSP with fast reroute (FRR) one-to-one enabled
- An LSP with FRR facility enabled

- An LSP including an admin group "blue" and an LSP excluding admin group "red"
- An LSP with a hop limit configured

There is no configuration example with bandwidth constraints configured in this chapter. See chapter "Automatic Bandwidth Adjustment" in P2P LSPs in the *7450 ESS, 7750 SR, 7950 XRS, and VSR MPLS Advanced Configuration Guide for Classic CLI* for a configuration with bandwidth constraint with or without automatic adjustment.

Initially, no traffic engineering is enabled in the ISIS context, but it will be enabled when required. For RSVP LSPs, MPLS and RSVP need to be enabled on each router and all network interfaces facing the MPLS domain. For PE-1, the following configuration is required:

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      admin-state enable
      interface "int-PE-1-PE-2" {
      }
      interface "int-PE-1-PE-4" {
      }
    }
    rsvp {
      admin-state enable
      interface "int-PE-1-PE-2" {
      }
      interface "int-PE-1-PE-4" {
      }
    }
  }
}
```

## Strict or loose path

On the iLER, first the definition of a path is required. A path is a sequence of MPLS routers (hops) through which the LSP using that path has to pass. It is not uniquely bound to a particular LSP; it can be used by any LSP originating in that node. A hop in a path can be strict or loose: strict or loose meaning that the LSP must take either a direct path from the previous hop router to this router (strict) or can traverse through other routers (loose). The hops not explicitly defined in the loose path definition are created by calculating the IGP shortest path. A third possibility is an empty path implying not a single node is required to be present in the LSP path and the shortest path from the IGP is used to define the LSP path. Other techniques, such as the use of admin groups or shared risk link groups, can also be used to influence the decision which hops to include in the path. Three paths will be configured, respectively:

1. Only strict hops
2. Mixed strict and loose hops
3. Empty path

To find a valid path, the last hop in the path sequence needs to be the system IP or an interface address of the terminating router (eLER). The IP addresses in the hop command can be the system IP addresses or the interface addresses of the node. However, it is recommended to use the system IP addresses with keyword **loose** because this allows more flexibility when finding new paths in failover scenarios (because the upstream node could use any of multiple paths to the system address, whereas specifying the interface address would restrict the upstream node to a single entry-point). The recommendation when using the keyword **strict** in the **hop** command context, is to use the physical link addresses.



**Note:**

The last hop in the path should be a system address to make it appear in the list on the 5620 SAM (service-aware manager).

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      path "dyn" {
        admin-state enable
      }
      path "path-PE-1-PE-6-semiLoose" {
        admin-state enable
        hop 10 {
          ip-address 192.0.2.5
          type loose
        }
        hop 20 {
          ip-address 192.168.56.2
          type strict
        }
      }
      path "path-PE-1-PE-6-strict" {
        admin-state enable
        hop 10 {
          ip-address 192.168.12.2
          type strict
        }
        hop 20 {
          ip-address 192.168.25.2
          type strict
        }
        hop 30 {
          ip-address 192.168.56.2
          type strict
        }
      }
    }
  }
}
```

The paths can be checked with the **show router mpls path** command.

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

=====
MPLS Path:
=====
```

Path Name	Admin	PathIdx
Hop Index	IP Address/SID-Label	Strict/Loose
dyn	Up	1
no hops	n/a	
path-PE-1-PE-6-semiLoose	Up	2
10	192.0.2.5	Loose
20	192.168.56.2	Strict
path-PE-1-PE-6-strict	Up	3
10	192.168.12.2	Strict
20	192.168.25.2	Strict
30	192.168.56.2	Strict

```
-----
```

```
Total Paths : 3
=====
```

## Simple RSVP LSP with strict primary path

The configuration of a simple LSP using RSVP signaling contains at least on the iLER:

- System IP address of the terminating node (to)
- Path to the eLER (primary)
- Administratively enabled

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-6" {
        admin-state enable
        type p2p-rsvp
        to 192.0.2.6
        primary "path-PE-1-PE-6-strict" {
        }
        secondary "dyn" {
        }
      }
    }
  }
}
```

All the hops in the strict path are already defined and there is no need to look up the IGP best route. The configuration of secondary paths is optional. In case the primary path fails, the secondary path can be signaled to take over the traffic. It can even be signaled as standby while the primary path is operational for a faster switchover when the keyword `standby` is added, which is not the case here. The secondary path has no hops defined. The hops will be calculated based on the IGP best route. The nodes through which the LSP will pass (LSRs and eLER) require no additional configuration: enabling MPLS and RSVP on their interfaces suffices.

An overview of all LSPs configured on a particular node is given by the **show router mpls lsp** command. More details about a particular LSP can be retrieved by adding the keyword `detail` to the previous command.

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

```
=====
MPLS LSPs (Originating)
=====
```

LSP Name To	Tun Id	Fastfail Config	Adm	Opr
LSP-PE-1-PE-6 192.0.2.6	1	No	Up	Up

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

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-6" detail
```

```

MPLS LSPs (Originating) (Detail)
=====
Legend :
+ - Inherited
=====
-----
Type : Originating
-----
LSP Name      : LSP-PE-1-PE-6
LSP Type      : RegularLsp           LSP Tunnel ID      : 1
LSP Index     : 1                   TTM Tunnel Id      : 1
From          : 192.0.2.1
To            : 192.0.2.6
Adm State     : Up                   Oper State          : Up
LSP Up Time   : 0d 00:00:37         LSP Down Time      : 0d 00:00:00
Transitions   : 1                   Path Changes        : 1
Retry Limit   : 0                   Retry Timer         : 30 sec
Signaling     : RSVP                Resv. Style         : SE
Hop Limit     : 255                 Negotiated MTU      : 1564
Adaptive      : Enabled              ClassType           : 0
FastReroute   : Disabled            Oper FR             : Disabled
PathCompMethod : none               ADSPEC              : Disabled
FallbkPathComp : not-applicable
Metric        : N/A
Load Bal Wt   : N/A
Include Grps  :                      ClassForwarding    : Disabled
None          :                      Exclude Grps        :
None          :                      None
Least Fill    : Disabled
BFDD Template : None                BFD Ping Intvl     : 60
BFDD Enable   : False               BFD Failure-action  : None
WaitForUpTimer : 4

Revert Timer  : Disabled             Next Revert In     : N/A
Entropy Label : Enabled+             Oper Entropy Label : Enabled
Negotiated EL : Disabled
Auto BW       : Disabled
LdpOverRsvp  : Enabled
VprnAutoBind : Enabled
IGP Shortcut  : Enabled             BGP Shortcut       : Enabled
IGP LFA       : Disabled            IGP Rel Metric     : Disabled
BGPTransTun  : Enabled
Oper Metric   : 16777215
Prop Adm Grp  : Disabled
PCE Report    : Disabled+
PCE Control   : Disabled
Path Profile  : None
Admin Tags    : None
Lsp Self Ping : Disabled+             Self Ping Timeouts : 0
SelfPingOAMFail* : 0

Secondary     : dyn
Down Time     : 0d 00:00:37
Bandwidth     : 0 Mbps
Primary(a)    : path-PE-1-PE-6-strict
Up Time       : 0d 00:00:37
Bandwidth     : 0 Mbps
=====
* indicates that the corresponding row element may have been truncated.

```

In each hop (originating, transit and terminate), the RSVP sessions can be verified as follows:

```

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

```

```

=====
RSVP Sessions
=====
RSVP Session Name
  From           To           Tunnel ID   LSP ID      State
-----
LSP-PE-1-PE-6::path-PE-1-PE-6-strict
192.0.2.1       192.0.2.6    1           13824       Up
-----
Sessions : 1
=====

```

The detailed output of this command includes among others the session type (here: originate), the incoming and outgoing labels, the previous and next hop, and - for originating LSPs - also the list of hops):

```

[/]
A:admin@PE-1# show router rsvp session detail
=====
RSVP Sessions (Detailed)
=====
LSP : LSP-PE-1-PE-6::path-PE-1-PE-6-strict
-----
From           : 192.0.2.1           To           : 192.0.2.6
Tunnel ID      : 1                       LSP ID       : 13824
Style          : SE                       State        : Up
Session Type   : Originate
In Interface   : n/a                   Out Interface : 1/1/1
In IF Name     : n/a
Out IF Name    : int-PE-1-PE-2
In Label       : n/a                   Out Label    : 524287
Previous Hop   : n/a                   Next Hop     : 192.168.12.2
Hops           :
                : 192.168.12.2(S)
                : -> 192.168.25.2(S)
                : -> 192.168.56.2(S)
SetupPriority  : 7                       Hold Priority : 0
Class Type    : 0
SubGrpOrig ID : 0                       SubGrpOrig Addr:
P2MP ID       : 0
FrrAvailType  : N/A
FrrSrlgStrict : N/A                       SrlgDisjoint : N/A

Path Recd     : 0                       Path Sent    : 4
Resv Recd     : 3                       Resv Sent    : 0
Summary msgs  :
SPath Recd   : 0                       SPath Sent   : 0
SResv Recd   : 0                       SResv Sent   : 0
LSP Attr Flags : N/A
=====

```

The following RSVP LSP is in the tunnel table on PE-1:

```

[/]
A:admin@PE-1# show router tunnel-table
=====
IPv4 Tunnel Table (Router: Base)
=====

```

Destination Color	Owner	Encap	TunnelId	Pref	Nexthop	Metric
192.0.2.6/32	rsvp	MPLS	1	7	192.168.12.2	16777215

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

In order to signal PHP with RSVP, implicit-null must be configured on the eLER.

```
# on PE-6:
configure {
  router "Base" {
    rsvp {
      implicit-null-label true
    }
  }
}
```

The implicit-null is signaled and is shown on PE-5 as an egress label of 3. This label is not actually sent toward PE-6.

```
[/]
A:admin@PE-5# show router rsvp session detail

=====
RSVP Sessions (Detailed)
=====
-----
LSP : LSP-PE-1-PE-6::path-PE-1-PE-6-strict
-----
From           : 192.0.2.1           To           : 192.0.2.6
Tunnel ID      : 1               LSP ID       : 13828
Style          : SE              State         : Up
Session Type   : Transit
In Interface   : 1/1/3            Out Interface : 1/1/2
In IF Name     : int-PE-5-PE-2
Out IF Name    : int-PE-5-PE-6
In Label       : 524287          Out Label     : 3
Previous Hop   : 192.168.25.1    Next Hop      : 192.168.56.2
---snip---
```

The use of implicit-null can also be enabled/disabled on a per interface basis:

```
# on PE-6:
configure {
  router "Base" {
    rsvp {
      interface "int-PE-6-PE-5"
        implicit-null-label true
      }
    }
}
```

In the remainder of the chapter, LSPs with empty paths will be used. LSP "LSP-PE-1-PE-6" is administratively disabled.



## Simple RSVP LSP with dynamic path

In this section, an LSP is configured from PE-1 to PE-3 with a dynamic path that is empty. There is no secondary path. LSP "LSP-PE-1-PE-3" is configured as follows:

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-3" {
        admin-state enable
        type p2p-rsvp
        to 192.0.2.3
        primary "dyn" {
        }
      }
    }
  }
}
```

Interfaces with a lower metric will be preferred over links with a high metric. The default IGP metric in this example is 10. The metric is lower for higher speed links, but can be configured manually; as follows:

```
# on PE-1:
configure {
  router "Base" {
    isis 0 {
      interface "int-PE-1-PE-2" {
        level 1 {
          metric 1000
        }
      }
    }
  }
}
```

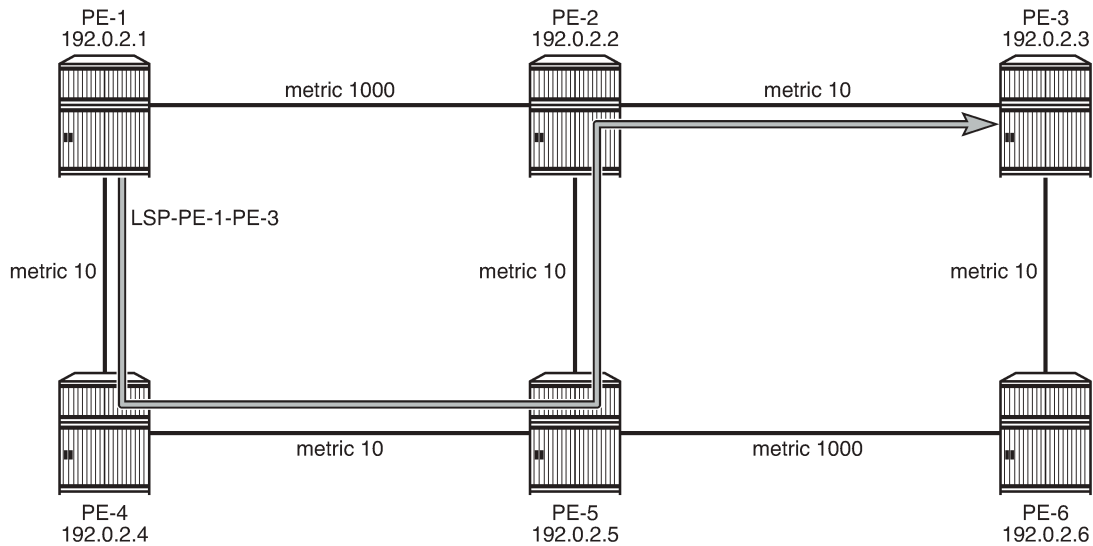
The link between PE-1 and PE-2 has a higher metric and will not be selected for forwarding traffic because the route via PE-4 has a lower metric. The routing table shows that the route to prefix 192.0.2.3 has PE-4 as next hop instead of PE-2 and that the metric is 40:

```
[/]
A:admin@PE-1# show router route-table 192.0.2.3

=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]                Type   Proto   Age           Pref
Next Hop[Interface Name]          Metric
-----
192.0.2.3/32                       Remote ISIS   00h02m45s  15
192.168.14.2                       40
-----
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
=====
```

Figure 96: LSP with dynamic path takes IGP best route shows the path used by the LSP:

Figure 96: LSP with dynamic path takes IGP best route



25765

The actual hops can be verified in the following output. The path is dynamic, therefore, no explicit hops are configured.

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-3" path detail

=====
MPLS LSP LSP-PE-1-PE-3 Path (Detail)
=====
Legend :
@ - Detour Available          # - Detour In Use
b - Bandwidth Protected      n - Node Protected
s - Soft Preemption          L - Loose
S - Strict                   + - Inherited
A - ABR

-----
LSP LSP-PE-1-PE-3
Path dyn
-----
LSP Name      : LSP-PE-1-PE-3
From          : 192.0.2.1
To            : 192.0.2.3
Admin State   : Up
Path Name     : dyn
Path LSP ID   : 45056
Path Admin    : Up
Out Interface : 1/1/2
Oper State    : Up
Path Type     : Primary
Path Oper     : Up
Out Label     : 524287
---snip---

Explicit Hops :
No Hops Specified
Actual Hops   :
192.168.14.1(192.0.2.1)      Record Label : N/A
-> 192.168.14.2(192.0.2.4)   Record Label : 524287
-> 192.168.45.2(192.0.2.5)   Record Label : 524287
-> 192.168.25.1(192.0.2.2)   Record Label : 524287
```

```
-> 192.168.23.2(192.0.2.3)          Record Label      : 524287
Resignal Eligible: False
Last Resignal      : n/a          CSPF Metric       : 0
=====
```

The tunnel table shows the RSVP LSP with PE-4 as the next hop and a metric of 40:

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

=====
IPv4 Tunnel Table (Router: Base)
=====
Destination          Owner      Encap TunnelId  Pref  Nexthop      Metric
  Color
-----
192.0.2.3/32         rsvp      MPLS  2          7    192.168.14.2  40
-----
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
=====
```

## RSVP-TE LSP with dynamic path

Traffic engineering is enabled in the ISIS context on all nodes; as follows:

```
# on all PEs:
configure {
  router "Base" {
    isis 0 {
      traffic-engineering
    }
  }
}
```

The LSP can be configured with constrained shortest path first (CSPF); as follows:

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-3" {
        path-computation-method local-cspf
      }
    }
  }
}
```

For this LSP, it will not make any difference. By default, the IGP metrics are used and the LSP path takes the IGP shortest path.

Besides IGP metrics, also TE metrics can be configured; as follows:

```
# on PE-2:
configure {
  router "Base" {
    mpls {
      interface "int-PE-2-PE-1" {
        te-metric 10
      }
      interface "int-PE-2-PE-3" {
        te-metric 500
      }
      interface "int-PE-2-PE-5" {

```

```
        te-metric 10
    }
```

In this example, all interfaces on all PEs get a TE metric of 10, except for the interfaces between PE-1 and PE-4, which get a TE metric of 100 and the interfaces between PE-2 and PE-3, which get a TE metric of 500. Even with these TE metrics configured, the LSP path will not change, because the IGP metric is used by default, as can be verified as follows:

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-3" path detail

=====
MPLS LSP LSP-PE-1-PE-3 Path (Detail)
=====
Legend :
@ - Detour Available          # - Detour In Use
b - Bandwidth Protected      n - Node Protected
s - Soft Preemption
S - Strict                    L - Loose
A - ABR                       + - Inherited
=====
-----
LSP LSP-PE-1-PE-3
Path dyn
-----
LSP Name      : LSP-PE-1-PE-3
From          : 192.0.2.1
To            : 192.0.2.3
Admin State   : Up
Path Name     : dyn
Path LSP ID   : 45058
Path Admin    : Up
Out Interface : 1/1/2
Out Label     : 524286
---snip---

Adaptive      : Enabled
Preference    : n/a
Path Trans    : 2
Failure Code  : noError
Failure Node  : n/a
Explicit Hops :
  No Hops Specified
Actual Hops   :
  192.168.14.1(192.0.2.1)
  -> 192.168.14.2(192.0.2.4)
  -> 192.168.45.2(192.0.2.5)
  -> 192.168.25.1(192.0.2.2)
  -> 192.168.23.2(192.0.2.3)
Computed Hops :
  192.168.14.1(S)
  -> 192.168.14.2(S)
  -> 192.168.45.2(S)
  -> 192.168.25.1(S)
  -> 192.168.23.2(S)
Resignal Eligible: False
Last Resignal : n/a
CSPF Metric   : 40
---snip---
```

The RSVP LSP in the tunnel table has next hop PE-4 and a metric of 40:

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

```

=====
IPv4 Tunnel Table (Router: Base)
=====
Destination          Owner      Encap TunnelId  Pref  Nexthop      Metric
  Color
-----
192.0.2.3/32        rsvp      MPLS    2       7    192.168.14.2  40
-----
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
=====
    
```

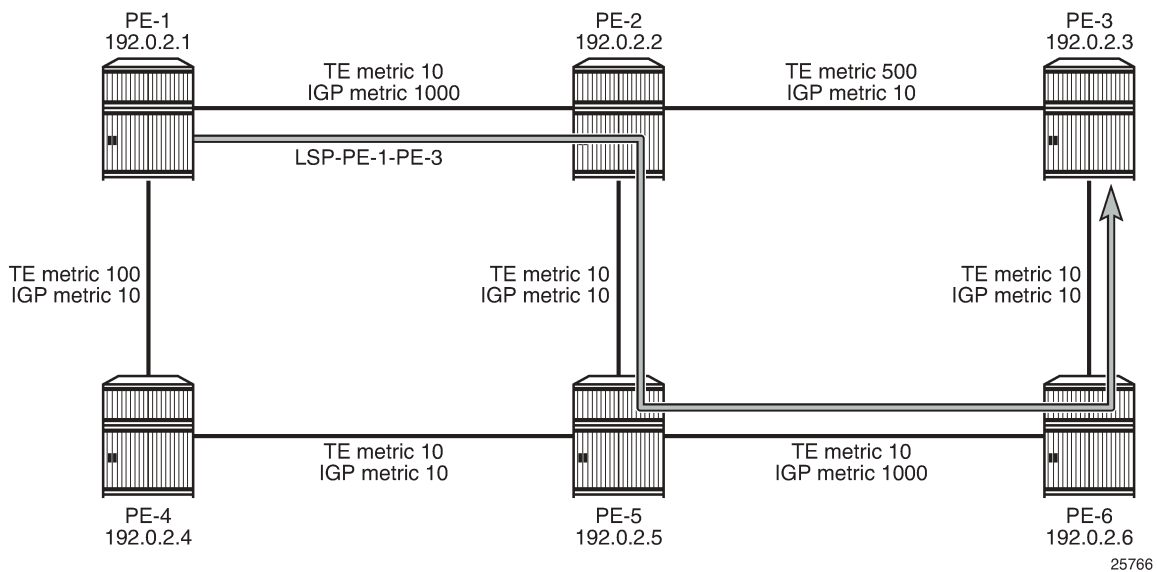
To force the LSP to use the TE metric, the LSP is reconfigured as follows:

```

# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-3" {
        path-computation-method local-cspf
        metric-type te
      }
    }
  }
}
    
```

The LSP path is shown in [Figure 97: RSVP-TE LSP with dynamic path using TE metric](#):

Figure 97: RSVP-TE LSP with dynamic path using TE metric



The LSP path goes from PE-1 to PE-2 and via PE-5 and PE-6 to PE-3, as can be seen in the following output. The CSPF metric is 40, which corresponds to the TE metric in this case:

```

[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-3" path detail
    
```

```

=====
MPLS LSP LSP-PE-1-PE-3 Path (Detail)
=====
    
```

```

Legend :
  @ - Detour Available          # - Detour In Use
  b - Bandwidth Protected      n - Node Protected
  s - Soft Preemption
  S - Strict                    L - Loose
  A - ABR                      + - Inherited
=====
-----
LSP LSP-PE-1-PE-3
Path dyn
-----
LSP Name      : LSP-PE-1-PE-3
From          : 192.0.2.1
To            : 192.0.2.3
Admin State   : Up                Oper State    : Up
Path Name     : dyn
Path LSP ID   : 45060             Path Type     : Primary
Path Admin    : Up                Path Oper     : Up
Out Interface : 1/1/1             Out Label     : 524287
---snip---

Adaptive      : Enabled           Oper Metric    : 16777215
Preference    : n/a
Path Trans    : 4                 CSPF Queries   : 2
Failure Code  : noError
Failure Node  : n/a
Explicit Hops :
  No Hops Specified
Actual Hops   :
  192.168.12.1(192.0.2.1)         Record Label   : N/A
  -> 192.168.12.2(192.0.2.2)     Record Label   : 524287
  -> 192.168.25.2(192.0.2.5)     Record Label   : 524287
  -> 192.168.56.2(192.0.2.6)     Record Label   : 524287
  -> 192.168.36.1(192.0.2.3)     Record Label   : 524287
Computed Hops :
  192.168.12.1(S)
  -> 192.168.12.2(S)
  -> 192.168.25.2(S)
  -> 192.168.56.2(S)
  -> 192.168.36.1(S)
Resignal Eligible: False
Last Resignal  : n/a              CSPF Metric    : 40
=====

```

The tunnel table shows the RSVP LSP with next hop PE-2 and a metric of 16777215 (infinity) because the IGP metric is not used:

```

[/]
A:admin@PE-1# show router tunnel-table
=====
IPv4 Tunnel Table (Router: Base)
=====
Destination      Owner      Encap TunnelId  Pref  Nexthop      Metric
  Color
-----
192.0.2.3/32     rsvp      MPLS  2        7        192.168.12.2  16777215
-----
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 IGP metric values are restored to their default value on all interfaces on all PEs. The TE metric is configured with a value of 10 on all interfaces on all PEs.

On PE-1, the configuration is as follows:

```
# on PE-1:
configure {
  router "Base" {
    isis 0 {
      interface "int-PE-1-PE-2" {
        level 1 {
          delete metric
        }
      }
    }
  }
  mpls {
    interface "int-PE-1-PE-4" {
      te-metric 10
    }
  }
}
```

When all metrics have the same value, it does not matter whether CSPF uses the IGP or TE metric. CSPF will use the IGP metric after the following command is executed.

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-3" {
        path-computation-method local-cspf
        delete metric-type # default metric-type = igp
      }
    }
  }
}
```

The primary path will go from PE-1 to PE-2 and then to PE-3 with a CSPF (IGP) metric of 20.

## Fast reroute for RSVP-TE LSPs

It is mandatory to have CSPF enabled for FRR.

Fast reroute can be configured on the RSVP LSP in two ways:

1. One-to-one: for each potential point of failure, the best detour tunnel to the eLER is signaled. This detour tunnel is signaled for this particular LSP only and cannot be shared among LSPs
2. Facility: local bypass tunnels are signaled from each point of local repair avoiding the next link or the next node. The bypass tunnels can be shared among LSPs.

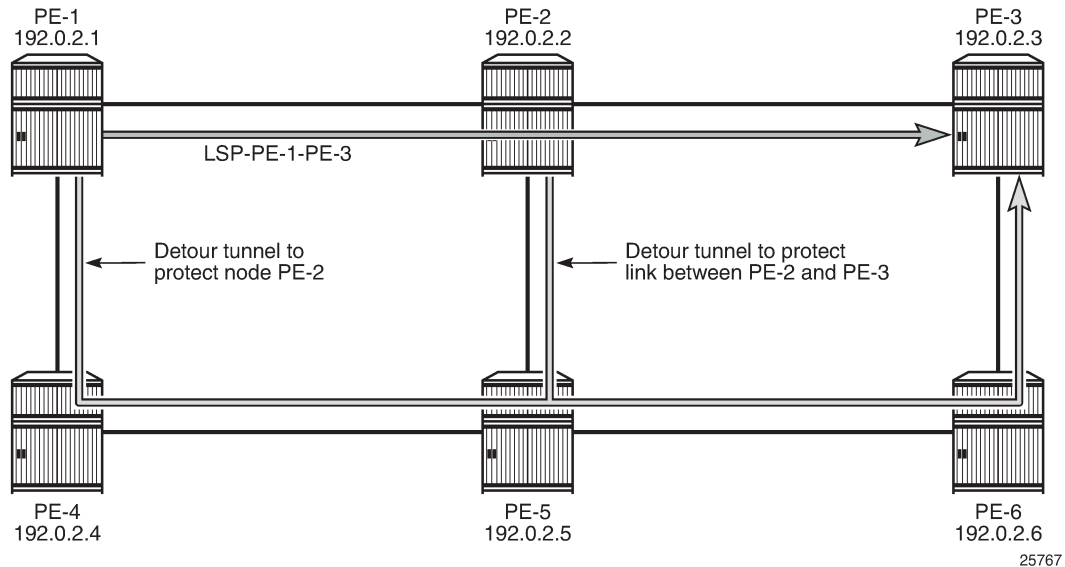
### FRR one-to-one

The LSP "LSP-PE-1-PE-3" is configured with FRR one-to-one; as follows:

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-3" {
        fast-reroute {
          frr-method one-to-one
        }
      }
    }
  }
}
```

The preferred path from PE-1 to PE-3 is via PE-2. There will be two detour tunnels: one originating in PE-1 to protect node PE-2, and a second detour tunnel originating in PE-2 to protect the link between PE-2 and PE-3. Both detour tunnels use the same path from PE-5 to PE-3 and there is no need to signal this path twice. One detour tunnel terminates in PE-5, and the diverted traffic in this tunnel will be sent to PE-6 and PE-3 via the established detour tunnel. Depending on which detour tunnel is established first, the other detour tunnel terminates in PE-5. The preferred tunnel and the detour tunnels are shown in [Figure 98: Fast reroute one-to-one detour tunnels](#):

Figure 98: Fast reroute one-to-one detour tunnels



The protection can be seen in the list of actual hops in the path. In PE-1, a detour tunnel for node protection originates (indicated by @ n; see legend) and in PE-2 a detour tunnel for link protection (indicated by @):

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-3" path detail

=====
MPLS LSP LSP-PE-1-PE-3 Path (Detail)
=====
Legend :
  @ - Detour Available          # - Detour In Use
  b - Bandwidth Protected      n - Node Protected
  s - Soft Preemption
  S - Strict                    L - Loose
  A - ABR                       + - Inherited

-----
LSP LSP-PE-1-PE-3
Path dyn
-----
LSP Name      : LSP-PE-1-PE-3
From          : 192.0.2.1
To            : 192.0.2.3
Admin State   : Up              Oper State    : Up
Path Name     : dyn
Path LSP ID   : 45064          Path Type     : Primary
```



```

Path Admin      : Up                Path Oper      : Up
Out Interface   : 1/1/1            Out Label      : 524287
---snip---

FRR             : Enabled          Oper FRR       : Enabled
FRR NodeProtect : Enabled          Oper FRR NP    : Enabled
FR Hop Limit    : 16              Oper FRHopLimit : 16
---snip---

Actual Hops     :
  192.168.12.1(192.0.2.1) @ n      Record Label   : N/A
-> 192.168.12.2(192.0.2.2) @      Record Label   : 524287
-> 192.168.23.2(192.0.2.3) @      Record Label   : 524287
---snip---

Detour Status   : Standby          Detour Type    : Originate
Detour Avoid Nod* : 192.0.2.2      Detour Origin  : 192.0.2.1
Setup Priority   : 7               Hold Priority   : 0
Class Type      : 0
Detour Active Ti* : n/a           Detour Up Time : 0d 00:02:55
In Interface     : n/a           In Label       : n/a
Out Interface    : 1/1/2         Out Label      : 524287
NextHop         : 192.168.14.2
Explicit Hops   :
  192.168.14.1(S)
-> 192.168.14.2(S)
-> 192.168.45.2(S)
-> 192.168.56.2(S)
-> 192.168.36.1(S)
=====
* indicates that the corresponding row element may have been truncated.

```

The output also contains information about the detour tunnel originating in PE-1 that protects node PE-2. Because the detour tunnel is dedicated for this LSP, that information can be included in the LSP information.

The RSVP detour sessions can be retrieved in the originating, transit, and terminating nodes. On originating node PE-1:

```

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

=====
RSVP Sessions
=====
RSVP Session Name
From           To           Tunnel ID   LSP ID      State
-----
LSP-PE-1-PE-3::dyn_detour
192.0.2.1     192.0.2.3   2           45064       Up
-----
Sessions : 1
=====

```

In the transit/terminating node PE-5:

```

[/]
A:admin@PE-5# show router rsvp session detour-transit

=====
RSVP Sessions
=====

```

```

=====
RSVP Session Name
  From           To           Tunnel ID   LSP ID     State
-----
LSP-PE-1-PE-3::dyn_detour
192.0.2.1       192.0.2.3     2           45064      Up
-----
Sessions : 1
=====

```

```

[/]
A:admin@PE-5# show router rsvp session detour-terminate

```

```

=====
RSVP Sessions
=====
RSVP Session Name
  From           To           Tunnel ID   LSP ID     State
-----
LSP-PE-1-PE-3::dyn_detour
192.0.2.1       192.0.2.3     2           45064      Up
-----
Sessions : 1
=====

```

More detailed information can be retrieved as follows:

```

[/]
A:admin@PE-5# show router rsvp session detail
=====
RSVP Sessions (Detailed)
=====
LSP : LSP-PE-1-PE-3::dyn_detour
-----
From           : 192.0.2.1           To           : 192.0.2.3
Tunnel ID      : 2                     LSP ID       : 45064
Style          : SE                   State         : Up
Session Type : Transit (Detour)
In Interface   : 1/1/1           Out Interface : 1/1/2
In IF Name     : int-PE-5-PE-4
Out IF Name    : int-PE-5-PE-6
In Label       : 524287           Out Label     : 524287
Previous Hop : 192.168.45.1   Next Hop    : 192.168.56.2
---snip---

LSP : LSP-PE-1-PE-3::dyn_detour
-----
From           : 192.0.2.1           To           : 192.0.2.3
Tunnel ID      : 2                     LSP ID       : 45064
Style          : SE                   State         : Up
Session Type : Terminate (Detour)
In Interface   : 1/1/3           Out Interface : 1/1/2
In IF Name     : int-PE-5-PE-2
Out IF Name    : int-PE-5-PE-6
In Label       : 524286           Out Label     : 524287
Previous Hop : 192.168.25.1   Next Hop    : 192.168.56.2
---snip---

```

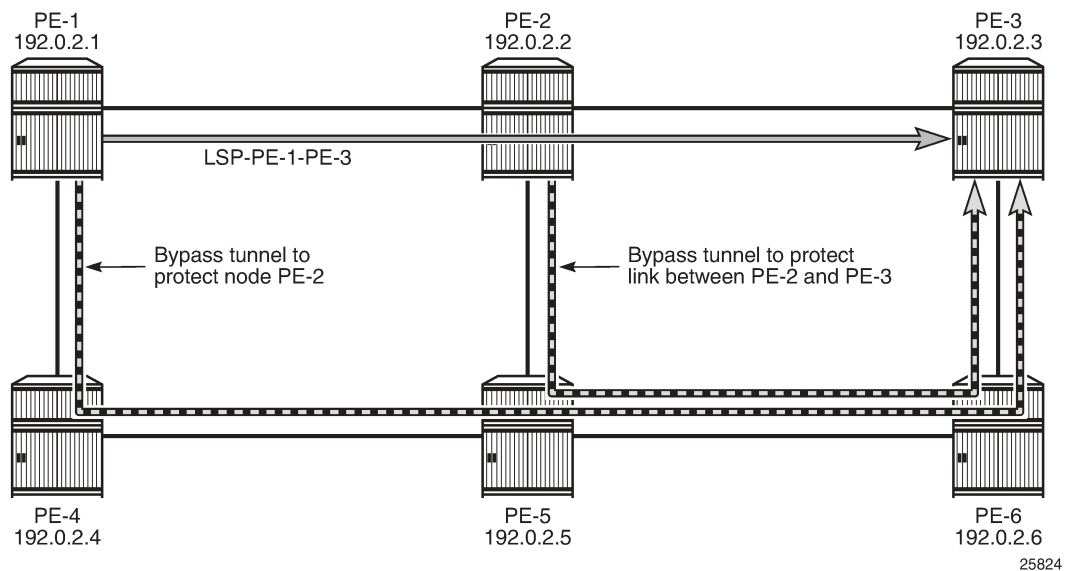
PE-5 is a transit node for the detour tunnel with previous hop PE-4 and a terminating node for the detour tunnel with previous hop PE-2. In both cases, the next hop is PE-6.

## FRR facility

The drawback of FRR one-to-one is that each LSP requires its own detour tunnels to be signaled. FRR facility does not have this issue, because it offers local repair for the next node or the next link that uses bypass tunnels that can be shared by LSPs. FRR facility bypass tunnels terminate in the merge point (MP), which is a hop in the primary path. FRR facility bypass tunnels for link protection terminate in the next hop in the primary path and FRR facility bypass tunnels for node protection terminate in the next hop of that next hop. FRR bypass tunnels are unaware of the final destination of the LSP and need not terminate in the final destination, but in this case they do, because the number of hops in the primary path is limited.

[Figure 99: FRR facility bypass tunnels](#) shows the FRR facility bypass tunnels for LSP "LSP-PE-1-PE-3":

Figure 99: FRR facility bypass tunnels



Fast reroute facility is enabled on LSP "LSP-PE-1-PE-3" as follows:

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-3" {
        admin-state enable
        type p2p-rsvp
        to 192.0.2.3
        path-computation-method local-cspf
        fast-reroute {
          frr-method facility
        }
        primary "dyn" {
        }
      }
    }
  }
}
```

The LSP path detail output shows that there is a bypass tunnel available in PE-1 that offers node protection for the next node in the primary path: PE-2. In PE-2, there is a bypass tunnel offering link protection for the next link, which is the link between PE-2 and PE-3; as follows:

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-3" path detail

=====
MPLS LSP LSP-PE-1-PE-3 Path (Detail)
=====
Legend :
  @ - Detour Available          # - Detour In Use
  b - Bandwidth Protected      n - Node Protected
  s - Soft Preemption
  S - Strict                    L - Loose
  A - ABR                      + - Inherited
=====

-----
LSP LSP-PE-1-PE-3
Path dyn
-----
LSP Name      : LSP-PE-1-PE-3
From          : 192.0.2.1
To            : 192.0.2.3
Admin State   : Up                Oper State    : Up
Path Name     : dyn
Path LSP ID   : 45064             Path Type     : Primary
Path Admin    : Up                Path Oper     : Up
Out Interface : 1/1/1             Out Label     : 524287
---snip---

FRR           : Enabled           Oper FRR      : Enabled
FRR NodeProtect : Enabled         Oper FRR NP   : Enabled
FR Hop Limit  : 16                Oper FRHopLimit : 16
---snip---

Actual Hops   :
  192.168.12.1(192.0.2.1) @ n      Record Label  : N/A
-> 192.168.12.2(192.0.2.2) @      Record Label  : 524287
-> 192.168.23.2(192.0.2.3)        Record Label  : 524287
---snip---

FRR           : Enabled           Oper FRR      : Enabled
FRR NodeProtect : Enabled         Oper FRR NP   : Enabled
FR Hop Limit  : 16                Oper FRHopLimit : 16
---snip---
```

In FRR facility mode, the bypass tunnels are shared. They are not included in the LSP information. The bypass tunnels can be retrieved as follows:

```
[/]
A:admin@PE-1# show router mpls bypass-tunnel protected-lsp detail

=====
MPLS Bypass Tunnels (Detail)
=====
-----
bypass-node192.0.2.2-61441
-----
To          : 192.168.36.1        State         : Up
Out I/F     : 1/1/2              Out Label     : 524287
Up Time     : 0d 00:02:54        Active Time    : n/a
```

```

Reserved BW      : 0 Kbps          Protected LSP Count : 1
Type             : Dynamic         Bypass Path Cost    : 40
Setup Priority   : 7               Hold Priority        : 0
Class Type      : 0
Exclude Node    : None            Inter-Area           : False
Computed Hops   :
  192.168.14.1(S)                Egress Admin Groups : None
-> 192.168.14.2(S)                Egress Admin Groups : None
-> 192.168.45.2(S)                Egress Admin Groups : None
-> 192.168.56.2(S)                Egress Admin Groups : None
-> 192.168.36.1(S)                Egress Admin Groups : None
Actual Hops     :
  192.168.14.1(192.0.2.1)         Record Label         : N/A
-> 192.168.14.2(192.0.2.4)         Record Label         : 524287
-> 192.168.45.2(192.0.2.5)         Record Label         : 524286
-> 192.168.56.2(192.0.2.6)         Record Label         : 524286
-> 192.168.36.1(192.0.2.3)         Record Label         : 524285
Last Resignal   :
Attempted At    : n/a             Resignal Reason      : n/a
Resignal Status: n/a             Reason               : n/a

Protected LSPs -
LSP Name        : LSP-PE-1-PE-3::dyn
From            : 192.0.2.1       To                    : 192.0.2.3
Avoid Node/Hop  : 192.0.2.2       Downstream Label     : 524287
Bandwidth       : 0 Kbps
=====

```

This is the bypass tunnel that originates in PE-1 to protect (avoid) PE-2. In this example, there is only one LSP protected by this bypass tunnel, but the list of protected LSPs can be longer. The same command can be launched on PE-2, where a bypass tunnel originates that protects the link between PE-2 and PE-3.

The RSVP sessions can be displayed as follows:

```

[/]
A:admin@PE-3# show router rsvp session

=====
RSVP Sessions
=====
RSVP Session Name
From          To          Tunnel ID  LSP ID    State
-----
LSP-PE-1-PE-3::dyn
192.0.2.1    192.0.2.3    2          45066     Up

bypass-link192.168.23.2-61441
192.0.2.2    192.168.36.1 61441     2          Up

bypass-node192.0.2.2-61441
192.0.2.1    192.168.36.1 61441     2          Up

-----
Sessions : 3
=====

```

In PE-3, there is an RSVP session for the regular LSP and two bypass tunnels. In this case, the bypass tunnels all go to PE-3, which is the terminating node for the LSP, but that need not be the case. All bypass tunnels are signaled from the point of local repair to the merge point on the LSP path.

To force a FRR facility switchover to a bypass tunnel, a failure is simulated by disabling port 1/1/1 on PE-2, as follows:

```
# on PE-2:
configure {
  port 1/1/1 {
    admin-state disable
```

The detailed output for the LSP path on PE-1 shows that the tunnel is locally repaired.

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-3" path detail

=====
MPLS LSP LSP-PE-1-PE-3 Path (Detail)
=====
Legend :
  @ - Detour Available          # - Detour In Use
  b - Bandwidth Protected      n - Node Protected
  s - Soft Preemption
  S - Strict                    L - Loose
  A - ABR                       + - Inherited
=====

LSP LSP-PE-1-PE-3
Path dyn
-----
LSP Name      : LSP-PE-1-PE-3
From          : 192.0.2.1
To           : 192.0.2.3
Admin State   : Up                Oper State    : Up
Path Name    : dyn
Path LSP ID  : 45066              Path Type     : Primary
Path Admin   : Up                Path Oper     : Up
Out Interface: 1/1/1              Out Label    : 524287
---snip---

Adaptive      : Enabled           Oper Metric   : 20
Preference    : n/a
Path Trans    : 8                 CSPF Queries  : 6
Failure Code  : tunnelLocallyRepaired
Failure Node  : 192.0.2.2
Explicit Hops :
  No Hops Specified
Actual Hops   :
  192.168.12.1(192.0.2.1) @ n      Record Label  : N/A
  -> 192.168.12.2(192.0.2.2) @ #   Record Label  : 524287
  -> 192.168.36.1(192.0.2.3)      Record Label  : 524287
---snip---
```

The failure code is tunnelLocallyRepaired and next to the actual hop 192.168.12.2 (PE-2), the symbol # indicates that the detour is in use.

## FRR facility without node protection

Node protection is by default enabled, but can be disabled as follows:

```
# on PE-1:
configure {
```

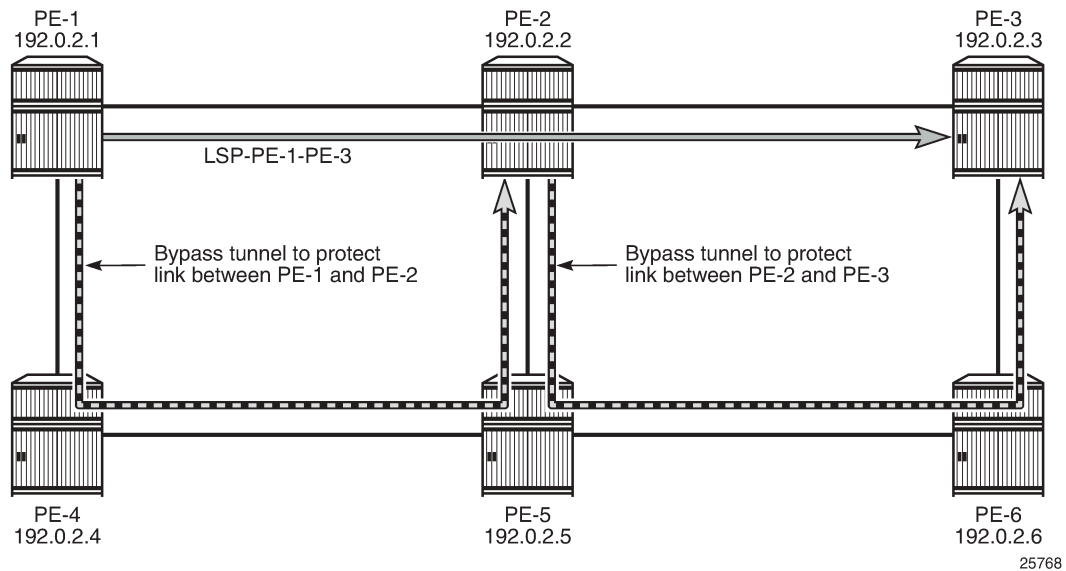
```

router "Base" {
  mpls {
    lsp "LSP-PE-1-PE-3" {
      admin-state enable
      type p2p-rsvp
      to 192.0.2.3
      fast-reroute {
        frr-method facility
        node-protect false
      }
      primary "dyn" {
      }
    }
  }
}

```

As a result, there is only link protection. The bypass tunnels from PE-1 and PE-2 terminate in the next hop in the primary path, as shown in [Figure 100: FRR facility without node protection](#):

Figure 100: FRR facility without node protection



The LSP path detail output shows that there is no node protection:

```

[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-3" path detail
=====
MPLS LSP LSP-PE-1-PE-3 Path (Detail)
=====
Legend :
@ - Detour Available          # - Detour In Use
b - Bandwidth Protected      n - Node Protected
s - Soft Preemption
S - Strict                   L - Loose
A - ABR                      + - Inherited
=====
LSP LSP-PE-1-PE-3
Path dyn
-----
LSP Name      : LSP-PE-1-PE-3

```

```

From          : 192.0.2.1
To            : 192.0.2.3
Admin State   : Up                Oper State    : Up
Path Name     : dyn
Path LSP ID   : 45070             Path Type     : Primary
Path Admin    : Up                Path Oper     : Up
Out Interface : 1/1/1             Out Label     : 524283
---snip---

FRR           : Enabled           Oper FRR      : Enabled
FRR NodeProtect : Disabled       Oper FRR NP   : Disabled
---snip---

Actual Hops   :
  192.168.12.1(192.0.2.1) @      Record Label  : N/A
-> 192.168.12.2(192.0.2.2) @      Record Label  : 524283
-> 192.168.23.2(192.0.2.3)       Record Label  : 524285
---snip---

```

The bypass tunnel originating in PE-1 is now terminating in PE-2 instead of PE-3; as follows:

```

[/]
A:admin@PE-1# show router mpls bypass-tunnel protected-lsp detail

=====
MPLS Bypass Tunnels (Detail)
=====
-----
bypass-link192.168.12.2-61443
-----
-----
To          : 192.168.25.1          State         : Up
Out I/F     : 1/1/2                Out Label     : 524287
Up Time     : 0d 00:02:14          Active Time   : n/a
Reserved BW : 0 Kbps               Protected LSP Count : 1
Type        : Dynamic              Bypass Path Cost : 30
Setup Priority : 7                  Hold Priority  : 0
Class Type  : 0
Exclude Node : None                 Inter-Area    : False
Computed Hops :
  192.168.14.1(S)                  Egress Admin Groups : None
-> 192.168.14.2(S)                  Egress Admin Groups : None
-> 192.168.45.2(S)                  Egress Admin Groups : None
-> 192.168.25.1(S)                  Egress Admin Groups : None
Actual Hops :
  192.168.14.1(192.0.2.1)          Record Label   : N/A
-> 192.168.14.2(192.0.2.4)          Record Label   : 524287
-> 192.168.45.2(192.0.2.5)          Record Label   : 524282
-> 192.168.25.1(192.0.2.2)          Record Label   : 524282
Last Resignal :
Attempted At : n/a                  Resignal Reason  : n/a
Resignal Status: n/a                Reason           : n/a
Protected LSPs -
LSP Name     : LSP-PE-1-PE-3::dyn
From          : 192.0.2.1            To              : 192.0.2.3
Avoid Node/Hop : 192.168.12.2      Downstream Label : 524283
Bandwidth     : 0 Kbps
=====

```

In the remainder of this chapter, this LSP is no longer used. Therefore, the LSP is disabled, as follows:

```

# on PE-1:
configure {

```



```
router "Base" {
  mpls {
    lsp "LSP-PE-1-PE-3" {
      admin-state disable
    }
  }
}
```

## Administrative groups for RSVP-TE LSPs

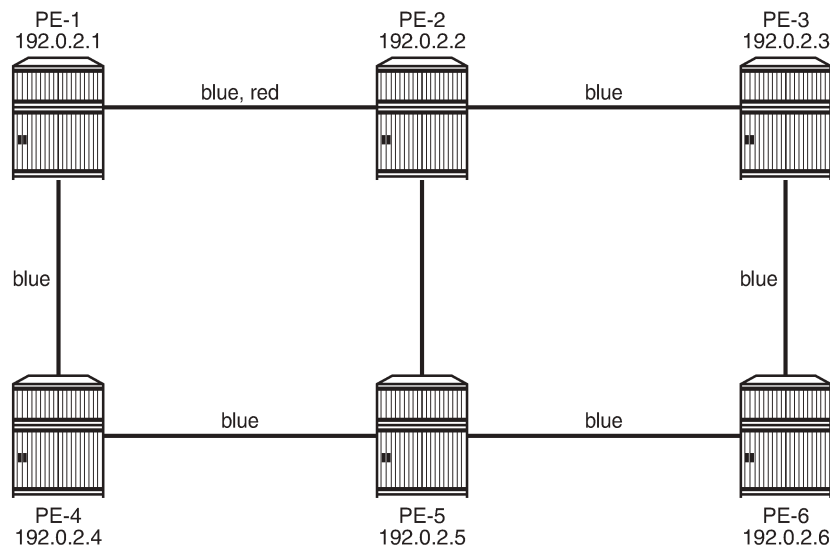
Administrative groups (link-coloring) can be used to calculate a path with the restriction to only include links of a particular admin group (color) or to exclude links of a particular admin group. Paths can be disjointed from each other, without the need for an explicit hops list.

Two admin groups are configured on all nodes; as follows:

```
# on all nodes:
configure {
  routing-options {
    if-attribute {
      admin-group "blue" {
        value 1
      }
      admin-group "red" {
        value 0
      }
    }
  }
}
```

Admin group "blue" is assigned to all MPLS interfaces, except for the link between PE-2 and PE-5 while admin group "red" is only assigned to the link between PE-1 and PE-2; see [Figure 101: Admin groups 'blue' and 'red'](#):

Figure 101: Admin groups 'blue' and 'red'



25825

The admin groups are assigned to the MPLS interfaces as follows:

```
# on PE-1:
configure {
```

```
router "Base" {
  mpls {
    interface "int-PE-1-PE-2" {
      admin-group ["blue" "red"]
    }
    interface "int-PE-1-PE-4" {
      admin-group ["blue"]
    }
  }
}
```

The configuration on the other nodes is similar.

To ensure that FRR bypass tunnels will adhere to the same admin group constraints as defined in the LSP, the following is configured on all nodes. It is required on all Points of Local Repair (PLRs):

```
# on all nodes (at least on all PLRs):
configure {
  router "Base" {
    mpls {
      admin-group-frr true
    }
  }
}
```

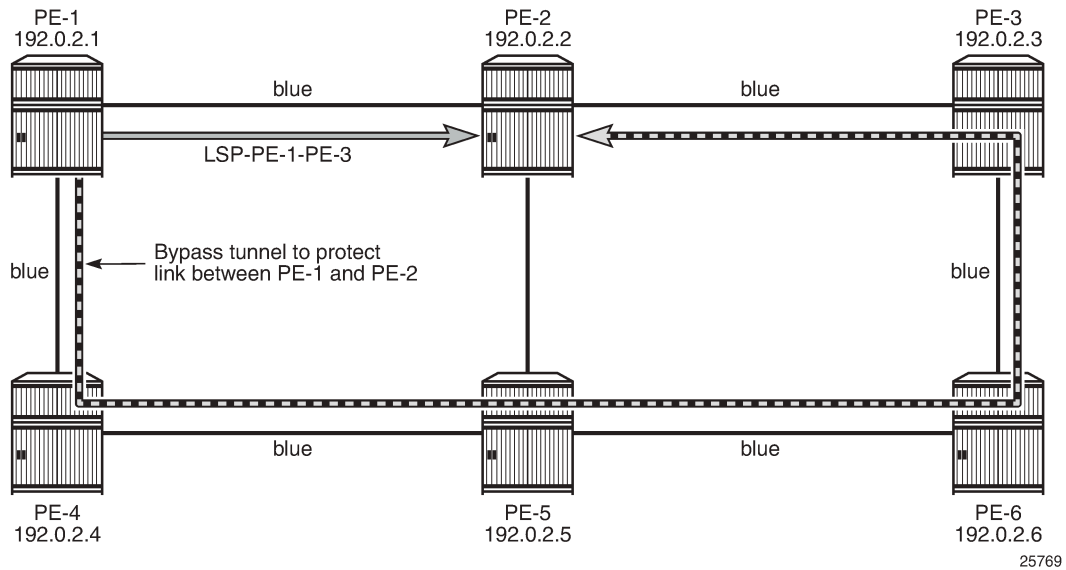
### LSP includes admin group 'blue'

LSP "LSP-PE-1-PE-2" is created on PE-1 with a dynamic primary path. FRR facility is enabled. The LSP includes admin group blue and both the primary path as the bypass tunnel must use links in admin group "blue" (**propagate-admin-group**). **Admin-group-frr** is enabled in the **mpls** context, to ensure that the admin group restriction is respected for FRR.

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      admin-group-frr true
      lsp "LSP-PE-1-PE-2" {
        admin-state enable
        type p2p-rsvp
        to 192.0.2.2
        propagate-admin-group true
        path-computation-method local-cspf
        include-admin-group ["blue"]
        fast-reroute {
          propagate-admin-group true
          frr-method facility
        }
        primary "dyn" {
        }
      }
    }
  }
}
```

The bypass tunnel cannot include the link between PE-2 and PE-5, because that link does not belong to admin group "blue". The LSP and its bypass tunnel are shown in [Figure 102: LSP and bypass within admin group 'blue'](#):

Figure 102: LSP and bypass within admin group 'blue'



The LSP path detailed information is as follows:

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-2" path detail

=====
MPLS LSP LSP-PE-1-PE-2 Path (Detail)
=====
Legend :
@ - Detour Available          # - Detour In Use
b - Bandwidth Protected      n - Node Protected
s - Soft Preemption          L - Loose
S - Strict                   + - Inherited
A - ABR

-----
LSP LSP-PE-1-PE-2
Path dyn
-----
LSP Name      : LSP-PE-1-PE-2
From          : 192.0.2.1
To            : 192.0.2.2
Admin State   : Up           Oper State    : Up
Path Name     : dyn
Path LSP ID   : 64000        Path Type     : Primary
Path Admin    : Up           Path Oper     : Up
Out Interface : 1/1/1        Out Label    : 524287
---snip---

FRR           : Enabled      Oper FRR      : Enabled
FRR NodeProtect : Enabled    Oper FRR NP   : Enabled
FR Hop Limit  : 16           Oper FRHopLimit : 16
FR Prop Admin Gr*: Enabled  Oper FRPropAdmGrp : Enabled
Propagate Adm Grp: Enabled  Oper Prop Adm Grp : Enabled
---snip---

Include Groups :              Oper IncludeGroups:
```

```

blue
Exclude Groups   :
None
---snip---

Actual Hops      :
  192.168.12.1(192.0.2.1) @
-> 192.168.12.2(192.0.2.2)
---snip---

blue
Oper ExcludeGroups:
None

Record Label     : N/A
Record Label     : 524287
    
```

There is a bypass tunnel originating in PE-1 that offers protection for the link between PE-1 and PE-2. More information about this bypass tunnel can be retrieved as follows:

```

[/]
A:admin@PE-1# show router mpls bypass-tunnel protected-lsp detail

=====
MPLS Bypass Tunnels (Detail)
=====
-----
bypass-link192.168.12.2-61444
-----
To           : 192.168.23.1      State          : Up
Out I/F      : 1/1/2           Out Label      : 524287
Up Time     : 0d 00:01:20     Active Time    : n/a
Reserved BW  : 0 Kbps         Protected LSP Count : 1
Type        : Dynamic         Bypass Path Cost : 50
Setup Priority : 7             Hold Priority   : 0
Class Type   : 0
Exclude Node : None           Inter-Area     : False
Computed Hops :
  192.168.14.1(S)             Egress Admin Groups :
-> 192.168.14.2(S)             blue
-> 192.168.45.2(S)            Egress Admin Groups :
-> 192.168.56.2(S)            blue
-> 192.168.36.1(S)           Egress Admin Groups :
-> 192.168.23.1(S)           blue
Actual Hops   :
  192.168.14.1(192.0.2.1)     Record Label      : N/A
-> 192.168.14.2(192.0.2.4)     Record Label      : 524287
-> 192.168.45.2(192.0.2.5)     Record Label      : 524287
-> 192.168.56.2(192.0.2.6)     Record Label      : 524287
-> 192.168.36.1(192.0.2.3)     Record Label      : 524287
-> 192.168.23.1(192.0.2.2)     Record Label      : 524286
Last Resignal :
Attempted At  : n/a           Resignal Reason   : n/a
Resignal Status: n/a         Reason            : n/a

Protected LSPs -
LSP Name     : LSP-PE-1-PE-2::dyn
From         : 192.0.2.1      To                 : 192.0.2.2
Avoid Node/Hop : 192.168.12.2 Downstream Label   : 524287
Bandwidth    : 0 Kbps
    
```

All egress links are in admin group blue on the originating and transit nodes.

## LSP excludes admin group 'red'

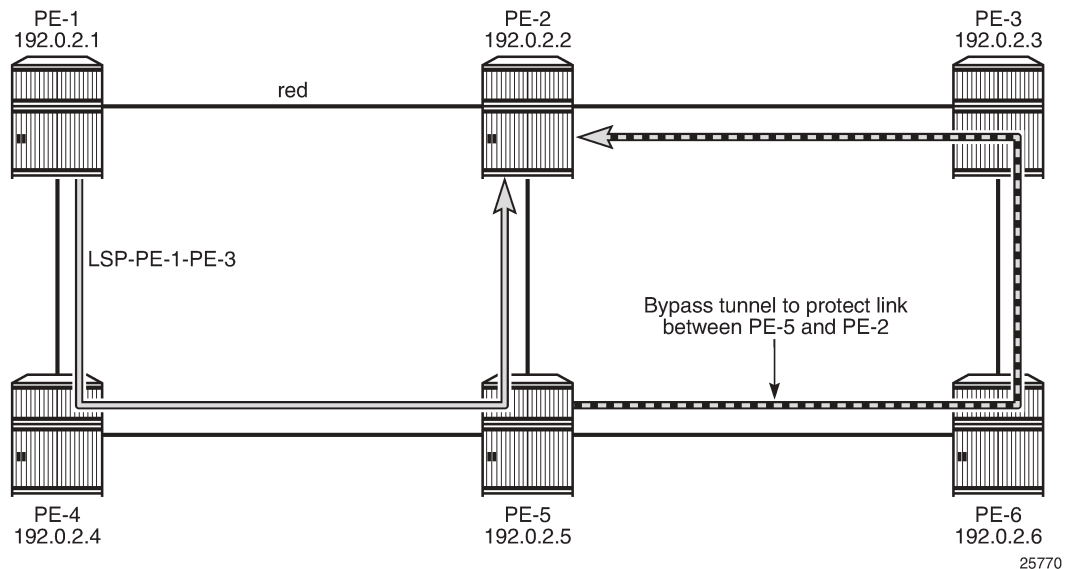
The LSP is reconfigured: instead of including admin group 'blue', it will exclude admin group 'red'. Nothing is changed to the configuration of FRR.

The MPLS configuration is modified as follows:

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-2" {
        delete include-admin-group ["blue"]
        exclude-admin-group ["red"]
      }
    }
  }
}
```

The LSP cannot use the red link between PE-1 and PE-2. The path that avoids the red link, is from PE-1 via PE-4 and PE-5 to PE-2. On all PLRs, **admin-group-frr** is configured, which implies that the originating FRR bypass tunnels need to respect the admin-group constraint of the LSP. There can be no node protection for PE-4 or PE-5 without using the red link between PE-1 and PE-2. The only link that can be protected without using the red link between PE-1 and PE-2, is the link between PE-5 and PE-2. The LSP and the FRR bypass tunnel are shown in [Figure 103: LSP and FRR bypass tunnel excluding admin group 'red'](#):

Figure 103: LSP and FRR bypass tunnel excluding admin group 'red'



The LSP path can be verified as follows:

```
[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-2" path detail

=====
MPLS LSP LSP-PE-1-PE-2 Path (Detail)
=====
Legend :
  @ - Detour Available          # - Detour In Use
```

```

b - Bandwidth Protected      n - Node Protected
s - Soft Preemption
S - Strict                    L - Loose
A - ABR                      + - Inherited
=====
-----
LSP LSP-PE-1-PE-2
Path dyn
-----
LSP Name      : LSP-PE-1-PE-2
From          : 192.0.2.1
To            : 192.0.2.2
Admin State   : Up
Path Name     : dyn
Path LSP ID   : 64002
Path Admin    : Up
Out Interface : 1/1/2
Oper State    : Up
Path Type     : Primary
Path Oper     : Up
Out Label     : 524286
---snip---

Include Groups :
None
Oper IncludeGroups:
None
Exclude Groups :
red
Oper ExcludeGroups:
red
---snip---

Actual Hops :
  192.168.14.1(192.0.2.1)
-> 192.168.14.2(192.0.2.4)
-> 192.168.45.2(192.0.2.5) @
-> 192.168.25.1(192.0.2.2)
Record Label : N/A
Record Label : 524286
Record Label : 524286
Record Label : 524285
---snip---

```

There is only link protection for the link from PE-5 to PE-2. The bypass tunnel originates in PE-5 and has no links belonging to admin group 'red':

```

[/]
A:admin@PE-5# show router mpls bypass-tunnel protected-lsp detail
=====
MPLS Bypass Tunnels (Detail)
=====
-----
bypass-link192.168.25.1-62074
-----
To          : 192.168.23.1      State          : Up
Out I/F     : 1/1/2            Out Label      : 524286
Up Time     : 0d 00:04:38      Active Time    : n/a
Reserved BW : 0 Kbps           Protected LSP Count : 1
Type        : Dynamic          Bypass Path Cost : 30
Setup Priority : 7              Hold Priority   : 0
Class Type  : 0
Exclude Node : None            Inter-Area     : False
Computed Hops :
  192.168.56.1(S)              Egress Admin Groups :
-> 192.168.56.2(S)              blue
-> 192.168.36.1(S)              Egress Admin Groups :
-> 192.168.23.1(S)              blue
Actual Hops :
  192.168.56.1(192.0.2.5)      Egress Admin Groups :
-> 192.168.56.2(192.0.2.6)      blue
Record Label : N/A
Record Label : 524286

```

```

-> 192.168.36.1(192.0.2.3)      Record Label      : 524286
-> 192.168.23.1(192.0.2.2)    Record Label      : 524284
Last Resignal      :
Attempted At      : n/a
Resignal Status   : n/a
Resignal Reason   : n/a
Reason            : n/a

Protected LSPs -
LSP Name          : LSP-PE-1-PE-2::dyn
From              : 192.0.2.1
To                : 192.0.2.2
Avoid Node/Hop    : 192.168.25.1
Bandwidth         : 0 Kbps
Downstream Label  : 524287

```

This configuration is preserved for the following example.

### Hop limit for RSVP-TE LSPs

Another constraint to influence the path selection, is hop limit. This can be configured on the LSP, on a secondary path, or on FRR in case the path should not contain too many hops. In this example, it will be configured on the LSP and later also for FRR on that LSP. By default, the LSP hop limit is 255, but it can be configured as follows:

```

# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-2" {
        hop-limit 5
      }
    }
  }
}

```

This hop limit of 5 is enough for the path via PE-4 and PE-5, but it will not be sufficient when the link between PE-2 and PE-5 is down:

```

# on PE-5:
configure {
  port 1/1/3 {
    admin-state disable
  }
}

```

In this case, the only possible path that excludes the 'red' link between PE-1 and PE-2, has to go to PE-2 via PE-4, PE-5, PE-6, and PE-3. There are too many hops. The FRR bypass tunnel can do a local repair, but no new LSP path can be signaled, with failure code: noCspfRouteToDestination:

```

[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-2" path detail

```

```

=====
MPLS LSP LSP-PE-1-PE-2 Path (Detail)
=====

```

```

Legend :
@ - Detour Available          # - Detour In Use
b - Bandwidth Protected      n - Node Protected
s - Soft Preemption
S - Strict                   L - Loose
A - ABR                      + - Inherited

```

```

-----
LSP LSP-PE-1-PE-2
Path dyn

```

```

-----
LSP Name      : LSP-PE-1-PE-2
From          : 192.0.2.1
To           : 192.0.2.2
Admin State   : Up                               Oper State    : Up
Path Name    : dyn
Path LSP ID   : 64004                             Path Type     : Primary
Path Admin   : Up                               Path Oper     : Up
Out Interface : 1/1/2                             Out Label    : 524287
---snip---

Include Groups :                               Oper IncludeGroups:
None                                                  None
Exclude Groups :                               Oper ExcludeGroups:
red                                                  red

Adaptive      : Enabled                         Oper Metric   : 30
Preference    : n/a
Path Trans    : 3                             CSPF Queries  : 4
Failure Code  : tunnelLocallyRepaired
Failure Node  : 192.0.2.5
Explicit Hops :
  No Hops Specified
Actual Hops   :
  192.168.14.1(192.0.2.1)                       Record Label  : N/A
  -> 192.168.14.2(192.0.2.4)                       Record Label  : 524287
  -> 192.168.45.2(192.0.2.5) @ #                   Record Label  : 524287
  -> 192.168.25.1(192.0.2.2)                       Record Label  : 524287
---snip---

In Prog MBB :
MBB Type     : GlobalRevert                     Next Retry In : 24 sec
Started At   : 02/13/2021 17:32:35             Retry Attempt  : 1
Failure Code : noCspfRouteToDestination        Failure Node   : 192.0.2.1
              n
Signaled BW  : 0 Mbps
=====
* indicates that the corresponding row element may have been truncated.

```

FRR tunnels also have a hop limit. The FRR hop limit is by default 16, but can be configured as follows:

```

# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-2" {
        fast-reroute {
          hop-limit 3
        }
      }
    }
  }
}

```

When the LSP is recalculated, it is impossible to establish the primary path with a hop limit of 5 and it is also impossible to establish a bypass tunnel protecting the link between PE-5 and PE-2 when the FRR hop limit is 3. The LSP will remain operationally down with failure code: noCspfRouteToDestination:

```

[/]
A:admin@PE-1# show router mpls lsp "LSP-PE-1-PE-2" path detail

=====
MPLS LSP LSP-PE-1-PE-2 Path (Detail)
=====
Legend :
  @ - Detour Available                # - Detour In Use

```



```

b - Bandwidth Protected      n - Node Protected
s - Soft Preemption
S - Strict                  L - Loose
A - ABR                     + - Inherited
=====
-----
LSP LSP-PE-1-PE-2
Path dyn
-----
LSP Name      : LSP-PE-1-PE-2
From          : 192.0.2.1
To            : 192.0.2.2
Admin State : Up                Oper State      : Down
Path Name     : dyn
Path LSP ID   : 64008              Path Type      : Primary
Path Admin  : Up                Path Oper    : Down
Out Interface : n/a                Out Label     : n/a
Path Up Time  : 0d 00:00:00        Path Down Time: 0d 00:00:56
---snip---

FRR           : Enabled              Oper FRR       : N/A
FRR NodeProtect : Enabled            Oper FRR NP    : N/A
FR Hop Limit  : 3                    Oper FRHopLimit : N/A
FR Prop Admin Gr* : Enabled          Oper FRPropAdmGrp : N/A
Propagate Adm Grp: Enabled           Oper Prop Adm Grp : N/A
---snip---

Neg MTU       : 0                    Oper MTU       : N/A
Bandwidth     : No Reservation        Oper Bandwidth : N/A
Hop Limit   : 5                  Oper HopLimit  : N/A
---snip---

Include Groups :                      Oper IncludeGroups:
None                                                  N/A
Exclude Groups :                      Oper ExcludeGroups:
red                                                  N/A
---snip---

Failure Code      : noCspfRouteToDestination
Failure Node      : 192.0.2.1
Explicit Hops     :
  No Hops Specified
Actual Hops       :
  No Hops Specified
---snip---

```

For the remainder of the examples, FRR is disabled and the hop limit is restored to the default value, which is 255:

```

# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-1-PE-2" {
        delete hop-limit
        delete fast-reroute
      }
    }
  }
}

```

On PE-5, port 1/1/3 is enabled, as follows:

```

# on PE-5:
configure {
  port 1/1/3 {

```

```
admin-state enable
```

## Manual resignal

Instead of waiting for the resignal timer to expire, one can manually trigger the resignal process.

The command to resignal the path "dyn" of LSP "LSP-PE-1-PE-2":

```
[/]
A:admin@PE-1# tools perform router mpls resignal lsp "LSP-PE-1-PE-2" path "dyn"
```

The command to resignal all RSVP LSPs originating at node PE-1:

```
[/]
A:admin@PE-1# tools perform router mpls resignal delay 0
WARNING: CLI #2006: Warning while processing command - WARNING: CLI Delay will not be in
effect, configure resignal-timer at config>router>mpls>resignal-timer .
```

The preceding command overrides the resignal timer in the **mpls** context, so it can only be launched after the resignal timer is configured.

```
[ex:configure router "Base" mpls]
A:admin@PE-1# resignal-timer ?

resignal-timer <number>
<number> - <30..10080> - minutes

Resignal timer for RSVP LSPs
```

The configuration timer is configured to 30 minutes as follows:

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      resignal-timer 30
```

Whenever an LSP is resigaled, the resignal timer is restarted.

## LSP OAM

The LSP diagnostics are modeled after ICMP echo request/reply which provides a mechanism to detect data plane failures in MPLS LSPs. For a given FEC, LSP ping verifies whether the packet reaches the egress label edge router (LER). The following OAM commands are in classic CLI.

```
[/]
A:admin@PE-1# //
INFO: CLI #2051: Switching to the classic CLI engine
A:PE-1# oam lsp-ping "LSP-PE-1-PE-2"
LSP-PING LSP-PE-1-PE-2: 92 bytes MPLS payload
Seq=1, send from intf int-PE-1-PE-4, reply from 192.0.2.2
udp-data-len=32 ttl=255 rtt=3.71ms rc=3 (EgressRtr)

---- LSP LSP-PE-1-PE-2 PING Statistics ----
1 packets sent, 1 packets received, 0.00% packet loss
```

```
round-trip min = 3.71ms, avg = 3.71ms, max = 3.71ms, stddev = 0.000ms
```

In LSP traceroute mode, the packet is sent to the control plane of each transit label switched router (LSR) which performs various checks to see if it is actually a transit LSR for the path.

```
A:PE-1# oam lsp-trace "LSP-PE-1-PE-2"
lsp-trace to LSP-PE-1-PE-2: 0 hops min, 0 hops max, 116 byte packets
1 192.0.2.4 rtt=2.60ms rc=8(DSRtrMatchLabel) rsc=1
2 192.0.2.5 rtt=3.82ms rc=8(DSRtrMatchLabel) rsc=1
3 192.0.2.2 rtt=3.54ms rc=3(EgressRtr) rsc=1
```

```
A:PE-1# oam lsp-trace "LSP-PE-1-PE-2" detail
lsp-trace to LSP-PE-1-PE-2: 0 hops min, 0 hops max, 116 byte packets
1 192.0.2.4 rtt=2.30ms rc=8(DSRtrMatchLabel) rsc=1
   DS 1: ipaddr=192.168.45.2 ifaddr=192.168.45.2 iftype=ipv4Numbered MRU=1564
        label[1]=524286 protocol=4(RSVP-TE)
2 192.0.2.5 rtt=3.70ms rc=8(DSRtrMatchLabel) rsc=1
   DS 1: ipaddr=192.168.25.1 ifaddr=192.168.25.1 iftype=ipv4Numbered MRU=1564
        label[1]=524286 protocol=4(RSVP-TE)
3 192.0.2.2 rtt=3.84ms rc=3(EgressRtr) rsc=1
```

## RSVP LSP statistics

Statistics can be collected for RSVP LSPs. For each accounting record, a file ID is configured; as follows:

```
# on PE-1:
configure {
  log {
    file 2 {
      rollover 5
      retention 1
      compact-flash-location {
        primary cfl
      }
    }
  }
}
```

An accounting policy is configured for each record type; as follows:

```
# on PE-1:
configure {
  log {
    accounting-policy 2 {
      admin-state enable
      record combined-mpls-lsp-ingress
      destination {
        file 2
      }
    }
  }
}
```

The collection of statistics is enabled in the **mpls** context as follows:

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      ingress-statistics {
        lsp sender 192.0.2.1 lsp-name "LSP-PE-1-PE-2" {
```

```

        admin-state enable
        collect-stats true
        accounting-policy 2
    }
}

```

To display the statistics, the following options are available for `lsp-ingress-stats`:

```

[/]
A:admin@PE-1# show router mpls lsp-ingress-stats ?

lsp-ingress-stats [type <keyword>] [active]
lsp-ingress-stats lsp <string> sender <ipv4 address>
lsp-ingress-stats [sender <ipv4 address>] [type <keyword>] [active] template-match
  SessionNameString <string>

SessionNameString    - [Max 64 chars]
active                - match on all stats enabled lsp
lsp                   - max 64 chars
sender                - ipv4 address '<d.d.d.d>'
template-match       - match on p2p/p2mp stats template
type                  - p2p|p2mp

```

The following command retrieves the LSP ingress statistics for LSP "LPS-PE-1-PE-2" with sender 192.0.2.1:

```

[/]
A:admin@ PE-1# show router mpls lsp-ingress-stats lsp "LSP-PE-1-PE-2"
                                                    sender 192.0.2.1

=====
MPLS LSP Ingress Statistics
=====
-----
LSP Name       : LSP-PE-1-PE-2
Sender        : 192.0.2.1
-----
Collect Stats  : Enabled                Accting Plcy.  : 2
Adm State     : Up                      PSB Match     : False
FC BE
InProf Pkts   : 0                       OutProf Pkts   : 0
InProf Octets : 0                       OutProf Octets : 0
FC L2
InProf Pkts   : 0                       OutProf Pkts   : 0
InProf Octets : 0                       OutProf Octets : 0
FC AF
InProf Pkts   : 0                       OutProf Pkts   : 0
InProf Octets : 0                       OutProf Octets : 0
FC L1
InProf Pkts   : 0                       OutProf Pkts   : 0
InProf Octets : 0                       OutProf Octets : 0
FC H2
InProf Pkts   : 0                       OutProf Pkts   : 0
InProf Octets : 0                       OutProf Octets : 0
FC EF
InProf Pkts   : 0                       OutProf Pkts   : 0
InProf Octets : 0                       OutProf Octets : 0
FC H1
InProf Pkts   : 0                       OutProf Pkts   : 0
InProf Octets : 0                       OutProf Octets : 0
FC NC
InProf Pkts   : 0                       OutProf Pkts   : 0

```

```
InProf Octets      : 0                      OutProf Octets    : 0
Aggregate Pkts    : 0                      Aggregate Octets  : 0
=====
```

Statistics can be cleared as follows:

```
[clear router mpls]
A:admin@PE-1# lsp-ingress-stats 192.0.2.1 lsp "LSP-PE-1-PE-2"
```

```
[clear router mpls]
A:admin@PE-1# lsp-ingress-stats
```

## Debug

A wide range of debug tools are available which can be tuned to the specific information of importance for a certain troubleshooting task. These tools are in classic CLI. In the **debug router mpls** context, the LSP object to trace or monitor can be selected by the following parameters:

- LSP name
- Source address of the LSP (the from parameter in the LSP definition)
- Termination point of the LSP (the to parameter in the LSP definition)
- Tunnel ID of the LSP
- LSP ID

```
A:PE-1# debug router rsvp ?
- no rsvp
- rsvp [lsp name>] [sender <sender-address>] [endpoint <endpoint-address>]
      [tunnel-id <tunnel-id>] [lsp-id <lsp-id>] [interface <ip-int-name>]

<name>          : [160 chars max]
<sender-address> : a.b.c.d
<endpoint-address> : a.b.c.d
<tunnel-id>     : [0..4294967295]
<lsp-id>       : [1..65535]
<ip-int-name>  : [32 chars max]

[no] event      + Enable/disable debugging for specific RSVP events
[no] packet     + Enable/disable debugging for specific RSVP packets
```

```
A:PE-1# debug router mpls ?
- mpls [lsp <name>] [sender <ip-address|ipv6-address>] [endpoint <ip-address|
      ipv6-address>] [tunnel-id <tunnel-id>] [lsp-id <lsp-id>]
- no mpls

<name>          : [160 chars max]
<ip-address|ipv6-a*> : [64 chars max]
<ip-address|ipv6-a*> : [64 chars max]
<tunnel-id>     : [0..4294967295]
<lsp-id>       : [1..65535]

[no] event      + Enable/disable debugging for specific MPLS events
```

```
[no] forwarding-pol* + Enable/disable debugging for MPLS Forwarding-Policies
```

In the **debug** command tree, the MPLS event type can be selected (tracing must be enabled):

```
A:PE-1# debug router mpls lsp "LSP-PE-1-PE-2" event ?
- event
- no event

[no] all          - Enable/disable debugging for MPLS all
[no] frr          - Enable/disable debugging for MPLS frr
[no] iom          - Enable/disable debugging for MPLS iom
[no] lsp-setup    - Enable/disable debugging for MPLS lsp setup
[no] mbb          - Enable/disable debugging for MPLS mbb
[no] misc         - Enable/disable debugging for MPLS misc
[no] pcc          - Enable/disable debugging for MPLS PCC
[no] te           - Enable/disable debugging for MPLS TE
[no] xc           - Enable/disable debugging for MPLS xc
```

As an example, the all keyword is entered, logging all MPLS events related to the selected LSP:

```
# on PE-1:
debug
  router "Base"
    mpls lsp "LSP-PE-1-PE-2"
      event
        all
```

```
A:PE-1# show debug
debug
  router "Base"
    mpls lsp "LSP-PE-1-PE-2"
      event
        iom
        lsp-setup
        xc
        frr
        mbb
        misc
        pcc
        te
      exit
    exit
  exit
exit
```

The last step is to create a log container which will gather all MPLS debugging information according to the criteria set in the debug context. The **from debug-trace** parameter must be configured but there are several options where the different captured entries will be stored: console, a syslog server, SNMP, local file on the compact flash card, a temporary circular memory buffer, or the telnet/SSH session from which you are logged into the node.

The ID of the log container is a local number without any other significance.

```
[ex:configure log log-id 2]
A:admin@PE-1# destination ?

destination
```

Note: 'configure log log-id 2' and all other elements in this context support single-model management only.

```

Immutable fields      - file, syslog, console

Choice: log-id-destination
cli                   :+ Enable the cli context
console               :- Log events to send to the system console port
file                  :- Log events to send to the specified file ID
memory                :+ Enable the memory context
netconf               :+ Enable the netconf context
snmp                  :+ Enable the snmp context
syslog                :- Log events to send to the specified syslog ID
    
```

For this example, the temporary buffer (with adjustable size) is chosen, as follows:

```

# on PE-1:
configure {
  log {
    log-id 2 {
      source {
        debug true
      }
      destination {
        memory {
        }
      }
    }
  }
}
    
```

All MPLS events related to the selected LSP are stored in the location (memory) specified. The content of this log container can be viewed through the **show log log-id 2** command. The following output is a subset of messages shown after port 1/1/2 on PE-2 is disabled, which causes LSP "LSP-PE-1-PE-2" to go down.

```

[/]
A:admin@PE-1# show log log-id 2 ascending

=====
Event Log 2 log-name 2
=====
Description : (Not Specified)
Memory Log contents [size=100  next event=19  (not wrapped)]

1 2021/02/13 20:16:29.034 CET MINOR: DEBUG #2001 Base MPLS
"MPLS: LSP Path
Signalling failure for LspPath LSP-PE-1-PE-2::dyn(LspId 64022)"

2 2021/02/13 20:16:29.034 CET MINOR: DEBUG #2001 Base MPLS
"MPLS: CSPF
Delete CSPF Hop list 18"

3 2021/02/13 20:16:29.034 CET MINOR: DEBUG #2001 Base MPLS
"MPLS: LSP Path
Set operational state for LspPath LSP-PE-1-PE-2::dyn(LspId 64022) to Down, previous state is
Up"

4 2021/02/13 20:16:29.034 CET MINOR: DEBUG #2001 Base MPLS
"MPLS: LSP Path
Set operational MTU for LspPath LSP-PE-1-PE-2::dyn(LspId 64022) to 0"

5 2021/02/13 20:16:29.034 CET MINOR: DEBUG #2001 Base MPLS
"MPLS: LSP Path
Set operational metric for LspPath LSP-PE-1-PE-2::dyn(LspId 64022) to 30"
---snip---
    
```

## Conclusion

MPLS provides the capability to establish connection oriented paths over a connectionless network. The LSP offers a mechanism to engineer network traffic on constraint-based paths rather than the IGP shortest path. This can greatly improve network resiliency. In this chapter, the configuration of several RSVP LSP features is given together with the associated show output which can be used to verify and troubleshoot.



# RSVP Signaled Point-to-Multipoint LSPs

This chapter provides information about RSVP-signaled point-to-multipoint LSPs.

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

## Applicability

This chapter was originally written for SR OS Release 7.0.R5, but the MD-CLI in the current edition corresponds to SR OS Release 24.10.R2.

## Overview

Point-to-MultiPoint (P2MP) Multi-Protocol Label Switching (MPLS) Label Switched Paths (LSPs) allow the source of multicast traffic to forward packets to one or many multicast receivers over a network without requiring a multicast protocol, such as Protocol Independent Multicast (PIM), to be configured in the network. A P2MP LSP tree is established in the control plane, and the path consists of a head-end node, one or many branch and bud nodes, and the leaf nodes. A bud node combines the roles of branch node and leaf node (for different source-to-leaf (S2L) LSPs). Packets injected by the head-end node are replicated in the data plane at the branching nodes before they are delivered to the leaf nodes.

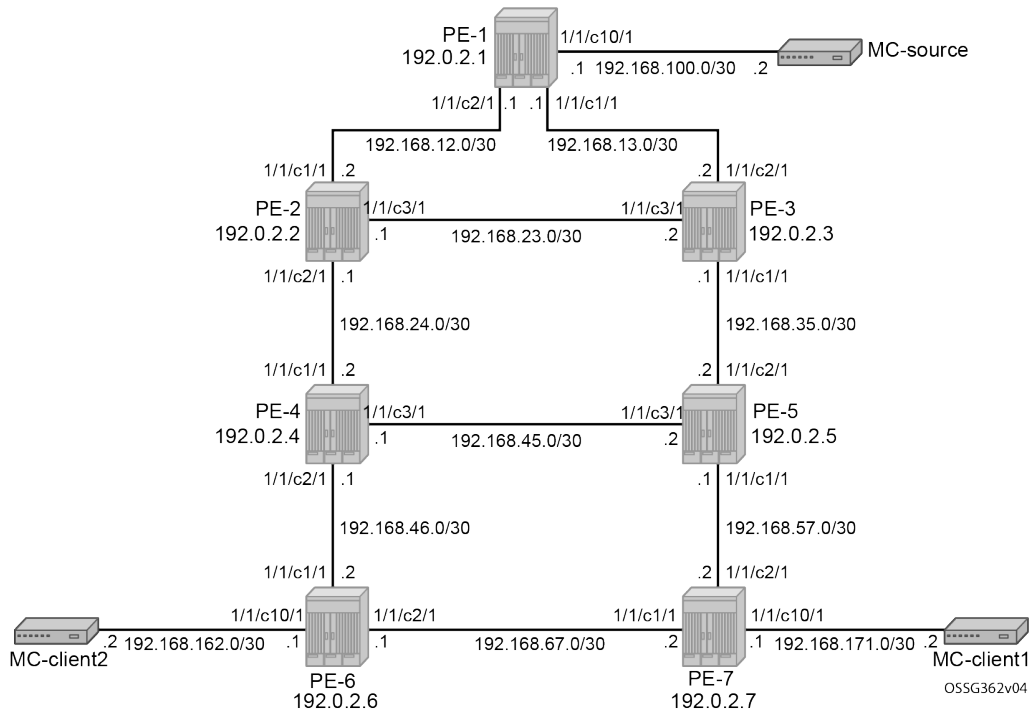
Similar to point-to-point (P2P) LSPs, also P2MP LSPs are unidirectional, originating on a head-end node (the ingress LER) and terminating on one or more leaf nodes (the egress LERs). Resource Reservation Protocol (RSVP) is used as signaling protocol. A P2MP LSP is modeled as a set of root-to-leaf sub LSPs (S2L). Each S2L is modeled as a point-to-point LSP in the control plane. This means that each S2L has its own PATH/RESV messages. This is called the de-aggregated method.

The forwarding of multicast packets to the LSP tree was initially based on static multicast routes, and has evolved to BGP-based VPN routes afterward (but the latter is beyond the scope of this chapter). In this example, forwarding multicast packets is done over P2MP RSVP LSPs in the base router instance.

RSVP signaled P2MP LSPs can have fast reroute (FRR) enabled. The FRR facility method (one-to-many) with link protection is supported.

**Figure 104: P2MP example topology** shows the P2MP example topology with seven PEs. The multicast source is connected to PE-1, multicast client 1 is attached to PE-7, and multicast client 2 to PE-6, as follows:

Figure 104: P2MP example topology



## Configuration

The following sections describe the steps that must be performed to configure RSVP signaled point-to-multipoint LSPs.

### Configuring the IP/MPLS network

The system addresses and Layer 3 interface addresses are configured according to [Figure 104: P2MP example topology](#). An Interior Gateway Protocol (IGP) is needed to distribute routing information to all PEs. In this case, the IGP is OSPF using the backbone area 0.0.0.0. A configuration example is shown for PE-1. A similar configuration is needed on all PEs.

```
# on PE-1:
configure {
  router "Base" {
    interface "int-PE-1-PE-2" {
      port 1/1/c2/1:1000
      ipv4 {
        primary {
          address 192.168.12.1
          prefix-length 30
        }
      }
    }
  }
  interface "int-PE-1-PE-3" {
```

```

port 1/1/c1/1:1000
ipv4 {
    primary {
        address 192.168.13.1
        prefix-length 30
    }
}
interface "system" {
    ipv4 {
        primary {
            address 192.0.2.1
            prefix-length 32
        }
    }
}
ospf 0 {
    admin-state enable
    traffic-engineering true
    area 0.0.0.0 {
        interface "int-PE-1-PE-2" {
            interface-type point-to-point
        }
        interface "int-PE-1-PE-3" {
            interface-type point-to-point
        }
        interface "system" {
        }
    }
}
}

```

Because FRR is enabled for the P2MP LSP, Traffic Engineering (TE) is needed on the IGP. With TE enabled, OSPF generates opaque link state advertisements (LSAs) which are collected in a Traffic Engineering Database (TED), separate from the traditional OSPF topology database. OSPF interfaces are set up as type **point-to-point** to improve convergence, because no Designated Router/Backup Designated Router (DR/BDR) election process is done. However, convergence is out of the scope of this chapter.

The following command verifies that the OSPF neighbors are up (state *Full*):

```

[/]
A:admin@PE-1# show router ospf neighbor
=====
Rtr Base OSPFv2 Instance 0 Neighbors
=====
Interface-Name          Rtr Id          State    Pri  RetxQ  TTL
Area-Id
-----
int-PE-1-PE-2          192.0.2.2      Full     1    0      33
0.0.0.0
int-PE-1-PE-3          192.0.2.3      Full     1    0      34
0.0.0.0
-----
No. of Neighbors: 2
=====

```

The route table shows the Layer 3 interface addresses/subnets that are known on the PEs. The route table on head-end PE-1 is as follows:

```

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

```

```

=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]                Type  Proto  Age           Pref
  Next Hop[Interface Name]          Metric
-----
192.0.2.1/32                      Local Local  00h03m06s    0
   system
192.0.2.2/32                      Remote OSPF   00h02m05s   10
   192.168.12.2
192.0.2.3/32                      Remote OSPF   00h01m25s   10
   192.168.13.2
192.0.2.4/32                      Remote OSPF   00h00m25s   10
   192.168.12.2
192.0.2.5/32                      Remote OSPF   00h00m12s   10
   192.168.13.2
192.0.2.6/32                      Remote OSPF   00h00m08s   10
   192.168.12.2
192.0.2.7/32                      Remote OSPF   00h00m02s   10
   192.168.13.2
192.168.12.0/30                   Local Local  00h03m06s    0
   int-PE-1-PE-2
192.168.13.0/30                   Local Local  00h03m06s    0
   int-PE-1-PE-3
192.168.23.0/30                   Remote OSPF   00h02m05s   10
   192.168.12.2
192.168.24.0/30                   Remote OSPF   00h02m05s   10
   192.168.12.2
192.168.35.0/30                   Remote OSPF   00h01m25s   10
   192.168.13.2
192.168.45.0/30                   Remote OSPF   00h00m25s   10
   192.168.12.2
192.168.46.0/30                   Remote OSPF   00h00m25s   10
   192.168.12.2
192.168.57.0/30                   Remote OSPF   00h00m12s   10
   192.168.13.2
192.168.67.0/30                   Remote OSPF   00h00m08s   10
   192.168.12.2
-----
No. of Routes: 16
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
=====
    
```

The following command shows the contents of the forwarding information base (FIB) for IOM card slot 1:

```

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

=====
FIB Display
=====
Prefix [Flags]                Protocol
  NextHop
-----
192.0.2.1/32                      LOCAL
   192.0.2.1 (system)
192.0.2.2/32                      OSPF
   192.168.12.2 (int-PE-1-PE-2)
192.0.2.3/32                      OSPF
   192.168.13.2 (int-PE-1-PE-3)
    
```

```

192.0.2.4/32                                OSPF
  192.168.12.2 (int-PE-1-PE-2)
192.0.2.5/32                                OSPF
  192.168.13.2 (int-PE-1-PE-3)
192.0.2.6/32                                OSPF
  192.168.12.2 (int-PE-1-PE-2)
192.0.2.7/32                                OSPF
  192.168.13.2 (int-PE-1-PE-3)
192.168.12.0/30                             LOCAL
  192.168.12.0 (int-PE-1-PE-2)
192.168.13.0/30                             LOCAL
  192.168.13.0 (int-PE-1-PE-3)
192.168.23.0/30                             OSPF
  192.168.12.2 (int-PE-1-PE-2)
192.168.24.0/30                             OSPF
  192.168.12.2 (int-PE-1-PE-2)
192.168.35.0/30                             OSPF
  192.168.13.2 (int-PE-1-PE-3)
192.168.45.0/30                             OSPF
  192.168.12.2 (int-PE-1-PE-2)
192.168.46.0/30                             OSPF
  192.168.12.2 (int-PE-1-PE-2)
192.168.57.0/30                             OSPF
  192.168.13.2 (int-PE-1-PE-3)
192.168.67.0/30                             OSPF
  192.168.12.2 (int-PE-1-PE-2)
-----
Total Entries : 16
-----
=====

```

On PE-1, the interface toward the multicast source is configured in an IES service. This could have been on a router interface instead.

```

# on PE-1:
configure {
  service {
    ies "IES 1" {
      admin-state enable
      service-id 1
      customer "1"
      interface "int-PE-1-MC-source" {
        sap 1/1/c10/1 {
        }
        ipv4 {
          primary {
            address 192.168.100.1
            prefix-length 30
          }
        }
      }
    }
  }
  router "Base"
  ospf 0 {
    area 0.0.0.0 {
      interface "int-PE-1-MC-source" {
        interface-type point-to-point
      }
    }
  }
}

```

Similar IES services are configured on PE-7 and PE-6 for multicast client 1 and multicast client 2.

```
# on PE-7:
configure {
  service {
    ies "IES 1" {
      admin-state enable
      service-id 1
      customer "1"
      interface "int-PE-7-MC-client1" {
        sap 1/1/c10/1 {
        }
        ipv4 {
          primary {
            address 192.168.171.1
            prefix-length 30
          }
        }
      }
    }
  }
}
```

```
# on PE-6:
configure {
  service {
    ies "IES 1" {
      admin-state enable
      service-id 1
      customer "1"
      interface "int-PE-6-MC-client2" {
        sap 1/1/c10/1 {
        }
        ipv4 {
          primary {
            address 192.168.162.1
            prefix-length 30
          }
        }
      }
    }
  }
}
```

The next step in the process of setting up a P2MP LSP, is enabling the L3 interfaces in the MPLS and RSVP context on all involved PE nodes (from PE-1 to PE-7). Explicit enabling of the MPLS and RSVP context is done by the **admin-state enable** command. The MPLS and RSVP configuration for PE-1 is as follows:

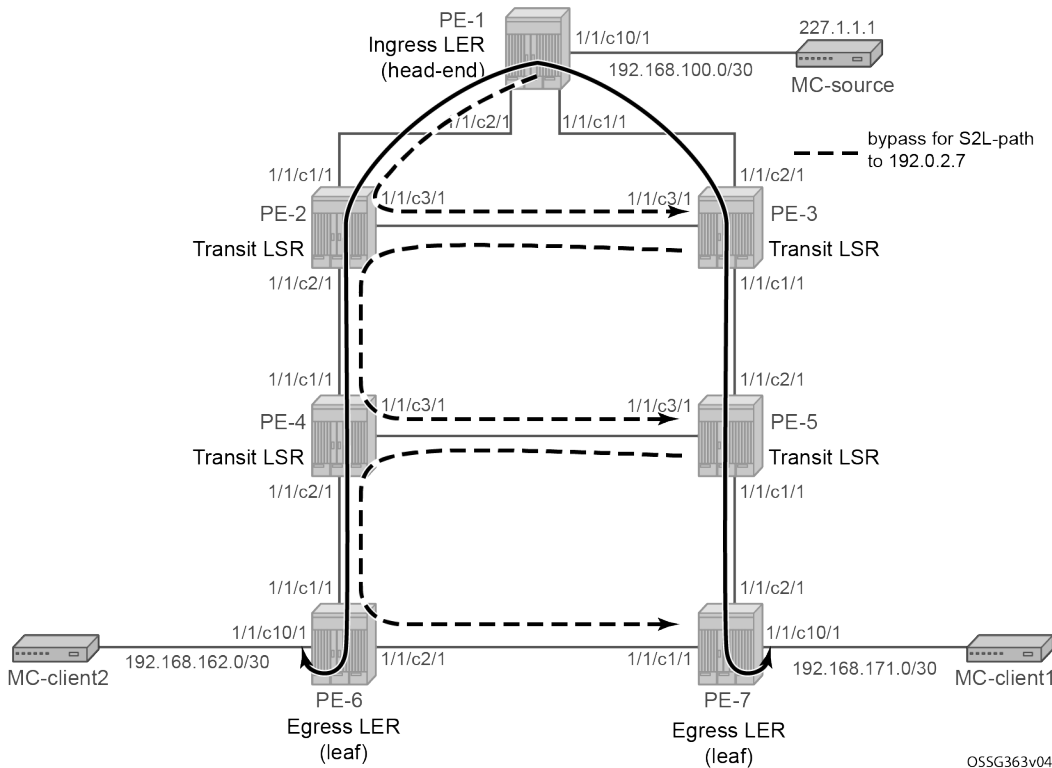
```
# on PE-1:
configure {
  router "Base" {
    mpls {
      admin-state enable
      interface "int-PE-6-PE-4" {
      }
      interface "int-PE-6-PE-7" {
      }
    }
    rsvp {
      admin-state enable
      interface "int-PE-6-PE-4" {
      }
      interface "int-PE-6-PE-7" {
      }
    }
  }
}
```

}

## Configuring P2MP RSVP LSP

Figure 105: P2MP LSP LSP-p2mp-1 with bypass tunnels shows the P2MP LSP LSP-p2mp-1 with FRR facility backup.

Figure 105: P2MP LSP LSP-p2mp-1 with bypass tunnels



A P2MP LSP (LSP-p2mp-1) is set up from PE-1 acting as head-end node and PE-6 and PE-7 acting as leaf nodes. Because FRR is enabled, Constrained Shortest Path First (CSPF) is enabled to calculate the routes on the Traffic Engineering Database (TED). FRR facility method is used without node protection; the facility method stands for one-to-many, meaning that one bypass tunnel can protect a set of primary LSPs with similar backup constraints. When a link failure occurs on one of the active S2L paths, the Point of Local Repair (PLR) node pushes an additional MPLS label on the incoming MPLS packet before sending it into the bypass tunnel downstream toward the Merge Point (MP) node.

In the first example, the IGP OSPF calculates the path to the two destinations PE-6 and PE-7. The intermediate hops of the LSP are dynamically assigned by OSPF best route selection, so the S2L paths follow the IGP least cost path. Therefore, an empty MPLS path is configured without specifying any hops.

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

```
}

```

The P2MP LSP "LSP-p2mp-1" is created on the ingress LER or head-end node (PE-1 in the example) and can be seen in following CLI output. The **type p2mp-rsvp** command is added to make a distinction in configuration between normal point-to-point LSPs and point-to-multipoint LSPs. A primary P2MP instance is initiated using the **primary-p2mp-instance** keyword accompanied with the P2MP instance name "p-LSP-p2mp-1". Within this primary P2MP instance, the different S2Ls are defined using the **s2l-path** keyword. The same MPLS path name can be used for different S2Ls as long as the destination is different (**to** command).

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-p2mp-1" {
        admin-state enable
        type p2mp-rsvp
        from 192.0.2.1
        path-computation-method local-cspf
        fast-reroute {
        }
        primary-p2mp-instance "p-LSP-p2mp-1" {
          s2l-path "empty" to 192.0.2.6 {
          }
          s2l-path "empty" to 192.0.2.7 {
          }
        }
      }
    }
  }
}
```

On the head-end LER node of the P2MP LSP, several **show** commands can be used. The following command shows that the only P2MP LSP on PE-1 is LSP-p2mp-1 and the administrative and operational state of this P2MP LSP is up and FRR is enabled:

```
[/]
A:admin@PE-1# show router mpls p2mp-lsp

=====
MPLS P2MP LSPs (Originating)
=====
LSP Name                               Tun    Fastfail  Adm  Opr
                                   Id      Config
-----
LSP-p2mp-1                             1      Yes       Up   Up
-----
LSPs : 1
=====
```

In this example, P2MP LSP "LSP-p2mp-1" has two active S2L paths: one toward leaf node PE-6 and one to leaf node PE-7. FRR is enabled and the FRR method is facility.

```
[/]
A:admin@PE-1# show router mpls p2mp-lsp "LSP-p2mp-1" detail

=====
MPLS P2MP LSPs (Originating) (Detail)
=====
Legend :
+ - Inherited
=====
-----
```



```

Type : Originating
-----
LSP Name      : LSP-p2mp-1
LSP Type     : P2mpLsp           LSP Tunnel ID      : 1
LSP Index    : 1                 TTM Tunnel Id     : 1
From         : 192.0.2.1
To           : Not configured
Adm State    : Up                Oper State         : Up
LSP Up Time  : 0d 00:00:06       LSP Down Time     : 0d 00:00:00
Transitions  : 1                 Path Changes      : 1
Retry Limit  : 0                 Retry Timer       : 30 sec
Signaling    : RSVP             Resv. Style       : SE
Hop Limit    : 255              Negotiated MTU    : n/a
Adaptive     : Enabled          ClassType         : 0
FastReroute  : Enabled          Oper FR           : Enabled
FR Method    : Facility         FR Hop Limit      : 16
FR Node Protect : Disabled      FR Prop Adm Grp   : Disabled
FR Object    : Enabled
PathCompMethod : local-cspf     ADSPEC            : Disabled
FallbkPathComp : not-applicable Metric Type        : igp
Metric       : Disabled         ClassForwarding   : Disabled
Load Bal Wt  : N/A             Exclude Grps      :
None
Least Fill   : Disabled         Soft Preemption   : Disabled

Revert Timer : Disabled         Next Revert In   : N/A
Auto BW      : Disabled
LdpOverRsvp : Disabled
VprnAutoBind : Disabled
IGP Shortcut : Disabled         BGP Shortcut     : Disabled
IGP LFA      : Disabled         IGP Rel Metric   : Disabled
AllowSrOverSrte : Disabled
BGPTransTun : Disabled
Oper Metric  : Disabled
Prop Adm Grp : Disabled

P2MPInstance : p-LSP-p2mp-1     P2MP-Inst-type   : Primary
S2L Cfg Counter : 2             S2L Oper Counter : 2
S2L-Name     : empty
To           : 192.0.2.6
S2L-Name     : empty
To           : 192.0.2.7
=====

```

```

[/]
A:admin@PE-1# show router mpls p2mp-info

=====
MPLS P2MP Cross Connect Information
=====
S2L:LSP-p2mp-1::empty
-----
Source IP Address : 192.0.2.1      Tunnel ID      : 1
P2MP ID          : 0              Lsp ID        : 11264
To               : 192.0.2.6
Out Interface    : 1/1/c2/1:1000   Out Label     : 524287
Num. of S2ls    : 1
-----
S2L:LSP-p2mp-1::empty
-----

```

```
Source IP Address : 192.0.2.1      Tunnel ID      : 1
P2MP ID          : 0              Lsp ID        : 11264
To               : 192.0.2.7
Out Interface    : 1/1/c1/1:1000   Out Label     : 524287
Num. of S2ls    : 1
```

```
-----
P2MP Cross-connect instances : 2
=====
```

```
[/]
A:admin@PE-1# show router mpls p2mp-lsp "LSP-p2mp-1" p2mp-instance p2mp-instance-name "p-LSP-p2mp-1"
```

```
=====
MPLS P2MP Instance (Originating)
=====
```

```
-----
Type : Originating
-----
```

```
LSP Name      : LSP-p2mp-1
P2MP ID       : 0
Adm State     : Up
LSP Tunnel ID : 1
Oper State    : Up

P2MPInstance  : p-LSP-p2mp-1
P2MP-Inst-type : Primary
P2MP Lsp Id   : 11264
Inst Admin    : Up
Inst Oper     : Up
Inst Up Time  : 0d 00:00:06
Inst Dn Time  : 0d 00:00:00
Hop Limit     : 255
Adaptive      : Enabled
Record Route  : Record
Record Label  : Record
Setup Priority : 7
Hold Priority  : 0
Exclude Grps  :
None          : None
Bandwidth     : No Reservation
Oper Bw       : 0 Mbps
S2L-Name      : empty
To            : 192.0.2.6
S2L Admin    : Up
S2L Oper     : Up
S2L-Name     : empty
To            : 192.0.2.7
S2L Admin    : Up
S2L Oper     : Up
```

```
-----
P2MP instances : 1
=====
```



**Note:**

As long as one S2L path is operationally up (**show router mpls p2mp-lsp <p2mp-lsp-name> p2mp-instance p2mp-instance-name <p2mp-instance-name>**), the operational state of the P2MP LSP is up.

FRR information can be displayed in detail for each S2L path. From this moment onward, the focus is on the S2L path toward PE-7. The following command shows that link protection is present for the link between PE-1 and PE-3, for the link between PE-3 and PE-5, and for the link between PE-5 and PE-7 (as indicated by the '@'-reference inside the list of actual hops).

```
[/]
A:admin@PE-1# show router mpls p2mp-lsp "LSP-p2mp-1" p2mp-instance p2mp-instance-name "p-LSP-p2mp-1" s2l s2l-name "empty" to 192.0.2.7 detail
```

```
=====
MPLS LSP LSP-p2mp-1 S2L empty (Detail)
```

```

=====
Legend :
  @ - Detour Available          # - Detour In Use
  b - Bandwidth Protected      n - Node Protected
  S - Strict                    L - Loose
  A - ABR
  s - Soft Preemption
=====
LSP Name       : LSP-p2mp-1
S2L LSP ID    : 11264
P2MP ID       : 0                S2L Grp Id       : 2
Admin State   : Up              Oper State      : Up
S2L State:    : Active          :
S2L Name      : empty
To            : 192.0.2.7
S2L Admin     : Up              S2L Oper       : Up
OutInterface  : 1/1/c1/1:1000   Out Label      : 524287
S2L Up Time   : 0d 00:02:04    S2L Dn Time    : 0d 00:00:00
RetryAttempt  : 0              NextRetryIn    : 0 sec
S2L Trans     : 1              CSPF Queries   : 1
Failure Code  : noError        Failure Node    : n/a
Inter-area    : False
ExplicitHops  :
  No Hops Specified
Actual Hops   :
  192.168.13.1(192.0.2.1) @    Record Label   : N/A
  -> 192.168.13.2(192.0.2.3) @ Record Label   : 524287
  -> 192.168.35.2(192.0.2.5) @ Record Label   : 524287
  -> 192.168.57.2(192.0.2.7)  Record Label   : 524287
ComputedHops  :
  192.168.13.1(S)
  -> 192.168.13.2(S)
  -> 192.168.35.2(S)
  -> 192.168.57.2(S)
LastResignal  : n/a
=====

```

More in detail, **show router mpls bypass-tunnel detail** can be used. The list of actual hops provides the explicit hops of the bypass tunnel used to avoid the direct link between PE-1 and PE-3. On node PE-1, the MPLS path from PE-1 to PE-3 via PE-2 is followed (see [Figure 105: P2MP LSP LSP-p2mp-1 with bypass tunnels](#)).

```

[/]
A:admin@PE-1# show router mpls bypass-tunnel detail
=====
MPLS Bypass Tunnels (Detail)
=====
-----
bypass-link192.168.13.2-61441
-----
To           : 192.168.23.2      State           : Up
Out I/F      : 1/1/c2/1:1000    Out Label      : 524285
Up Time      : 0d 00:03:35      Active Time    : n/a
Reserved BW  : 0 Kbps          Protected LSP Count : 1
Type         : P2mp            Bypass Path Cost : 2
Setup Priority : 7              Hold Priority    : 0
Class Type   : 0
Exclude Node : None            Inter-Area      : False
Computed Hops :
  192.168.12.1(S)              Egress Admin Groups : None
  -> 192.168.12.2(S)           Egress Admin Groups : None
  -> 192.168.23.2(S)           Egress Admin Groups : None

```

```

Actual Hops      :
  192.168.12.1(192.0.2.1)      Record Label      : N/A
-> 192.168.12.2(192.0.2.2)      Record Label      : 524285
-> 192.168.23.2(192.0.2.3)      Record Label      : 524284

-----
---snip---

```

On node PE-3, the MPLS path from PE-3 to PE-5 via PE-2 and PE-4 is followed (see [Figure 105: P2MP LSP LSP-p2mp-1 with bypass tunnels](#)) to avoid the direct link between PE-3 and PE-5.

```

[/]
A:admin@PE-3# show router mpls bypass-tunnel protected-lsp p2mp detail

=====
MPLS Bypass Tunnels (Detail)
=====
-----
bypass-link192.168.35.2-61441
-----
To          : 192.168.45.2      State          : Up
Out I/F     : 1/1/c3/1:1000    Out Label     : 524286
Up Time    : 0d 00:04:09      Active Time   : n/a
Reserved BW : 0 Kbps          Protected LSP Count : 1
Type       : P2mp             Bypass Path Cost : 3
Setup Priority : 7             Hold Priority   : 0
Class Type  : 0
Exclude Node : None           Inter-Area     : False
Computed Hops :
  192.168.23.2(S)             Egress Admin Groups : None
-> 192.168.23.1(S)             Egress Admin Groups : None
-> 192.168.24.2(S)             Egress Admin Groups : None
-> 192.168.45.2(S)             Egress Admin Groups : None
Actual Hops :
  192.168.23.2(192.0.2.3)      Record Label     : N/A
-> 192.168.23.1(192.0.2.2)      Record Label     : 524286
-> 192.168.24.2(192.0.2.4)      Record Label     : 524285
-> 192.168.45.2(192.0.2.5)      Record Label     : 524285

Protected LSPs -
LSP Name    : LSP-p2mp-1::empty
From        : 192.0.2.1        To              : 192.0.2.7
Avoid Node/Hop : 192.168.35.2  Downstream Label : 524287
Bandwidth   : 0 Kbps

=====

```

A similar output can be seen on PE-5 node also. To avoid the direct link from PE-5 to PE-7, the MPLS path from PE-5 to PE-7 via PE-4 and PE-6 is followed, as shown in [Figure 105: P2MP LSP LSP-p2mp-1 with bypass tunnels](#).

On the transit LSRs and egress LER/leaf node (see [Figure 105: P2MP LSP LSP-p2mp-1 with bypass tunnels](#)), the **show router mpls p2mp-info** command can be used. See the **show** command on node PE-3 for the S2L path to 192.0.2.7. Similar outputs are possible for nodes PE-5 and PE-7.

```

[/]
A:admin@PE-3# show router mpls p2mp-info ?

p2mp-info [type <keyword>] [s2l-endpoint <ipv4 address>]

s2l-endpoint      - [a.b.c.d]

```

```

type                - keywords

[/]
A:admin@PE-3# show router mpls p2mp-info

=====
MPLS P2MP Cross Connect Information
=====
-----
S2L:LSP-p2mp-1::empty
-----
Source IP Address   : 192.0.2.1           Tunnel ID   : 1
P2MP ID             : 0                 Lsp ID     : 11264
To                  : 192.0.2.7
Out Interface       : 1/1/c1/1:1000       Out Label  : 524287
Num. of S2ls       : 1
-----
P2MP Cross-connect instances : 1
=====

```

## Mapping multicast traffic

To map multicast traffic into the LSP tree from the head-end node until leaf node, PIM and Internet Group Management Protocol (IGMP) configurations are needed on the head-end node (PE-1) and leaf nodes (PE-6 and PE-7) of the P2MP RSVP LSP. The intermediate nodes (transit LSR or branch LSR) do not need any explicit configuration for that.

## Head-end node (ingress LER) PE-1

PIM must be enabled on the interface toward the multicast source, as follows:

```

# on PE-1:
configure {
  router "Base" {
    pim {
      interface "int-PE-1-MC-source" {
      }
    }
  }
}

```

A tunnel interface is an internal representation of a specific P2MP LSP. On head-end PE-1, this tunnel interface is configured with keyword **rsvp-p2mp-root**, as follows:

```

# on PE-1:
configure {
  router "Base" {
    tunnel-interface {
      rsvp-p2mp-root "LSP-p2mp-1" {
      }
    }
  }
}

```

It is not possible to show PIM tunnel interfaces when PIM is not enabled. The following error is raised when a tunnel interface is configured without enabling PIM:

```

[ex:/configure router "Base" tunnel-interface]
A:admin@PE-1# show router pim tunnel-interface

```

```
MINOR: CLI #2005: Error while processing command - PIM is not configured
```

For multicast packets received on an interface to pass through the data plane, a successful reverse path forwarding (RPF) check must be done on the source address, otherwise the packet is dropped.

Besides enabling PIM on the tunnel interface, also IGMP is enabled to do a static <S,G> or <\*,G> join of a multicast group address (227.1.1.1 in the example) to the tunnel interface for the P2MP LSP. There is always a one-to-one mapping between <S,G> or <\*,G> and the tunnel interface for the P2MP LSP. In the example an <S,G> is configured. A <\*,G> join scenario is included in the [Additional topics](#) section.

```
# on PE-1:
configure {
  router "Base" {
    igmp {
      tunnel-interface {
        rsvp-p2mp-root "LSP-p2mp-1" {
          static {
            group 227.1.1.1 {
              source 192.168.100.2 { }
            }
          }
        }
      }
    }
  }
}
```

The **show router pim tunnel-interface** command shows the administrative and operational state of the tunnel interface and an association to an internal local interface index (73728 in the example).

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

=====
PIM Interfaces ipv4
=====
Interface                               Originator Address  Adm  Opr  Transport Type
-----
mpls-if-73728                            N/A                 Up   Up   Tx-IPMSI
-----
Interfaces : 1
=====
```

The **show router igmp group <group-address>** command provides the configured <S,G> entry:

```
[/]
A:admin@PE-1# show router igmp group 227.1.1.1

=====
IGMP Interface Groups
=====
(192.168.100.2,227.1.1.1)                UpTime: 0d 00:00:17
  Fwd List :
-----
Entries : 1
=====
IGMP Host Groups
=====
No Matching Entries
=====
IGMP SAP Groups
=====
No Matching Entries
=====
```

```
IGMP SLA Profile Instance Groups
```

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

The **show router pim group <group-address> detail** command on PE-1 shows that multicast traffic is sent from source 192.168.100.2 via incoming interface int-PE-1-MC-source to outgoing tunnel interface mpls-if-73728, as follows:

```
[/]
A:admin@PE-1# show router pim group 227.1.1.1 detail

=====
PIM Source Group ipv4
=====
Group Address       : 227.1.1.1
Source Address      : 192.168.100.2
RP Address          : 0
Advt Router        : 192.0.2.1
Flags              :                               Type           : (S,G)
Mode               : sparse
MRIB Next Hop      : 192.168.100.2
MRIB Src Flags     : direct
Keepalive Timer    : Not Running
Up Time            : 0d 00:03:12          Resolved By         : rtable-u

Up JP State        : Joined              Up JP Expiry        : 0d 00:00:00
Up JP Rpt         : Not Joined StarG    Up JP Rpt Override  : 0d 00:00:00

Register State    : No Info
Reg From Anycast RP: No

Rpf Neighbor      : 192.168.100.2
Incoming Intf     : int-PE-1-MC-source
Outgoing Intf List : mpls-if-73728

Curr Fwding Rate  : 9729.280 kbps
Forwarded Packets : 7344                Discarded Packets   : 0
Forwarded Octets  : 10149408           RPF Mismatches     : 0
Spt threshold     : 0 kbps              ECMP opt threshold : 7
Admin bandwidth   : 1 kbps

-----
Groups : 1
=====
```

## Leaf node (egress LER)

On the leaf nodes, PIM must be enabled:

```
# on PE-7:
configure {
  router "Base" {
    pim {
      admin-state enable
    }
  }
}
```

The following command creates the tunnel interface with keyword **rsvp-p2mp-leaf**. An explicit reference to the head-end system address, using the **sender-address <systemIP\_head-end\_node>** parameter is needed.

```
# on PE-7:
configure {
  router "Base" {
    tunnel-interface {
      rsvp-p2mp-leaf "LSP-p2mp-1" sender-address 192.0.2.1 {
    }
  }
}
```

The **show router pim tunnel-interface** command provides the administrative and operational state of the tunnel interface and an association to an internal local interface index (73728 in this example, by coincidence the same interface index as the one on the head-end node PE-1).

```
[/]
A:admin@PE-7# show router pim tunnel-interface

=====
PIM Interfaces ipv4
=====
Interface                               Originator Address  Adm  Opr  Transport Type
-----
mpls-if-73728                            N/A                 Up   Up   Tx-IPMSI
-----
Interfaces : 1
=====
```

The main goal on the leaf nodes is to get traffic off the P2MP LSP tunnel interface. This is done using a multicast information policy (**multicast-info-policy**). Inside this multicast information policy, a range of multicast group addresses is defined under a **bundle** context (*bundle1*) in order to see traffic (**channel**). Also inside the bundle context, the P2MP LSP is presented by the tunnel interface (**primary-tunnel-interface**). The following configures the multicast information policy on leaf node PE-7:

```
# on PE-7:
configure {
  multicast-management {
    multicast-info-policy "p2mp-pol" {
      bundle "bundle1" {
        primary-tunnel-interface {
          rsvp-p2mp "LSP-p2mp-1"
          sender 192.0.2.1
        }
        channel start 227.1.1.1 end 227.1.1.1 {
        }
      }
    }
  }
}
```



**Note:**

The **channel** command is a range command with a start-mc-group-address and an end-mc-group-address. In this example, only one multicast group address, 227.1.1.1, is seen.

The configured multicast information policy must be applied to the base router instance:

```
# on PE-7:
configure {
  router "Base" {
    multicast-info-policy "p2mp-pol"
```



On the leaf nodes PE-7 and PE-6, multicast clients are connected. IGMP is enabled on those multicast clients with a static <S,G> join to redirect multicast traffic downstream to the multicast client. This is configured as follows:

```
# on PE-7:
configure {
  router "Base" {
    igmp {
      interface "int-PE-7-MC-client1" {
        static {
          group 227.1.1.1 {
            source 192.168.100.2 { }
          }
        }
      }
    }
  }
}
```

The **show router igmp group** provides the configured <S,G> entry and outgoing interface "int-PE-7-MC-client1".

```
[/]
A:admin@PE-7# show router igmp group 227.1.1.1
=====
IGMP Interface Groups
=====
(192.168.100.2,227.1.1.1)                               UpTime: 0d 00:00:03
  Fwd List  : int-PE-7-MC-client1
-----
Entries : 1
=====
IGMP Host Groups
=====
No Matching Entries
=====
IGMP SAP Groups
=====
No Matching Entries
=====
IGMP SLA Profile Instance Groups
=====
No Matching Entries
=====
```

Now, users can verify if multicast traffic is sent to the multicast client using the **show router pim group group-address detail** command

```
[/]
A:admin@PE-7# show router pim group 227.1.1.1 detail
=====
PIM Source Group ipv4
=====
Group Address      : 227.1.1.1
Source Address     : 192.168.100.2
RP Address         : 0
Advt Router       :
Flags              :
Mode               : sparse
MRIB Next Hop     :
MRIB Src Flags    : remote
Keepalive Timer   : Not Running
Type               : (S,G)
```

```

Up Time           : 0d 00:00:10           Resolved By      : unresolved
Up JP State      : Joined                Up JP Expiry     : 0d 00:00:49
Up JP Rpt       : Not Joined StarG      Up JP Rpt Override : 0d 00:00:00

Register State   : No Info
Reg From Anycast RP: No

Rpf Neighbor     :
Incoming Intf    : mpls-if-73728
Outgoing Intf List : int-PE-7-MC-client1

Curr Fwding Rate : 9729.280 kbps
Forwarded Packets : 9057                Discarded Packets : 0
Forwarded Octets  : 12516774           RPF Mismatches    : 0
Spt threshold    : 0 kbps              ECMP opt threshold : 7
Admin bandwidth  : 1 kbps

-----
Groups : 1
=====

```

## OAM tools



**Note:** For SR OS Release 24.10.R2, the OAM tools are in classic CLI.

P2P LSP Operation, Administration, and Maintenance (OAM) commands (**oam lsp-ping** and **oam lsp-trace**) are extended for P2MP LSP. The user can instruct the head-end node to generate an P2MP LSP ping or a P2MP LSP trace by entering the command **oam p2mp-lsp-ping** or **oam p2mp-lsp-trace**. The P2MP OAM extensions are defined in *draft-ietf-mpls-p2mp-lsp-ping*.

For P2MP LSP ping, the echo request is sent on the active P2MP instance and replicated in the data path over all branches of the P2MP LSP instance. By default, all egress LER nodes which are leaves of the P2MP LSP instance reply. Echo reply messages can be reduced by configuring the **s2l-dest-address** (a maximum of five egress nodes in a single run of the OAM command). Replies are sent by IP.

```

<in classic CLI>
*A:PE-1# oam p2mp-lsp-ping ?
  - p2mp-lsp-ping <lsp-name> [p2mp-instance <instance-name> [s2l-dest-address <ipv4-address>
    [... up to 5]]]
    [ttl <label-ttl>]
  - p2mp-lsp-ping ldp <p2mp-identifier> [vpn-recursive-fec] [sender-addr <ipv4-address>] [leaf-
addr
  <ipv4-address> [... up to 5]]
  - p2mp-lsp-ping ldp-ssm source <ip-address> group <ip-address> [router <router-instance>]
service-name
  <service-name>] [sender-addr <ipv4-address>] [leaf-addr <ipv4-address> [... up to 5]]
  - p2mp-lsp-ping p2mp-policy root-address <ipv4-address> root-tree-id <tree-id> instance-id
<instance-id>
  [sender-addr <ipv4-address>] [leaf-addr <ipv4-address> [... up to 5]]
  - options common to all p2mp-lsp-ping cases: [detail] [fc <fc-name> [profile {in|out}]]
[size <octets>]
[timeout <timeout>]

<lsp-name>           : [64 chars max]
<instance-name>     : [32 chars max]
<in|out>            : in|out - Default: out
<fc-name>           : be|l2|af|l1|h2|ef|h1|nc - Default: be
<octets>            : [1..9786] - Default: 1

```

```

<label-ttl>      : [1..255] hops - Default: 255
<timeout>       : [1..120] seconds - Default: 10
<detail>        : keyword - displays detailed information
<p2mp-identifier> : [1..4294967295]
<ldp-ssm>       : keyword - Label Distribution Protocol, Source-Specific Multicast
<ip-address>    : ipv4-address - a.b.c.d
                  ipv6-address - x:x:x:x:x:x:x:x (eight 16-bit pieces)
                  x:x:x:x:x:x:d.d.d.d
                  x - [0..FFFF]H
                  d - [0..255]D

<ipv4-address>  : a.b.c.d
<router-instance> : <router-name>|<vprn-svc-id>
                  router-name - "Base" Default - Base
                  vprn-svc-id - [1..2147483647]

<service-name>  : [64 chars max]
<vpn-recursive-fec> : keyword - add a VPN Recursive FEC element to the launched packet
                  (useful for pinging a
                  VPN BGP inter-AS Option B leaf)
<p2mp-policy>   : keyword - ping a P2MP segment routing policy
<tree-id>       : [8193..16286]
<instance-id>   : [1..4294967295]
    
```

```

<in classic CLI>
*A:PE-1# oam p2mp-lsp-ping "LSP-p2mp-1" detail
P2MP LSP LSP-p2mp-1: 92 bytes MPLS payload

=====
S2L Information
=====
From          RTT          Return Code
-----
192.0.2.6     =3.79ms     EgressRtr(3)
192.0.2.7     =3.88ms     EgressRtr(3)
=====

Total S2L configured/up/responded = 2/2/2,
round-trip min/avg/max = 3.79 / 3.83 / 3.88 ms

Responses based on return code:
EgressRtr(3)=2
    
```

Return codes are based on RFC 4379. Value 3 means the replying router is an egress for the FEC at stack depth.

P2MP LSP trace allows the user to trace the path of a single S2L path of a P2MP LSP from head-end node to leaf node. Using the downstream mapping TLV, each node along the S2L path can fill in the appropriate flags: B or E flag. The B-flag is set when the responding node is a branch LSR and the E-flag is set when the responding node is an egress LER.

```

<in classic CLI>
*A:PE-1# oam p2mp-lsp-trace ?
- p2mp-lsp-trace <lsp-name> p2mp-instance <instance-name> s2l-dest-address <ip-address> [fc
<fc-name> [profile
  {in|out}]] [size <octets>] [max-fail <no-response-count>] [probe-count <probes-per-hop>]
[min-ttl
  <min-label-ttl>] [max-ttl <max-label-ttl>] [timeout <timeout>] [interval <interval>]
[detail]

<lsp-name>      : [64 chars max]
<instance-name> : [32 chars max]
<ip-address>    : ipv4 address a.b.c.d
<fc-name>       : be|l2|af|l1|h2|ef|h1|nc - Default: be
    
```

```
<in|out>          : in|out - Default: out
<octets>          : [1..9786] - Default: 1
<no-response-count> : [1..10] - Default: 5
<probes-per-hop>   : [1..10] - Default: 1
<min-label-ttl>    : [1..255] hops - Default: 1
<max-label-ttl>    : [1..255] hops - Default: 30
<timeout>         : [1..60] seconds - Default: 3
<detail>          : keyword - displays detailed information
<interval>        : [1..10] seconds - Default: 1
```

```
<in classic CLI>
*A:PE-1# oam p2mp-lsp-trace "LSP-p2mp-1" p2mp-instance "p-LSP-p2mp-1" s2l-dest-address
192.0.2.7 detail
P2MP LSP LSP-p2mp-1: 132 bytes MPLS payload
P2MP Instance p-LSP-p2mp-1, S2L Egress 192.0.2.7

 1 192.0.2.3 rtt=1.48 ms rc=8(DSRtrMatchLabel)
   DS 1: ipaddr=192.168.35.2 ifaddr=192.168.35.2 iftype=ipv4Numbered MRU=8982 label=524287
   proto=4(RSVP-TE) B/E flags:0/0
 2 192.0.2.5 rtt=2.33 ms rc=8(DSRtrMatchLabel)
   DS 1: ipaddr=192.168.57.2 ifaddr=192.168.57.2 iftype=ipv4Numbered MRU=8982 label=524287
   proto=4(RSVP-TE) B/E flags:0/0
 3 192.0.2.7 rtt=3.13 ms rc=3(EgressRtr)
```

Return codes (**rc**) are based on RFC 4279. Value 8 means that the label is switched at stack depth. This is the case for a transit LSR doing MPLS label swapping. No B or E flag is set.

## Additional topics

### <\*,G> IGMP join instead of <S,G> IGMP join

In the [Head-end node \(ingress LER\) PE-1](#) and [Leaf node \(egress LER\)](#) steps, a source specific IGMP join (<S,G> join) is used at the head-end node and leaf nodes. Another possibility is to use a source unknown or starg IGMP join (<\*,G> join). When doing the latter, a rendezvous point (RP) must be defined in the PIM network. The RP allows multicast data flows between sources and receivers to meet at a predefined network location (in this example, the loopback address of node PE-1). It must be seen as an intermediate device to establish a multicast flow.

The RP can be defined in a dynamic way (bootstrap router (BSR) protocol) or a static way. In this example, the static way is chosen meaning that on all involved PIM nodes, the RP address is statically configured. The following configuration is needed on head-end and leaf nodes.

```
# on PE-1, PE-6, PE-7:
configure {
  router "Base" {
    pim {
      rp {
        ipv4 {
          static {
            address 192.0.2.1 {
              group-prefix 227.1.1.1/32 { }
            }
          }
        }
      }
    }
  }
}
```

The **group-prefix** is a mandatory keyword that references a group address or group address range for which this rendezvous point is used.

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

=====
PIM RP Set ipv4
=====
Group Address          RP Address          Type      Prio  Hold Expiry
-----
227.1.1.1/32          192.0.2.1          Static    1     N/A  N/A
-----
Group Prefixes : 1
=====
```

As previously mentioned, the configuration of the <\*,G> IGMP join is done on the head-end node (PE-1) and leaf nodes (PE-6 and PE-7)

```
# on PE-1:
configure {
  router "Base" {
    igmp {
      tunnel-interface {
        rsvp-p2mp-root "LSP-p2mp-1" {
          static {
            group 227.1.1.1 {
              starg
            }
          }
        }
      }
    }
  }
}
```

```
# on PE-6:
configure {
  router "Base" {
    igmp {
      interface "int-PE-6-MC-client2" {
        static {
          group 227.1.1.1 {
            starg
          }
        }
      }
    }
  }
}
```

```
# on PE-7:
configure {
  router "Base" {
    igmp {
      interface "int-PE-7-MC-client1" {
        static {
          group 227.1.1.1 {
            starg
          }
        }
      }
    }
  }
}
```

The same preceding **show** commands can be used to verify the multicast traffic on head-end node and leaf nodes, **show router igmp group 227.1.1.1** and **show router pim group 227.1.1.1 detail**.

```
[/]
A:admin@PE-7# show router igmp group 227.1.1.1
=====
IGMP Interface Groups
=====
(*,227.1.1.1)                               UpTime: 0d 00:06:58
  Fwd List  : int-PE-7-MC-client1
-----
Entries : 1
=====
IGMP Host Groups
=====
No Matching Entries
=====
IGMP SAP Groups
=====
No Matching Entries
=====
IGMP SLA Profile Instance Groups
=====
No Matching Entries
=====
```

```
[/]
A:admin@PE-7# show router pim group 227.1.1.1 detail
=====
PIM Source Group ipv4
=====
Group Address      : 227.1.1.1
Source Address     : *
RP Address         : 192.0.2.1
Advt Router       :
Flags              :                               Type           : (*,G)
Mode               : sparse
MRIB Next Hop     :
MRIB Src Flags    : remote
Keepalive Timer   : Not Running
Up Time           : 0d 07:48:39      Resolved By         : unresolved

Up JP State       : Joined           Up JP Expiry        : 0d 00:00:52
Up JP Rpt         : Not Joined StarG  Up JP Rpt Override  : 0d 00:00:00

Rpf Neighbor      :
Incoming Intf     : mpls-if-73728
Outgoing Intf List : int-PE-7-MC-client1

Curr Fwding Rate  : 0.000 kbps
Forwarded Packets : 9                Discarded Packets   : 0
Forwarded Octets  : 12438           RPF Mismatches      : 0
Spt threshold     : 0 kbps           ECMP opt threshold  : 7
Admin bandwidth   : 1 kbps

=====
PIM Source Group ipv4
=====
Group Address      : 227.1.1.1
Source Address     : 192.168.100.2
RP Address         : 192.0.2.1
```

```

Advt Router      :
Flags           : spt                Type           : (S,G)
Mode           : sparse
MRIB Next Hop   :
MRIB Src Flags  : remote
Keepalive Timer Exp: 0d 00:03:27
Up Time        : 0d 00:00:04         Resolved By    : unresolved

Up JP State     : Joined             Up JP Expiry   : 0d 00:00:55
Up JP Rpt      : Not Pruned         Up JP Rpt Override : 0d 00:00:00

Register State  : No Info
Reg From Anycast RP: No

Rpf Neighbor    :
Incoming Intf   : mpls-if-73728
Outgoing Intf List : int-PE-7-MC-client1

Curr Fwding Rate : 9729.280 kbps
Forwarded Packets : 3838             Discarded Packets : 0
Forwarded Octets  : 5304116         RPF Mismatches    : 0
Spt threshold    : 0 kbps           ECMP opt threshold : 7
Admin bandwidth  : 1 kbps

-----
Groups : 2
=====

```

## Influence IGP metric

The IGP metric is increased on all links pointing to/from PE-2 and on the link between PE-5 and PE-7.

```

# on PE-1:
configure {
  router "Base" {
    ospf {
      area 0.0.0.0 {
        interface "int-PE-1-PE-2" {
          metric 10000
        }
      }
    }
  }
}

```

```

# on PE-2:
configure {
  router "Base" {
    ospf 0 {
      area 0.0.0.0 {
        interface "int-PE-2-PE-1" {
          metric 10000
        }
        interface "int-PE-2-PE-3" {
          metric 10000
        }
        interface "int-PE-2-PE-4" {
          metric 10000
        }
      }
    }
  }
}

```

```

# on PE-3:
configure {
  router "Base" {
    ospf 0 {
      area 0.0.0.0
    }
  }
}

```

```

interface "int-PE-3-PE-2"
  metric 10000

# on PE-4:
configure {
  router "Base" {
    ospf 0 {
      area 0.0.0.0 {
        interface "int-PE-4-PE-2" {
          metric 10000
        }
      }
    }
  }
}

# on PE-5:
configure {
  router "Base" {
    ospf 0 {
      area 0.0.0.0 {
        interface "int-PE-5-PE-7" {
          metric 10000
        }
      }
    }
  }
}

# on PE-7:
configure {
  router "Base" {
    ospf 0 {
      area 0.0.0.0 {
        interface "int-PE-7-PE-5" {
          metric 10000
        }
      }
    }
  }
}

```

The existing P2MP LSP *LSP-p2mp-1* does not take into account these new constraints. The two S2L paths (one *empty* toward PE-6 and another *empty* toward PE-7) are calculated using the default OSPF metric. To trigger MPLS to recompute the S2L paths, configure a *p2mp-resignal-timer* on the head-end node inside the global MPLS context. Each time this timer expires (in the example, every 60 minutes), MPLS triggers CSPF to recompute the whole set of S2L paths of all active P2MP instances. MPLS performs a global Make-Before-Break (MBB) and moves each S2L sub-LSP in the instance into its new path using a new P2MP LSP ID if the global MBB is successful. **show router mpls status** gives an indication when the P2MP resignal timer will expire and which types of LSPs are set up on the node.

```

# on PE-1:
configure {
  router "Base" {
    mpls {
      p2mp-resignal-timer 60      # in minutes
    }
  }
}

[/]
A:admin@PE-1# show router mpls status

=====
MPLS Status
=====
Admin Status           : Up
Oper(V4) State         : Up      Oper(V6) State           : Down
IPv4 Oper Down Reason  : n/a
IPv6 Oper Down Reason  : ipv6TeRtrDown
FRR Object             : Enabled  Resignal Timer           : Disabled
Hold Timer             : 1 seconds Next Resignal           : N/A
Srlg Frr               : Disabled Srlg Frr Strict          : Disabled
Admin Group Frr        : Disabled
Dynamic Bypass         : Enabled  User Srlg Database       : Disabled

```



```

BypassResignalTimer      : Disabled   BypassNextResignal      : N/A
LeastFill Min Thd        : 5 percent  LeastFill Reopti Thd    : 10 percent
Local TTL Prop           : Enabled     Transit TTL Prop        : Enabled
P2mp TTL Prop           : Enabled
AB Sample Multiplier     : 1         AB Adjust Multiplier    : 288
Exp Backoff Retry        : Disabled    CSPF On Loose Hop       : Disabled
Lsp Init RetryTimeout    : 30 seconds MBB Pref Current Hops   : Disabled
Logger Event Bundling    : Disabled
Retry on IGP Overload    : Disabled    Resignal on IGP Overload : Disabled
Resignal on IGP Event    : Disabled
StrictEroNhopDirectRes  : Disabled

P2mp Resignal Timer      : 60 minutes  P2mp Next Resignal      : 59 minutes
Sec FastRetryTimer       : Disabled    Static LSP FR Timer     : 30 seconds
P2P Max Bypass Association: 1000
Max Bypass PLR Association: 16
P2PActPathFastRetry     : Disabled    P2MP S2L Fast Retry     : Disabled
In Maintenance Mode     : No
MplsTp                   : Disabled
Next Available Lsp Index : 2
Entropy Label RSVP-TE   : Enabled     Entropy Label SR-TE     : Enabled
PCE Report RSVP-TE      : Disabled    PCE Report SR-TE        : Disabled
PCE Init LSP            : Disabled
PCC Oper Status RSVP-TE : Down        PCE Oper Status RSVP-TE : Down
PCC Oper Status SR-TE   : Down        PCE Oper Status SR-TE   : Down
SR-TE Resignal Timer    : Disabled    SR-TE Next Resignal     : N/A
SR-TE Resig on IGP Event : Disabled    SR-TE Resig on IGP Overlo*: Disabled
LSP BSID Block          : N/A
LSP History              : Disabled
LSP Self Ping Timeout   : 300 seconds LSP Self Ping Interval  : 1 seconds
RSVP-TE LSP Self Ping   : Disabled    Self Ping Timeout Action : retry
RSVP-TE Tunnel Table Pref : 7         SR-TE Tunnel Table Pref : 8
    
```

=====

MPLS LSP Count

=====

	Originate	Transit	Terminate
Static LSPs	0	0	0
Dynamic LSPs	2	0	0
P2P LSPs	0	N/A	N/A
Detour LSPs	0	0	0
P2MP S2Ls	2	0	0
MPLS-TP LSPs	0	0	0
Mesh-P2P LSPs	0	N/A	N/A
One Hop-P2P LSPs	0	N/A	N/A
SR-TE LSPs	0	N/A	N/A
Mesh-P2P SR-TE LSPs	0	N/A	N/A
One Hop-P2P SR-TE LSPs	0	N/A	N/A
PCE Init SR-TE LSPs	0	N/A	N/A
On-Demand SR-TE LSPs	0	N/A	N/A

=====

As an alternative, users can also perform a manual resignal of a P2MP instance on the head-end node using the following **tools** command:

```

# on PE-1:
tools
  perform
    router "Base"
      mpls
    
```

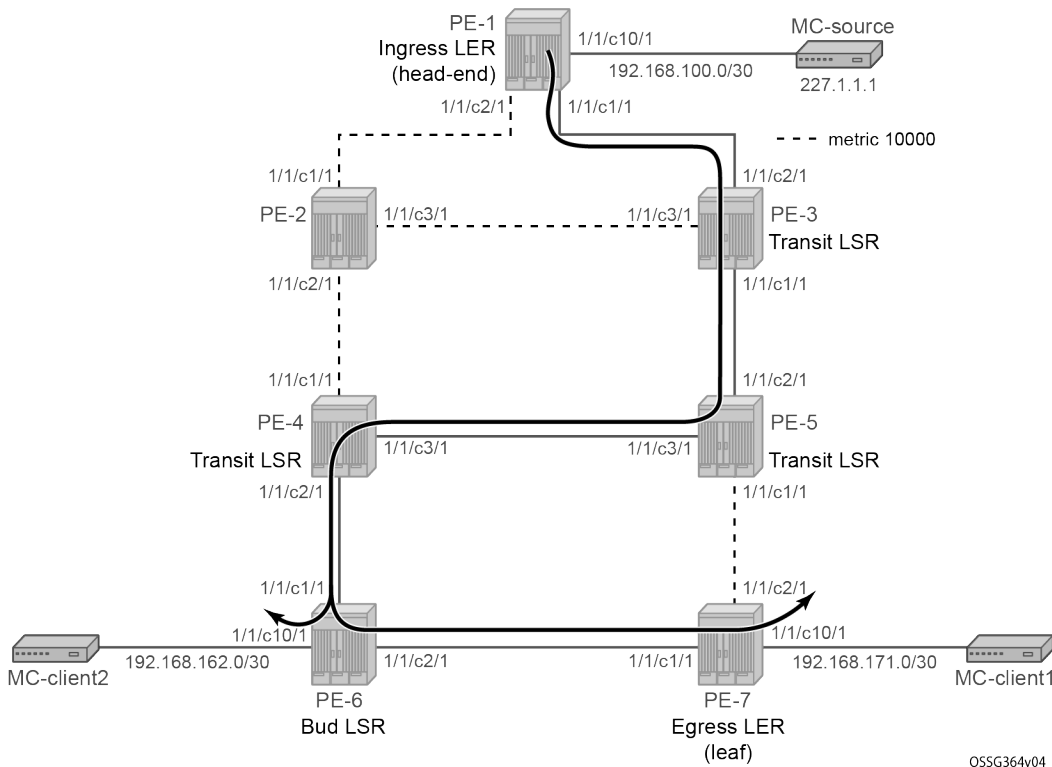
```

resignal p2mp-lsp "LSP-p2mp-1" p2mp-instance "p-LSP-p2mp-1"

# on PE-1:
tools
  perform
    router "Base"
      mpls
        resignal p2mp-delay 0
  
```

Figure 106: P2MP LSP p-to-mp-1 with metric change shows the resigaled S2L paths. Node PE-6 is now a bud LSR node (instead of egress LER before).

Figure 106: P2MP LSP p-to-mp-1 with metric change



The following command shows the new actual and computed hops in the resigaled S2L path to PE-7:

```

[/]
A:admin@PE-1# show router mpls p2mp-lsp "LSP-p2mp-1" p2mp-instance p2mp-instance-name "p-LSP-p2mp-1" s2l s2l-name "empty" to 192. detail
=====
MPLS LSP LSP-p2mp-1 S2L empty (Detail)
=====
Legend :
@ - Detour Available          # - Detour In Use
b - Bandwidth Protected      n - Node Protected
S - Strict                   L - Loose
A - ABR
s - Soft Preemption
=====
LSP Name      : LSP-p2mp-1
  
```

```

S2L LSP ID      : 11268
P2MP ID        : 0
Admin State    : Up
S2L State:     : Active
S2L Name       : empty
To             : 192.0.2.7
S2L Admin      : Up
OutInterface   : 1/1/c1/1:1000
S2L Up Time    : 0d 00:18:01
RetryAttempt   : 0
S2L Trans      : 3
Failure Code   : noError
Inter-area     : False
ExplicitHops   :
  No Hops Specified
Actual Hops    :
  192.168.13.1(192.0.2.1) @
-> 192.168.13.2(192.0.2.3) @
-> 192.168.35.2(192.0.2.5) @
-> 192.168.45.1(192.0.2.4) @
-> 192.168.46.2(192.0.2.6) @
-> 192.168.67.2(192.0.2.7) @
ComputedHops   :
  192.168.13.1(S)
-> 192.168.13.2(S)
-> 192.168.35.2(S)
-> 192.168.45.1(S)
-> 192.168.46.2(S)
-> 192.168.67.2(S)
LastResignal   : n/a
=====

```

The following command shows the new actual and computed hops in the resigaled S2L path to PE-6:

```

[/]
A:admin@PE-1# show router mpls p2mp-lsp "LSP-p2mp-1" p2mp-instance p2mp-instance-name "p-LSP-
p2mp-1" s2l s2l-name "empty" to 192. detail

=====
MPLS LSP LSP-p2mp-1 S2L empty (Detail)
=====
Legend :
  @ - Detour Available          # - Detour In Use
  b - Bandwidth Protected      n - Node Protected
  S - Strict                   L - Loose
  A - ABR
  s - Soft Preemption

=====
LSP Name       : LSP-p2mp-1
S2L LSP ID     : 11268
P2MP ID        : 0
Admin State    : Up
S2L State:     : Active
S2L Name       : empty
To             : 192.0.2.6
S2L Admin      : Up
OutInterface   : 1/1/c1/1:1000
S2L Up Time    : 0d 00:18:01
RetryAttempt   : 0
S2L Trans      : 3
Failure Code   : noError
Inter-area     : False
ExplicitHops   :
S2L Grp Id     : 2
Oper State     : Up
S2L Oper       : Up
Out Label      : 524282
S2L Dn Time    : 0d 00:00:00
NextRetryIn    : 0 sec
CSPF Queries   : 3
Failure Node   : n/a

```

```

    No Hops Specified
Actual Hops      :
  192.168.13.1(192.0.2.1) @           Record Label      : N/A
-> 192.168.13.2(192.0.2.3) @           Record Label      : 524282
-> 192.168.35.2(192.0.2.5) @           Record Label      : 524281
-> 192.168.45.1(192.0.2.4) @           Record Label      : 524280
-> 192.168.46.2(192.0.2.6) @           Record Label      : 524282
ComputedHops    :
  192.168.13.1(S)
-> 192.168.13.2(S)
-> 192.168.35.2(S)
-> 192.168.45.1(S)
-> 192.168.46.2(S)
LastResignal    : n/a
=====

```

An **oam p2mp-lsp-trace** command (in classic CLI) toward PE-7 now sets the E flag on PE-6 because PE-6 acts also as an egress LER node.

```

<in classic CLI>
*A:PE-1# /oam p2mp-lsp-trace LSP-p2mp-1 p2mp-instance p-LSP-p2mp-1 s2l-dest-address 192.0.2.7
detail
P2MP LSP LSP-p2mp-1: 132 bytes MPLS payload
P2MP Instance p-LSP-p2mp-1, S2L Egress 192.0.2.7

  1 192.0.2.3 rtt=2.56 ms rc=8(DSRtrMatchLabel)
    DS 1: ipaddr=192.168.35.2 ifaddr=192.168.35.2 iftype=ipv4Numbered MRU=8982 label=524281
    proto=4(RSVP-TE) B/E flags:0/0
  2 192.0.2.5 rtt=3.91 ms rc=8(DSRtrMatchLabel)
    DS 1: ipaddr=192.168.45.1 ifaddr=192.168.45.1 iftype=ipv4Numbered MRU=8982 label=524280
    proto=4(RSVP-TE) B/E flags:0/0
  3 192.0.2.4 rtt=4.31 ms rc=8(DSRtrMatchLabel)
    DS 1: ipaddr=192.168.46.2 ifaddr=192.168.46.2 iftype=ipv4Numbered MRU=8982 label=524282
    proto=4(RSVP-TE) B/E flags:0/0
  4 192.0.2.6 rtt=4.83 ms rc=8(DSRtrMatchLabel)
    DS 1: ipaddr=192.168.67.2 ifaddr=192.168.67.2 iftype=ipv4Numbered MRU=8982 label=524281
    proto=4(RSVP-TE) B/E flags:0/1
  5 192.0.2.7 rtt=5.80 ms rc=3(EgressRtr)

INFO: CLI #2052: Switching to the MD-CLI engine

```

In the next step, the S2L path toward PE-7 is changed from an empty path to a strict direct MPLS path (*path-strict-to-PE-7*). In that way, OSPF is not calculating the shortest path to the leaf node anymore.

```

# on PE-1:
configure {
  router "Base" {
    mpls {
      path "path-strict-to-PE-7" {
        admin-state enable
        hop 10 {
          ip-address 192.168.13.2
          type strict
        }
        hop 20 {
          ip-address 192.168.35.2
          type strict
        }
        hop 30 {
          ip-address 192.168.57.2
          type strict
        }
      }
    }
  }
}

```

```
}

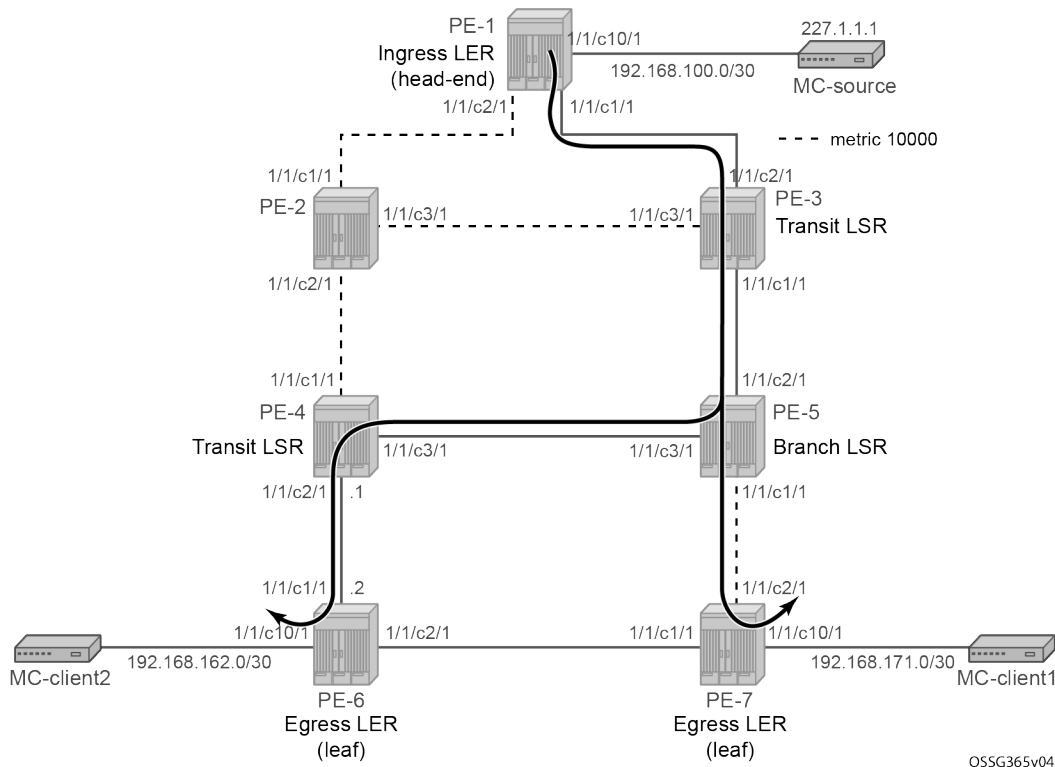
```

Before applying this new S2L path to the existing P2MP LSP (*LSP-p2mp-1*), the existing S2L path toward PE-7 is removed.

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      lsp "LSP-p2mp-1" {
        admin-state enable
        type p2mp-rsvp
        from 192.0.2.1
        path-computation-method local-cspf
        fast-reroute {
        }
      }
      primary-p2mp-instance "p-LSP-p2mp-1" {
        s2l-path "empty" to 192.0.2.6 {
        }
        delete s2l-path "empty" to 192.0.2.7
        s2l-path "path-strict-to-PE-7" to 192.0.2.7 {
        }
      }
    }
  }
}
```

As a consequence of this, only the S2L group ID changes while S2L LSP ID remains the same. [Figure 107: P2MP LSP LSP-p2mp-1 with strict S2L path toward PE-7](#) shows the P2MP LSP LSP-p2mp-1 with strict S2L path to leaf PE-7.

Figure 107: P2MP LSP LSP-p2mp-1 with strict S2L path toward PE-7



OSSG365v04

S2L paths can be verified according to [Figure 107: P2MP LSP LSP-p2mp-1 with strict S2L path toward PE-7](#). PE-5 is now a branch LSR node (instead of a transit LSR before).

```
[/]
A:admin@PE-1# show router mpls p2mp-lsp "LSP-p2mp-1" p2mp-instance p2mp-instance-name "p-LSP-p2mp-1" s2l s2l-name "path-strict-to-PE-7" to 192.0.2.7 detail

=====
MPLS LSP LSP-p2mp-1 S2L path-strict-to-PE-7 (Detail)
=====
Legend :
@ - Detour Available          # - Detour In Use
b - Bandwidth Protected      n - Node Protected
S - Strict                    L - Loose
A - ABR
s - Soft Preemption

=====
LSP Name       : LSP-p2mp-1
S2L LSP ID    : 11268
P2MP ID       : 0
Admin State   : Up
S2L State     : Active
S2L Name      : path-strict-to-PE-7
To            : 192.0.2.7
S2L Admin     : Up
OutInterface  : 1/1/c1/1:1000
S2L Up Time   : 0d 00:01:06
RetryAttempt  : 0
S2L Trans     : 1
Failure Code  : noError
Inter-area    : False
ExplicitHops  :
                192.168.13.2(S)
                -> 192.168.35.2(S)
                -> 192.168.57.2(S)
Actual Hops   :
                192.168.13.1(192.0.2.1) @
                -> 192.168.13.2(192.0.2.3) @
                -> 192.168.35.2(192.0.2.5) @
                -> 192.168.57.2(192.0.2.7)
ComputedHops  :
                192.168.13.1(S)
                -> 192.168.13.2(S)
                -> 192.168.35.2(S)
                -> 192.168.57.2(S)
LastResignal  : n/a
=====
```

An **oam p2mp-lsp-trace** command (in classic CLI) toward PE-7 now sets the B flag on PE-5 because PE-5 acts as a branch LSR now.

```
[/]
A:admin@PE-1# //oam p2mp-lsp-trace "LSP-p2mp-1" p2mp-instance "p-LSP-p2mp-1" s2l-dest-address 192.0.2.7 detail
INFO: CLI #2051: Switching to the classic CLI engine
INFO: CLI #2050: Classic CLI modification of the configuration is not allowed - 'model-driven' management interface configuration mode active
*A:PE-1# /oam p2mp-lsp-trace "LSP-p2mp-1" p2mp-instance "p-LSP-p2mp-1" s2l-dest-address 192.0.2.7 detail
P2MP LSP LSP-p2mp-1: 132 bytes MPLS payload
P2MP Instance p-LSP-p2mp-1, S2L Egress 192.0.2.7
```

```

1 192.0.2.3 rtt=2.39 ms rc=8(DSRtrMatchLabel)
  DS 1: ipaddr=192.168.35.2 ifaddr=192.168.35.2 iftype=ipv4Numbered MRU=8982 label=524281
  proto=4(RSVP-TE) B/E flags:0/0
2 192.0.2.5 rtt=3.38 ms rc=8(DSRtrMatchLabel)
  DS 1: ipaddr=192.168.57.2 ifaddr=192.168.57.2 iftype=ipv4Numbered MRU=8982 label=524287
  proto=4(RSVP-TE) B/E flags:1/0
3 192.0.2.7 rtt=5.15 ms rc=3(EgressRtr)

```

INFO: CLI #2052: Switching to the MD-CLI engine

## Intelligent remerge

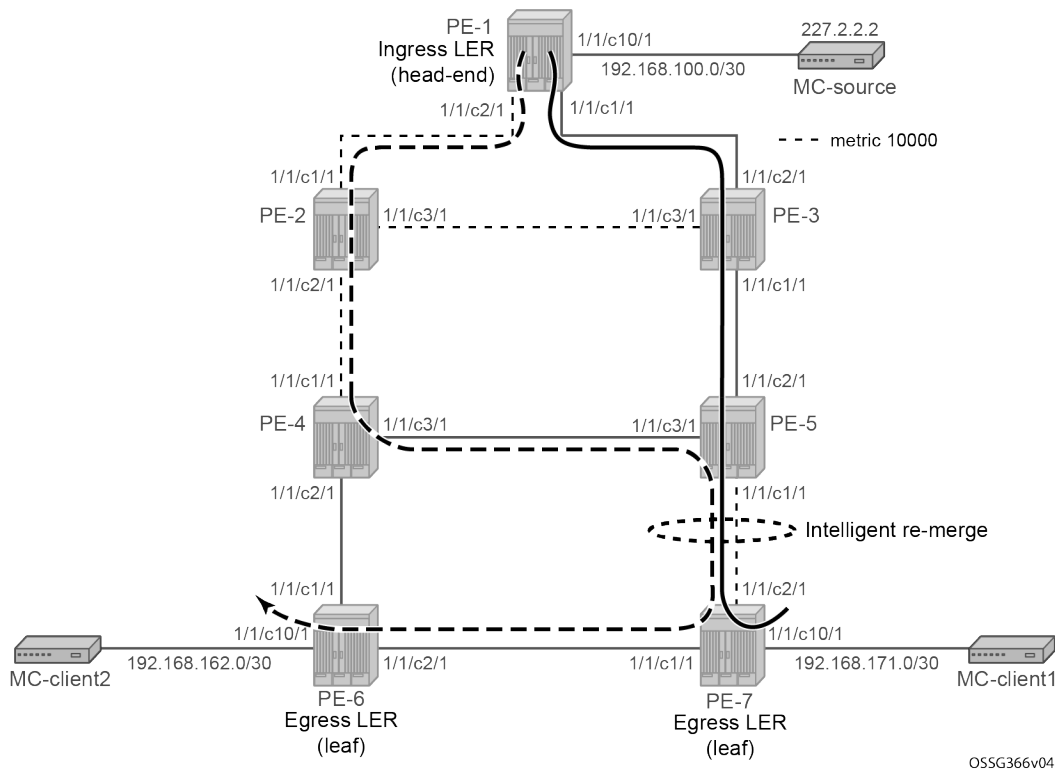
Intelligent remerge protects users from receiving duplicate multicast traffic during convergence. It also protects against duplicate traffic in case of badly designed S2L paths. Three cases are described for which intelligent remerge is implemented.

### Intelligent remerge case 1 - Badly designed S2L paths

In this example, the paths of two different S2Ls of the same P2MP LSP instance have ingress label maps (ILMs) on different ports but go out on the same Next-Hop Label Forwarding Entry (NHLFE).

[Figure 108: Intelligent remerge, case 1](#) shows P2MP LSP *LSP-p2mp-2* with two incoming S2L paths at PE-5.

Figure 108: Intelligent remerge, case 1



OSSG366v04

On the head-end node (PE-1), P2MP LSP *LSP-p2mp-2* is created with two strict direct MPLS paths (*path-strict-to-PE-7* and *path-strict-to-PE-6*), as follows. Intelligent remerge is performed at node PE-5.

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      path "path-strict-to-PE-6" {
        admin-state enable
        hop 10 {
          ip-address 192.168.12.2
          type strict
        }
        hop 20 {
          ip-address 192.168.24.2
          type strict
        }
        hop 30 {
          ip-address 192.168.45.2
          type strict
        }
        hop 40 {
          ip-address 192.168.57.2
          type strict
        }
        hop 50 {
          ip-address 192.168.67.1
          type strict
        }
      }
      path "path-strict-to-PE-7" {
        admin-state enable
        hop 10 {
          ip-address 192.168.13.2
          type strict
        }
        hop 20 {
          ip-address 192.168.35.2
          type strict
        }
        hop 30 {
          ip-address 192.168.57.2
          type strict
        }
      }
    }
    lsp "LSP-p2mp-2" {
      admin-state enable
      type p2mp-rsvp
      from 192.0.2.1
      primary-p2mp-instance "p-LSP-p2mp-2" {
        s2l-path "path-strict-to-PE-6" to 192.0.2.6 {
        }
        s2l-path "path-strict-to-PE-7" to 192.0.2.7 {
        }
      }
    }
  }
}
```

```
[/]
A:admin@PE-1# show router mpls p2mp-lsp "LSP-p2mp-2" p2mp-instance p2mp-instance-name "p-LSP-p2mp-2"
```

```
=====
MPLS P2MP Instance (Originating)
```



```

=====
-----
Type : Originating
-----
LSP Name      : LSP-p2mp-2
P2MP ID       : 0
Adm State     : Up
LSP Tunnel ID : 2
Oper State    : Up

P2MPInstance  : p-LSP-p2mp-2
P2MP Inst Id  : 2
Inst Admin    : Up
Inst Up Time  : 0d 00:00:03
Hop Limit     : 255
Record Route  : Record
Setup Priority : 7
Include Grps  :
None
Bandwidth     : No Reservation
S2L-Name      : path-strict-to-PE-7
S2L Admin    : Up
S2L-Name      : path-strict-to-PE-6
S2L Admin    : Up

P2MP-Inst-type : Primary
P2MP Lsp Id    : 30720
Inst Oper      : Up
Inst Dn Time   : 0d 00:00:00
Adaptive       : Enabled
Record Label   : Record
Hold Priority   : 0
Exclude Grps   :
None
Oper Bw        : 0 Mbps
To             : 192.0.2.7
S2L Oper      : Up
To             : 192.0.2.6
S2L Oper      : Up
-----
P2MP instances : 1
=====

```

To verify that node PE-5 is not sending duplicate multicast traffic downstream toward PE-7 while it receives two incoming multicast streams, a new tunnel interface and a new static <S,G> IGMP join are configured on head-end node (PE-1) and leaf nodes (PE-6 and PE-7). Also on the leaf nodes, an extension to the existing multicast information policy is needed. The configuration is as follows:

```

# on PE-1:
configure {
  router "Base" {
    tunnel-interface {
      rsvp-p2mp-root "LSP-p2mp-2" { # PIM already enabled on PE-1
      }
    }
  }
  igmp {
    tunnel-interface {
      rsvp-p2mp-root "LSP-p2mp-2" {
        static {
          group 227.2.2.2 {
            source 192.168.100.2 { }
          }
        }
      }
    }
  }
}

```

```

# on PE-6:
configure {
  multicast-management {
    multicast-info-policy "p2mp-pol" {
      bundle "bundle2" {
        primary-tunnel-interface {
          rsvp-p2mp "LSP-p2mp-2"
          sender 192.0.2.1
        }
      }
    }
  }
  channel start 227.2.2.2 end 227.2.2.2 {
  }
}

```

```

    }
  }
}
router "Base" {
  tunnel-interface {
    rsvp-p2mp-leaf "LSP-p2mp-2" sender-address 192.0.2.1 { # PIM enabled
    }
  }
  igmp {
    interface "int-PE-6-MC-client2" {
      static {
        group 227.2.2.2 {
          source 192.168.100.2 { }
        }
      }
    }
  }
}

```

```

# on PE-7:
configure {
  multicast-management {
    multicast-info-policy "p2mp-pol" {
      bundle "bundle2" {
        primary-tunnel-interface {
          rsvp-p2mp "LSP-p2mp-2"
          sender 192.0.2.1
        }
        channel start 227.2.2.2 end 227.2.2.2 {
        }
      }
    }
  }
}
router "Base" {
  tunnel-interface {
    rsvp-p2mp-leaf "LSP-p2mp-2" sender-address 192.0.2.1 { # PIM enabled
    }
  }
  igmp {
    interface "int-PE-7-MC-client1" {
      static {
        group 227.2.2.2 {
          source 192.168.100.2 { }
        }
      }
    }
  }
}

```

For verification of incoming/outgoing multicast traffic at node PE-5, the **monitor** command is used.

```

[/]
A:admin@PE-5# monitor port 1/1/c1/1 port-string 1/1/c2/1 port-string-2 1/1/c3/1 rate interval 3
repeat 10

```

```

=====
Monitor statistics for Ports
=====

```

```

-----
Input                                     Output
-----

```

```

---snip---
-----
At time t = 12 sec (Mode: Rate)
-----
Port 1/1/c1/1
-----
Octets          140          1239113
Packets         1           881
Errors          0           0
Bits           1120        9912904
Utilization (% of port capacity)  ~0.00      0.01

Port 1/1/c2/1
-----
Octets          1239582        140
Packets         881           1
Errors          0           0
Bits           9916656        1120
Utilization (% of port capacity)  0.01      ~0.00

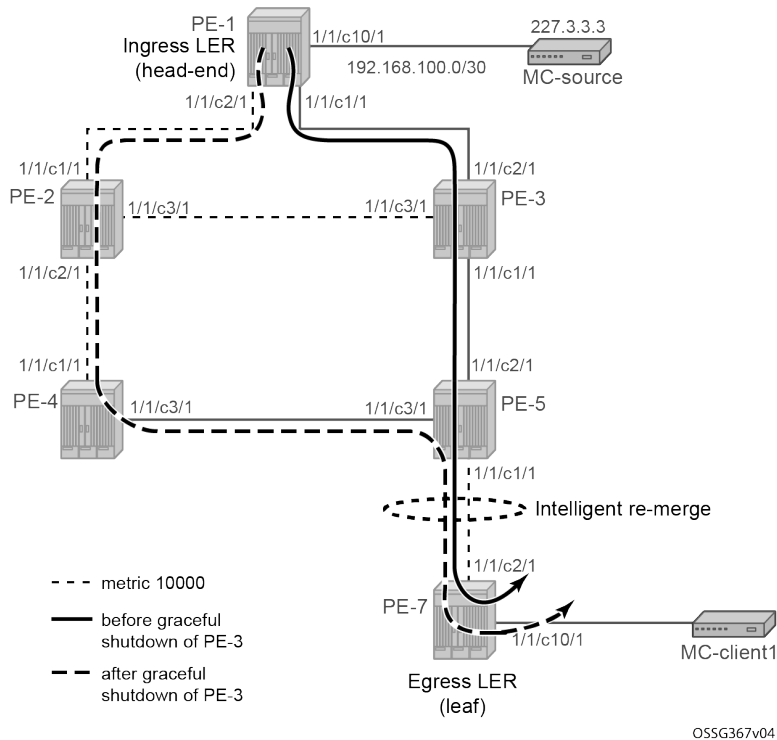
Port 1/1/c3/1
-----
Octets          1239643        110
Packets         881           1
Errors          0           0
Bits           9917144        880
Utilization (% of port capacity)  0.01      ~0.00
-----
---snip---
    
```

Two incoming multicast streams are seen at PE-5 node (*port 1/1/c2/1* and *port 1/1/c3/1*) and only one outgoing multicast stream (*port 1/1/c1/1*) is sent. No traffic duplication is seen.

### Intelligent remerge case 2 - Make-before-break after graceful shutdown

[Figure 109: Intelligent remerge, case 2](#) shows two paths of the same S2L that have Ingress Label Map (ILM) entries on different incoming ports and go out on the same NHLFE. This is the case during a make-before-break (MBB) operation on an S2L path after a graceful shutdown or global revertive. This is only a temporary situation until the original path is torn down.

Figure 109: Intelligent remerge, case 2



For this test, only multicast client MC-client1 is used (the one connected to leaf node PE-7). On nodes PE-4 and PE-7, the port to PE-6 is disabled to isolate PE-6. On the head-end node PE-1, a new P2MP LSP *LSP-p2mp-3* is created with an empty MPLS path and with CSPF enabled. The metric type is TE to ensure that CSPF uses the TE metric instead of the IGP metric. In this case, the TE metric has the same value on all MPLS interfaces and the path has the hops PE-1, PE-3, PE-5, and PE-7. Also in this case, intelligent remerge is performed at node PE-5.

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      path "empty" {
        admin-state enable
      }
      lsp "LSP-p2mp-3" {
        admin-state enable
        type p2mp-rsvp
        from 192.0.2.1
        path-computation-method local-cspf
        metric-type te
        primary-p2mp-instance "p-LSP-p2mp-3" {
          s2l-path "empty" to 192.0.2.7 {
          }
        }
      }
    }
  }
}
```

```
[/]
A:admin@PE-1# show router mpls p2mp-lsp "LSP-p2mp-3" p2mp-instance p2mp-instance-name "p-LSP-p2mp-3" s2l s2l-name "empty" to 192.0.2.7 detail
```

```

=====
MPLS LSP LSP-p2mp-3 S2L empty (Detail)
=====
Legend :
  @ - Detour Available          # - Detour In Use
  b - Bandwidth Protected      n - Node Protected
  S - Strict                    L - Loose
  A - ABR
  s - Soft Preemption
=====
LSP Name       : LSP-p2mp-3
S2L LSP ID    : 37376
P2MP ID       : 0
Admin State   : Up
S2L State:    : Active
S2L Name      : empty
To            : 192.0.2.7
S2L Admin     : Up
OutInterface  : 1/1/c1/1:1000
S2L Up Time   : 0d 00:00:10
RetryAttempt  : 0
S2L Trans     : 1
Failure Code  : noError
Inter-area    : False
ExplicitHops  :
  No Hops Specified
Actual Hops   :
  192.168.13.1(192.0.2.1)
  -> 192.168.13.2(192.0.2.3)
  -> 192.168.35.2(192.0.2.5)
  -> 192.168.57.2(192.0.2.7)
ComputedHops :
  192.168.13.1(S)
  -> 192.168.13.2(S)
  -> 192.168.35.2(S)
  -> 192.168.57.2(S)
LastResignal  : n/a
=====
S2L Grp Id    : 1
Oper State    : Up
S2L Oper     : Up
Out Label     : 524286
S2L Dn Time  : 0d 00:00:00
NextRetryIn  : 0 sec
CSPF Queries : 1
Failure Node  : n/a
Record Label  : N/A
Record Label  : 524286
Record Label  : 524283
Record Label  : 524282

```

In a normal situation, the P2MP LSP follows the nodes PE-1, PE-3, PE-5, and PE-7. This can be verified with multicast traffic. Therefore, a new tunnel interface and a new static <S,G> IGMP join are configured on head-end node PE-1 and leaf node PE-7. On the leaf node, an extension to the existing multicast information policy is needed. The configuration is as follows:

```

# on PE-1:
configure {
  router "Base" {
    tunnel-interface {
      rsvp-p2mp-root "LSP-p2mp-3" {
      }
    }
  }
  igmp {
    tunnel-interface {
      rsvp-p2mp-root "LSP-p2mp-3" {
        static {
          group 227.3.3.3 {
            source 192.168.100.2 { }
          }
        }
      }
    }
  }
}

```



```

Utilization (% of port capacity)          0.01          ~0.00
-----
Port 1/1/c3/1
-----
Octets                21                21
Packets               0                0
Errors                0                0
Bits                  168               168
Utilization (% of port capacity)        ~0.00          ~0.00
-----

```

An RSVP graceful shutdown is performed on node PE-3, as follows:

```

# on PE-3:
configure {
  router "Base" {
    rsvp {
      graceful-shutdown true
    }
  }
}

```

The graceful shutdown on PE-3 triggers a global revertive on head-end node PE-1. A new MPLS path is calculated (see the dashed line in [Figure 109: Intelligent remerge, case 2](#)). For a few seconds or even less than a second, the old path and new path are active (two incoming multicast streams on node PE-5). Node PE-5 is doing intelligent remerge, not sending duplicate multicast traffic downstream toward PE-7:

```

[/]
A:admin@PE-5# monitor port 1/1/c1/1 port-string 1/1/c2/1 port-string-2 1/1/c3/1 rate interval 3
repeat 20

=====
Monitor statistics for Ports
=====
-----
Input                Output
-----
---snip---
-----
At time t = 6 sec (Mode: Rate)
-----
Port 1/1/c1/1
-----
Octets                83                1239610
Packets               1                 881
Errors                0                 0
Bits                  664               9916880
Utilization (% of port capacity)  ~0.00             0.01

Port 1/1/c2/1
-----
Octets                1239531           89
Packets               881               1
Errors                0                 0
Bits                  9916248           712
Utilization (% of port capacity)  0.01             ~0.00

Port 1/1/c3/1
-----
Octets                218371            119
Packets               156               1
Errors                0                 0

```

```

Bits                               1746968                952
Utilization (% of port capacity)    ~0.00                  ~0.00
-----
---snip---

```

The granularity of the monitoring command is 3 seconds. The graceful shutdown takes less than 3 seconds. However, it is clear that the number of outgoing packets on port 1/1/c1/1 equals the number of incoming packets on port 1/1/c2/1. A number of these incoming packets also arrived on port 1/1/c3/1, but no duplicate packets were sent on the outgoing port. No traffic duplication is seen.

```

<continued>
-----
At time t = 9 sec (Mode: Rate)
-----
Port 1/1/c1/1
-----
Octets                               63                1239257
Packets                              0                 881
Errors                               0                 0
Bits                                 504               9914056
Utilization (% of port capacity)    ~0.00              0.01

Port 1/1/c2/1
-----
Octets                               90799             63
Packets                              65                0
Errors                               0                 0
Bits                                 726392            504
Utilization (% of port capacity)    ~0.00             ~0.00

Port 1/1/c3/1
-----
Octets                               1239769           169
Packets                              881                0
Errors                               0                 0
Bits                                 9918152           1352
Utilization (% of port capacity)    0.01              ~0.00
-----
---snip---

```

Eventually, port 1/1/c3/1 takes over all traffic.

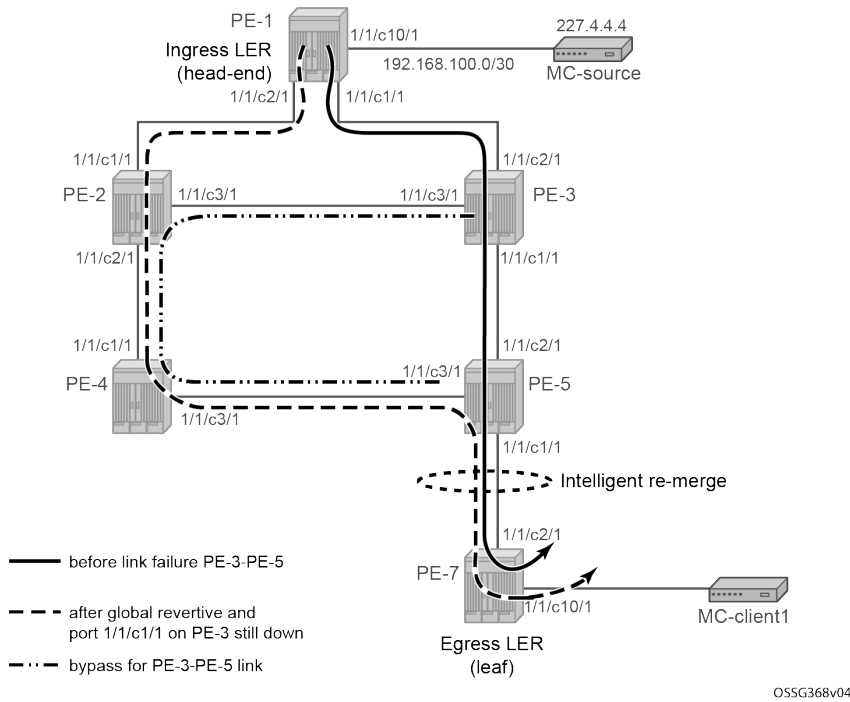
### Intelligent remerge case 3 - Bypass and new global revertive path

When a bypass is active on the S2L path and the new global revertive path of the same S2L arrives on the same incoming interface as the original path (interface flapped) at the FRR merge point node, the node performs intelligent remerge. The implementation recognizes this specific case and signals a different label from the original S2L path coming on that same interface.

**Figure 110: Intelligent remerge, case 3** shows the initial S2L path before the link failure (solid line), the bypass path to protect the link between PE-3 and PE-5, and the new global revertive path of the S2L (dotted line). Both the bypass path and the global revertive path share the links between PE-2 and PE-4 and between PE-4 and PE-5.



Figure 110: Intelligent remerge, case 3



For this test, all the non-default OSPF metrics are removed from the interfaces on all the PEs; on PE-1, as follows:

```
# on PE-1:
configure {
  router "Base" {
    ospf 0 {
      area 0.0.0.0 {
        interface "int-PE-1-PE-2" {
          delete metric
        }
      }
    }
  }
}
```

Only one MC-client is used (MC-client1 connected to leaf node PE-7). On nodes PE-4 and PE-7, the port toward PE-6 is disabled to isolate PE-6. On the head-end node PE-1, a new P2MP LSP "LSP-p2mp-4" is created with an empty MPLS path and FRR enabled. Also in this case, intelligent remerge is performed at node PE-5.

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      path "empty"
      admin-state enable
    }
    lsp "LSP-p2mp-4" {
      admin-state enable
      type p2mp-rsvp
      from 192.0.2.1
      path-computation-method local-cspf
      fast-reroute {
      }
      primary-p2mp-instance "p-LSP-p2mp-4" {
      }
    }
  }
}
```

```

        s2l-path "empty" to 192.0.2.7 {
        }
    }

[/]
A:admin@PE-1# show router mpls p2mp-lsp "LSP-p2mp-4" p2mp-instance p2mp-instance-name "p-LSP-
p2mp-4" s2l s2l-name "empty" to 192.0.2.7 detail

=====
MPLS LSP LSP-p2mp-4 S2L empty (Detail)
=====
Legend :
  @ - Detour Available          # - Detour In Use
  b - Bandwidth Protected      n - Node Protected
  S - Strict                    L - Loose
  A - ABR
  s - Soft Preemption

=====
LSP Name       : LSP-p2mp-4
S2L LSP ID     : 57344
P2MP ID        : 0
Admin State    : Up
S2L State      : Active
S2L Name       : empty
To             : 192.0.2.7
S2L Admin      : Up
OutInterface   : 1/1/c1/1:1000
S2L Up Time    : 0d 00:01:09
RetryAttempt   : 0
S2L Trans      : 1
Failure Code    : noError
Inter-area     : False
ExplicitHops   :
  No Hops Specified
Actual Hops    :
  192.168.13.1(192.0.2.1) @
  -> 192.168.13.2(192.0.2.3) @
  -> 192.168.35.2(192.0.2.5)
  -> 192.168.57.2(192.0.2.7)
ComputedHops   :
  192.168.13.1(S)
  -> 192.168.13.2(S)
  -> 192.168.35.2(S)
  -> 192.168.57.2(S)
LastResignal   : n/a

=====
S2L Grp Id     : 1
Oper State     : Up
S2L Oper       : Up
Out Label      : 524285
S2L Dn Time    : 0d 00:00:00
NextRetryIn   : 0 sec
CSPF Queries   : 1
Failure Node   : n/a
Record Label   : N/A
Record Label   : 524285
Record Label   : 524284
Record Label   : 524286

```

In the normal situation, the P2MP LSP follows the nodes PE-1, PE-3, PE-5, and PE-7. This can be verified with multicast traffic. Therefore, a new tunnel interface and a new static <S,G> IGMP join are configured on head-end node PE-1 and leaf node PE-7. On the leaf node, an extension to the existing multicast information policy is needed. This is configured as follows:

```

# on PE-1:
configure {
  router "Base" {
    tunnel-interface {
      rsvp-p2mp-root "LSP-p2mp-4" {
      }
    }
  }
  igmp {
    tunnel-interface {
      rsvp-p2mp-root "LSP-p2mp-4" {

```

```

        static {
            group 227.4.4.4 {
                source 192.168.100.2 { }
            }
        }
    }
}

# on PE-7:
configure {
    multicast-management {
        multicast-info-policy "p2mp-pol" {
            bundle "bundle4" {
                primary-tunnel-interface {
                    rsvp-p2mp "LSP-p2mp-4"
                    sender 192.0.2.1
                }
                channel start 227.4.4.4 end 227.4.4.4 {
                }
            }
        }
    }
}
router "Base" {
    tunnel-interface {
        rsvp-p2mp-leaf "LSP-p2mp-4" sender-address 192.0.2.1 {
        }
    }
}
igmp {
    interface "int-PE-7-MC-client1" {
        static {
            group 227.4.4.4 {
                source 192.168.100.2 { }
            }
        }
    }
}
}

```

The initial path goes from PE-1 via PE-3 and PE-5 to PE-7. The following command emulates a link failure on the interface from PE-3 to PE-5:

```

# on PE-3:
configure {
    port 1/1/c1/1 {
        admin-state disable
    }
}

```

As a consequence of this, traffic flows from PE-1 to PE-3 and then over the bypass link from PE-3 via PE-2 and PE-4 to PE-5, as shown in [Figure 110: Intelligent remerge, case 3](#). The following **show** command shows the '#' symbol for actual hop 192.168.13.2 and the failure code is tunnelLocallyRepaired:

```

[/]
A:admin@PE-1# show router mpls p2mp-lsp "LSP-p2mp-4" p2mp-instance p2mp-instance-name "p-LSP-p2mp-4" s2l s2l-name "empty" to 192.0.2.7 detail

=====
MPLS LSP LSP-p2mp-4 S2L empty (Detail)
=====
Legend :
  @ - Detour Available           # - Detour In Use
  b - Bandwidth Protected       n - Node Protected
  S - Strict                     L - Loose
  A - ABR
  s - Soft Preemption

=====
LSP Name          : LSP-p2mp-4

```

```

S2L LSP ID      : 57344
P2MP ID        : 0
Admin State    : Up
S2L State:     : Active
S2L Name      : empty
To            : 192.0.2.7
S2L Admin     : Up
OutInterface  : 1/1/c1/1:1000
S2L Up Time   : 0d 00:03:30
RetryAttempt  : 0
S2L Trans     : 1
Failure Code  : tunnelLocallyRepaired
Inter-area    : False
ExplicitHops  :
  No Hops Specified
Actual Hops    :
  192.168.13.1(192.0.2.1) @
-> 192.168.13.2(192.0.2.3) @ #
-> 192.168.35.2(192.0.2.5)
-> 192.168.57.2(192.0.2.7)
ComputedHops   :
  192.168.13.1(S)
-> 192.168.13.2(S)
-> 192.168.35.2(S)
-> 192.168.57.2(S)
LastResignal  : n/a
In Prog MBB   :
  MBB Type    : GlobalRevert
  Started At  : 02/11/2025 10:16:59
  FailureCode : noError
S2L Grp Id    : 1
Oper State    : Up
S2L Oper     : Up
Out Label     : 524285
S2L Dn Time   : 0d 00:00:00
NextRetryIn  : 0 sec
CSPF Queries  : 1
Failure Node  : 192.0.2.3
Record Label  : N/A
Record Label  : 524285
Record Label  : 524284
Record Label  : 524286
  
```

In the meantime, PE-3 triggers a global revertive action by sending a PathErr message toward the head-end node PE-1. After the global revert, the path goes from PE-1 via PE-2, PE-4, and PE-5 to PE-7.

```

[/]
A:admin@PE-1# show router mpls p2mp-lsp "LSP-p2mp-4" p2mp-instance p2mp-instance-name "p-LSP-
p2mp-4" s2l s2l-name empty to 192.0.2.7 detail

=====
MPLS LSP LSP-p2mp-4 S2L empty (Detail)
=====
Legend :
  @ - Detour Available          # - Detour In Use
  b - Bandwidth Protected      n - Node Protected
  S - Strict                    L - Loose
  A - ABR
  s - Soft Preemption

=====
LSP Name      : LSP-p2mp-4
S2L LSP ID    : 57344
P2MP ID      : 0
Admin State   : Up
S2L State:    : Active
S2L Name     : empty
To          : 192.0.2.7
S2L Admin    : Up
OutInterface : 1/1/c2/1:1000
S2L Up Time  : 0d 00:03:56
RetryAttempt  : 0
S2L Trans    : 2
Failure Code  : noError
Inter-area   : False
S2L Grp Id   : 2
Oper State   : Up
S2L Oper     : Up
Out Label    : 524287
S2L Dn Time  : 0d 00:00:00
NextRetryIn  : 0 sec
CSPF Queries : 2
Failure Node  : n/a
  
```

```

ExplicitHops      :
  No Hops Specified
Actual Hops       :
  192.168.12.1(192.0.2.1) @           Record Label      : N/A
-> 192.168.12.2(192.0.2.2)           Record Label      : 524287
-> 192.168.24.2(192.0.2.4)           Record Label      : 524287
-> 192.168.45.2(192.0.2.5)           Record Label      : 524286
-> 192.168.57.2(192.0.2.7)           Record Label      : 524286
ComputedHops      :
  192.168.12.1(S)
-> 192.168.12.2(S)
-> 192.168.24.2(S)
-> 192.168.45.2(S)
-> 192.168.57.2(S)
LastResignal     : n/a
Last MBB         :
MBB Type         : GlobalRevert          MBB State          : Success
Ended At         : 02/11/2025 10:17:33
=====
    
```

For the very short time that the bypass path is used together with the new calculated path, PE-5 receives two incoming MC streams (both arriving on port 1/1/c3/1): one from the bypass path (PE-3 => PE-2 => PE-4 => PE-5) and one from the new MPLS path (PE-1 => PE-2 => PE-4 => PE-5 => PE-7). Port 1/1/c1/1 on PE-5 performs intelligent remerge, so only one MC stream is sent downstream toward leaf node PE-7.

## Conclusion

From a configuration point of view, a P2MP LSP is only configured on the head-end node of that P2MP LSP; no explicit configuration is needed on the transit LSRs, branch LSRs, bud LSRs, and egress LERs/leaf nodes.

Because the PIM protocol is only needed on the head-end node and the leaf nodes, the core network can remain PIM-free. Although convergence is not covered in this chapter, failures in the core will be resolved by MPLS (in case of FRR, traffic loss for less than 50ms is expected), which is a major improvement compared to PIM convergence.

# Seamless MPLS: Isolated IGP/LDP Domains and Labeled BGP

This chapter provides information about Seamless MPLS: Isolated IGP/LDP domains and Labeled BGP.

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

## Applicability

This chapter is applicable to SR OS routers and was initially written for SR OS Release 13.0.R7. The CLI in the current edition is based on SR OS Release 23.3.R1.

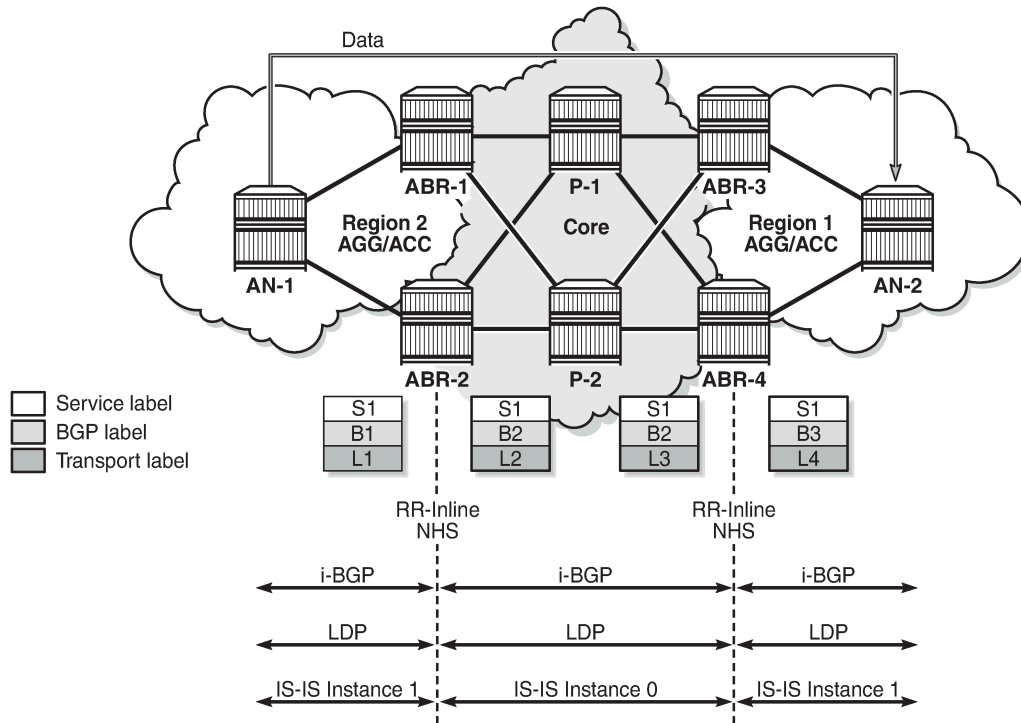
## Overview

Seamless Multi-Protocol Label Switching (MPLS) is a network architecture that extends MPLS networks to integrate access and aggregation networks into a single MPLS domain, to solve the scaling problems in flat MPLS-based deployments. The Seamless MPLS transport concept described in this chapter partitions the core, aggregation, and access networks into isolated IGP/LDP domains. Seamless MPLS does not define any new protocols or technologies and is based on existing and well-known ones. Seamless MPLS provides end-to-end service-independent transport, separating the service and transport plane. Therefore, it removes the need for service-specific configurations in network transport nodes. Service provisioning is restricted only at the points of the network where it is required.

When BGP is used to distribute a route, it can also distribute an MPLS label that is mapped to that route. The label mapping information is appended to the BGP update message that is used to distribute the route. This is described in RFC 3107, *Carrying Label Information in BGP-4*.

[Figure 111: Seamless MPLS - network topology, control and data plane](#) shows a network with a core area and regional areas. [Figure 111: Seamless MPLS - network topology, control and data plane](#) also shows the control plane used in this Seamless MPLS implementation. For simplification, the control plane is displayed from right to left and the data plane from left to right. In this example, LDP is used as the underlying transport inside each IGP domain. Alternatively, RSVP-TE could be used.

Figure 111: Seamless MPLS - network topology, control and data plane



25635

In typical Seamless MPLS solutions, multiple ABRs are in place that result in some specific BGP configurations to send/receive multiple paths, such as the add-path feature. Because of this, ANs and ABRs have several next hops for the same prefix, allowing the use of redundancy mechanisms such as BGP Prefix Independent Convergence (PIC) edge, also known as BGP Fast ReRoute (FRR). These mechanisms are beyond the scope of this chapter.

AN routers in a regional area learn the reachability of AN routers in other regional areas through BGP labeled routes redistributed by the local ABRs (RFC 3107).

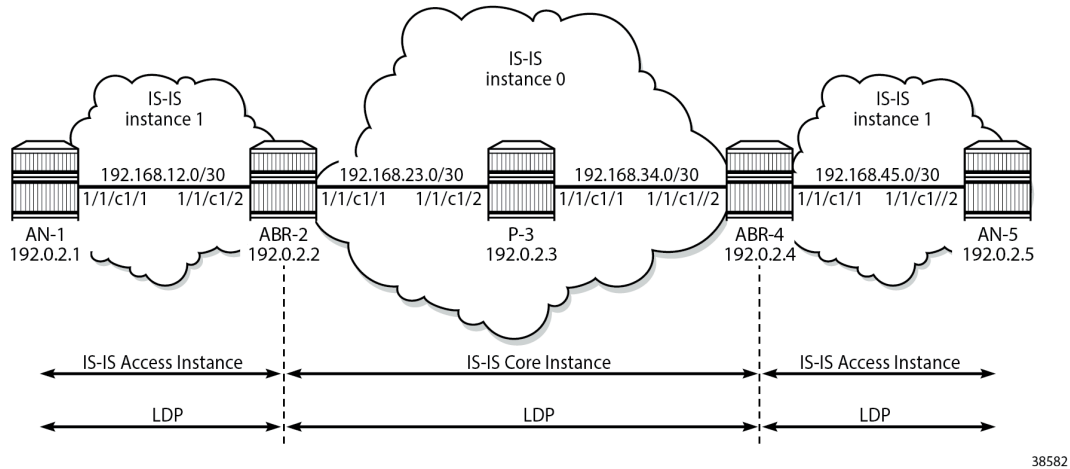
The label stack contains three labels for packets sent in a VPN service between the access nodes:

- The ANs push a service label to the packets sent in the VPN service. The service label remains unchanged end-to-end between ANs. The service label is popped by the remote AN and is the inner label of the label stack.
- The BGP label is the middle label of the label stack and should be regarded as a transport label. The transport label stack is increased to two labels: BGP and LDP transport labels. The BGP label is pushed by the iLER AN and is swapped at the BGP next hop, which can be one of the two local ABRs. Both ABRs are configured with next-hop-self. The BGP label is also swapped by the remote ABR.
- The iLER AN pushes an LDP transport label to the packets sent to the remote AN to reach the BGP next hop. At the local ABR, the LDP transport label is popped and a new LDP transport label is pushed to reach the BGP next hop (remote ABR). The LDP transport label is swapped in every label switching router (LSR) and popped by the ABR nearest to the remote AN. That ABR pops the LDP transport label, swaps the BGP label, and pushes an LDP transport label to reach the remote eLER AN.

## Configuration

Figure 112: Seamless MPLS - IGP/LDP domains shows the example topology that is used in this chapter. An Epipe and VPRN are established between the access nodes AN-1 and AN-5. In the regional areas, and in the core area, IS-IS L2 capability is used.

Figure 112: Seamless MPLS - IGP/LDP domains



38582

## Initial configuration

All nodes have the following initial configuration:

- Cards, media dependent adapters (MDAs), ports
  - Router interfaces:

```
# on ABR-2:
configure exclusive
router "Base" {
  interface "int-ABR-2-AN-1" {
    port 1/1/c1/2
    ipv4 {
      primary {
        address 192.168.12.2
        prefix-length 30
      }
    }
  }
  interface "int-ABR-2-P-3" {
    port 1/1/c1/1
    ipv4 {
      primary {
        address 192.168.23.1
        prefix-length 30
      }
    }
  }
}
interface "system" {
```



```

        ipv4 {
            primary {
                address 192.0.2.2
                prefix-length 32
            }
        }
    }
}

```

- IS-IS (alternatively, OSPF could be used). Core area and regional areas run an isolated IS-IS instance. ABRs run two IS-IS instances: instance 0 belongs to the core and instance 1 belongs to the access network.
  - **Core instance.** All ABRs and Ps have level 2 (L2) capability, as follows:

```

# on ABR-2:
configure exclusive
router "Base" {
    isis 0 {
        admin-state enable
        level-capability 2
        area-address [49.0001]
        interface "int-ABR-2-P-3" {
            interface-type point-to-point
        }
        interface "system" {
        }
    }
}

```

- **Access instance.** All ABRs and ANs have L2 capability, as follows:

```

# on ABR-2:
configure exclusive
router "Base" {
    isis 1 {
        admin-state enable
        level-capability 2
        interface "int-ABR-2-AN-1" {
            interface-type point-to-point
        }
        interface "system" {
        }
    }
}

```

- LDP

Link LDP is enabled on all router interfaces on all nodes, as follows:

```

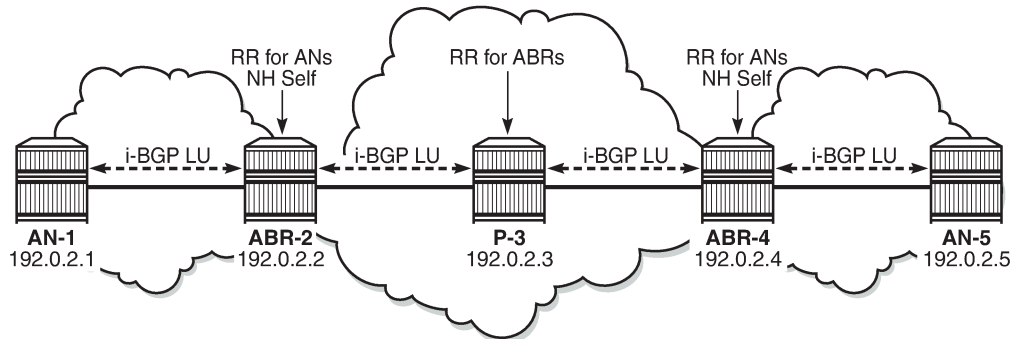
# on ABR-2:
configure exclusive
router "Base" {
    ldp {
        interface-parameters {
            interface "int-ABR-2-AN-1" {
                ipv4 {
                }
            }
            interface "int-ABR-2-P-3" {
                ipv4 {
                }
            }
        }
    }
}

```

## Configure BGP

BGP is configured on all ABRs and all ANs. P-3 acts as a core Route Reflector (RR). To allow for separation of core/access IGP domains, the ABRs become RRs inline and implement next-hop-self on labeled IPv4 BGP prefixes. [Figure 113: Seamless MPLS - BGP](#) shows the exchange of iBGP Labeled Unicast (LU) routes.

Figure 113: Seamless MPLS - BGP



25637

## BGP configuration on ABRs

There are two BGP groups on the ABRs: one group toward the core RR and another group toward the AN, as follows:

```
# on ABR-2:
configure exclusive
router "Base" {
  autonomous-system 64496
  bgp {
    ebgp-default-reject-policy {
      import false
      export false
    }
    group "core" {
      peer-as 64496
      advertise-inactive true
      family {
        vpn-ipv4 true
        label-ipv4 true
      }
    }
  }
  neighbor "192.0.2.3" {
    description "coreRR_P-3"
    group "core"
    next-hop-self true
  }
}
```

**Advertise-inactive** must be enabled on the BGP group toward the core. The /32 system IP addresses, learned in labeled BGP, are also learned in IS-IS. Because IS-IS has a lower preference compared to iBGP, the IS-IS routes are installed in the routing table. BGP default behavior only advertises those prefixes

that were elected by RTM and used. The VPN IPv4 address family is also included, along with labeled IPv4, to allow setting up L3 VPN services, as shown in next sections. The next-hop attribute of VPN IPv4 prefixes remains unchanged.

```
# on ABR-2:
configure exclusive
router "Base" {
  bgp {
    group "ANs_Label_IPv4+VPN_IPv4" {
      peer-as 64496
      family {
        vpn-ipv4 true
        label-ipv4 true
      }
      cluster {
        cluster-id 2.2.2.2
      }
    }
    neighbor "192.0.2.1" {
      description "AN-1"
      group "ANs_Label_IPv4+VPN_IPv4"
      next-hop-self true
    }
  }
}
```

## BGP configuration on the core RR

```
# on P-3:
configure exclusive
router "Base" {
  autonomous-system 64496
  bgp {
    ebgp-default-reject-policy {
      import false
      export false
    }
    group "core" {
      peer-as 64496
      advertise-inactive true
      family {
        vpn-ipv4 true
        label-ipv4 true
      }
      cluster {
        cluster-id 3.3.3.3
      }
    }
    neighbor "192.0.2.2" {
      description "ABR-2"
      group "core"
    }
    neighbor "192.0.2.4" {
      description "ABR-4"
      group "core"
    }
  }
}
```

## BGP configuration on ANs toward ABRs

```
# on AN-1:
configure exclusive
router "Base" {
    autonomous-system 64496
    bgp {
        ebgp-default-reject-policy {
            import false
            export false
        }
        group "ABRs_Label_IPv4+VPN_IPv4" {
            peer-as 64496
            family {
                vpn-ipv4 true
                label-ipv4 true
            }
        }
        neighbor "192.0.2.2" {
            group "ABRs_Label_IPv4+VPN_IPv4"
        }
    }
}
```

Configuring address family **label-ipv4** implies that all advertised IPv4 prefixes are sent to the remote BGP peer as an RFC 3107 formatted label. The **next-hop-self** command only applies to labeled IPv4 prefixes, not to VPN-IPv4.

The BGP sessions can be shown with the following command:

```
[/]
A:admin@P-3# show router bgp summary all

=====
BGP Summary
=====
Legend : D - Dynamic Neighbor
=====

Neighbor
Description
ServiceId          AS PktRcvd InQ  Up/Down  State|Rcv/Act/Sent (Addr Family)
                   PktSent OutQ
-----
192.0.2.2
ABR-2
Def. Inst          64496      6   0 00h00m32s 0/0/0 (VpnIPv4)
                   4   0          0/0/0 (Lbl-IPv4)
192.0.2.4
ABR-4
Def. Inst          64496      3   0 00h00m23s 0/0/0 (VpnIPv4)
                   3   0          0/0/0 (Lbl-IPv4)
-----
```

```
[/]
A:admin@AN-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	6	0	00h00m38s	0/0/0	(VpnIPv4)
			6	0		0/0/0	(Lbl-IPv4)
-----							

## Export policies for BGP

A policy is required on the ANs to advertise the system IP address in labeled BGP toward the ABRs. The same policy is required on the ABRs to advertise their system IP address in labeled BGP toward the core and the AN.

## Policy configuration on ANs and ABRs

```
# on AN-1 and ABR-2:
configure exclusive
  policy-options {
    prefix-list "system" {
      prefix 192.0.2.1/32 type exact {
      }
    }
  }
  policy-statement "export-system" {
    entry 10 {
      from {
        prefix-list ["system"]
        protocol {
          name [direct]
        }
      }
      action {
        action-type accept
      }
    }
  }
}
```

This export policy must be applied in the **bgp** context on AN-1: either in the general settings or per **group** or per **neighbor**, as follows:

```
# on AN-1:
configure exclusive
  router "Base" {
    bgp {
      group "ABRs_Label_IPv4+VPN_IPv4" {
        export {
          policy ["export-system"]
        }
      }
    }
  }
}
```

The same export policy is applied in the group "core" on ABR-2, as follows:

```
# on ABR-2:
```

```
configure exclusive
  router "Base" {
    bgp {
      group "core" {
        export {
          policy ["export-system"]
        }
      }
    }
  }
```

A similar export policy is defined to export prefix 192.0.2.5 from AN-5 to ABR-4 and from ABR-4 to the RR in the core network, P-3.

The prefix of the remote AN is added to the routing table in AN-1 and services can be configured in the ANs. No service configuration is required in the transit nodes.

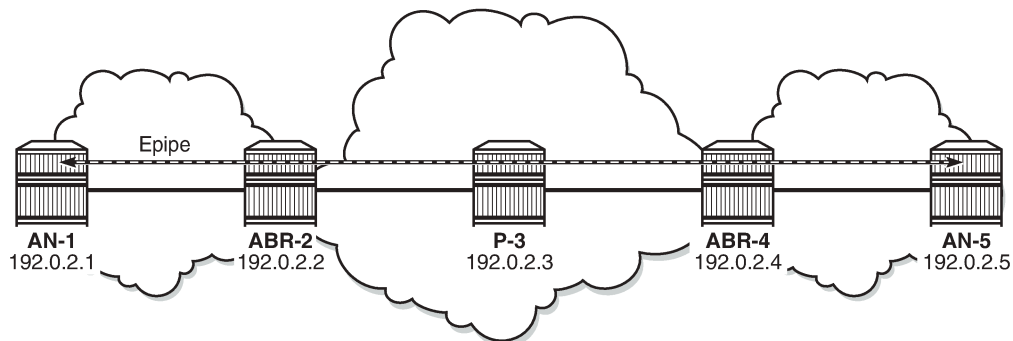
```
[/]
A:admin@AN-1# show router route-table

=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]          Type  Proto  Age           Pref
Next Hop[Interface Name]   Metric
-----
192.0.2.1/32                Local  Local   00h05m09s    0
  system                    0
192.0.2.2/32                Remote ISIS(1) 00h04m57s   18
  192.168.12.2              10
192.0.2.5/32              Remote BGP_LABEL 00h00m25s 170
  192.0.2.2 (tunneled)    10
192.168.12.0/30            Local  Local   00h05m09s    0
  int-AN-1-ABR-2            0
-----
No. of Routes: 4
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
=====
```

## Configure SDP and Epipe

An end-to-end Epipe service is established between AN-1 and AN-5, as shown in [Figure 114: End-to-End Epipe service](#).

Figure 114: End-to-End Epipe service



25638

The SDP is configured on AN-1 and AN-5, as follows:

```
# on AN-1:
configure exclusive
  service {
    sdp 15 {
      admin-state enable
      delivery-type mpls
      bgp-tunnel true
      far-end {
        ip-address 192.0.2.5
      }
    }
  }
```

```
# on AN-5:
configure exclusive
  service {
    sdp 51 {
      admin-state enable
      delivery-type mpls
      bgp-tunnel true
      far-end {
        ip-address 192.0.2.1
      }
    }
  }
```

Epipe 1 is configured on AN-1 and AN-5, as follows:

```
# on AN-1:
configure exclusive
  service {
    epipe "Epipe 1" {
      admin-state enable
      service-id 1
      customer "1"
      spoke-sdp 15:1 {
      }
      sap 1/1/c1/3:1 {
      }
    }
  }
```

```
# on AN-5:
```

```
configure exclusive
  service {
    epipe "Epipe 1" {
      admin-state enable
      service-id 1
      customer "1"
      spoke-sdp 51:1 {
      }
      sap 1/1/c1/3:1 {
      }
    }
  }
}
```

The state of the SDP and of the Epipe service can be verified on AN-1, as follows:

```
[/]
A:admin@AN-1# show service sdp

=====
Services: Service Destination Points
=====
SdpId  AdmMTU  OprMTU  Far End          Adm  Opr          Del  LSP  Sig
-----
15     0        1552    192.0.2.5        Up  Up           MPLS B    TLDP
-----
Number of SDPs : 1
-----
Legend: R = RSVP, L = LDP, B = BGP, M = MPLS-TP, n/a = Not Applicable
        I = SR-ISIS, 0 = SR-OSPF, T = SR-TE, F = FPE
=====
```

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

=====
Service Basic Information
=====
Service Id       : 1                Vpn Id          : 0
Service Type     : Epipe
---snip---
Admin State      : Up                Oper State      : Up
---snip---
-----
Service Access & Destination Points
-----
Identifier                Type          AdmMTU  OprMTU  Adm  Opr
-----
sap:1/1/c1/3:1            q-tag        1518    1518    Up   Up
sdp:15:1 S(192.0.2.5)     Spok         0        1552    Up   Up
-----
* indicates that the corresponding row element may have been truncated.
```

The state of the SDP and of the Epipe service can be verified on AN-5, as follows:

```
[/]
A:admin@AN-5# show service sdp

=====
Services: Service Destination Points
=====
SdpId  AdmMTU  OprMTU  Far End          Adm  Opr          Del  LSP  Sig
-----
51     0        1552    192.0.2.1        Up  Up           MPLS B    TLDP
-----
```



```

-----
Number of SDPs : 1
-----
Legend: R = RSVP, L = LDP, B = BGP, M = MPLS-TP, n/a = Not Applicable
       I = SR-ISIS, O = SR-OSPF, T = SR-TE, F = FPE
=====

[/]
A:admin@AN-5# show service id 1 base

=====
Service Basic Information
=====
Service Id       : 1                Vpn Id          : 0
Service Type     : Epipe
---snip---
Admin State      : Up              Oper State      : Up
---snip---

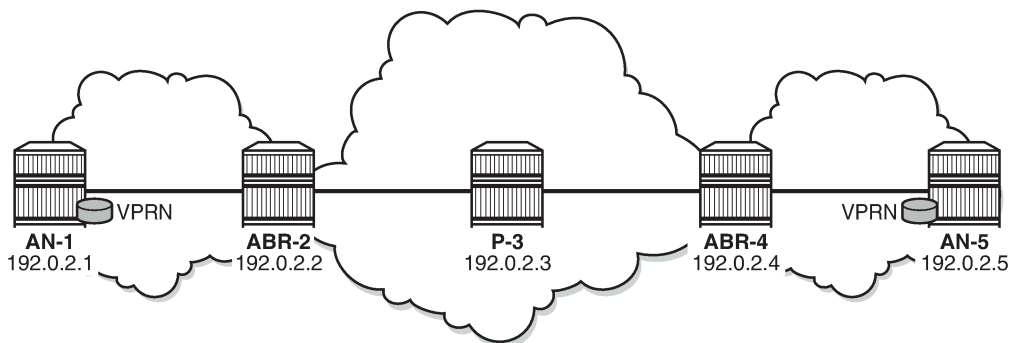
Service Access & Destination Points
-----
Identifier                               Type            AdmMTU  OprMTU  Adm  Opr
-----
sap:1/1/c1/3:1                           q-tag           1518   1518   Up   Up
sdp:51:1 S(192.0.2.1)                     Spok            0      1552   Up   Up
=====
* indicates that the corresponding row element may have been truncated.

```

## Configure VPRN

An L3 VPN service is established on AN-1 and AN-5, as shown in [Figure 115: L3 VPN service](#).

Figure 115: L3 VPN service



25639

The VPRN service is configured on AN-1 and AN-5, as follows. For simplicity, no CEs are attached to the ANs and only one loopback is created for verification.

```

# on AN-1:
configure exclusive
service {
  vprn "VPRN 2" {
    admin-state enable
    service-id 2
  }
}

```

```

customer "1"
  bgp-ipvpn {
    mpls {
      admin-state enable
      route-distinguisher "64496:2"
      vrf-target {
        community "target:64496:2"
      }
      auto-bind-tunnel {
        resolution any
      }
    }
  }
  interface "loopback" {
    loopback true
    ipv4 {
      primary {
        address 192.0.1.1
        prefix-length 32
      }
    }
  }
}

```

```

# on AN-5:
configure exclusive
  service {
    vprn "VPRN 2" {
      admin-state enable
      service-id 2
      customer "1"
      bgp-ipvpn {
        mpls {
          admin-state enable
          route-distinguisher "64496:2"
          vrf-target {
            community "target:64496:2"
          }
          auto-bind-tunnel {
            resolution any
          }
        }
      }
      interface "loopback" {
        loopback true
        ipv4 {
          primary {
            address 192.0.1.5
            prefix-length 32
          }
        }
      }
    }
  }
}

```

The routing table for VPRN 2 contains the local and the remote loopback addresses. On AN-1, this can be verified as follows:

```

[/]
A:admin@AN-1# show router "2" route-table

=====
Route Table (Service: 2)
=====
Dest Prefix[Flags]                                Type    Proto    Age      Pref

```

```

Next Hop[Interface Name]                               Metric
-----
192.0.1.1/32                                           Local  Local  00h02m26s  0
  loopback
192.0.1.5/32                                           Remote  BGP VPN 00h01m08s 170
  192.0.2.5 (tunneled:BGP)                             1000
-----
No. of Routes: 2
Flags: n = Number of times nexthop is repeated
      B = BGP backup route available
      L = LFA nexthop available
      S = Sticky ECMP requested
=====

```

On AN-5, this can be verified as follows:

```

[/]
A:admin@AN-5# show router "2" route-table

=====
Route Table (Service: 2)
=====
Dest Prefix[Flags]                                     Type  Proto   Age           Pref
  Next Hop[Interface Name]                             Metric
-----
192.0.1.1/32                                           Remote  BGP VPN 00h01m31s 170
  192.0.2.1 (tunneled:BGP)                             1000
192.0.1.5/32                                           Local  Local  00h02m19s  0
  loopback
-----
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
=====

```

Ping messages can be sent from the loopback address in VPRN 2 on AN-1 to the remote loopback address in VPRN 2 on AN-5, as follows:

```

[/]
A:admin@AN-1# ping 192.0.1.5 router-instance "VPRN 2"
PING 192.0.1.5 56 data bytes
64 bytes from 192.0.1.5: icmp_seq=1 ttl=64 time=2.69ms.
64 bytes from 192.0.1.5: icmp_seq=2 ttl=64 time=3.18ms.
. 64 bytes from 192.0.1.5: icmp_seq=3 ttl=64 time=2.87ms.
. 64 bytes from 192.0.1.5: icmp_seq=4 ttl=64 time=2.64ms.
. 64 bytes from 192.0.1.5: icmp_seq=5 ttl=64 time=2.36ms.

---- 192.0.1.5 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 2.36ms, avg = 2.75ms, max = 3.18ms, stddev = 0.271ms

```

In a similar way, ping messages are sent from the loopback address in VPRN 2 on AN-5 to the loopback address in VPRN 2 on AN-1, as follows:

```

[/]
A:admin@AN-5# ping 192.0.1.1 router-instance "VPRN 2"
PING 192.0.1.1 56 data bytes
64 bytes from 192.0.1.1: icmp_seq=1 ttl=64 time=2.05ms.
64 bytes from 192.0.1.1: icmp_seq=2 ttl=64 time=1.82ms.

```

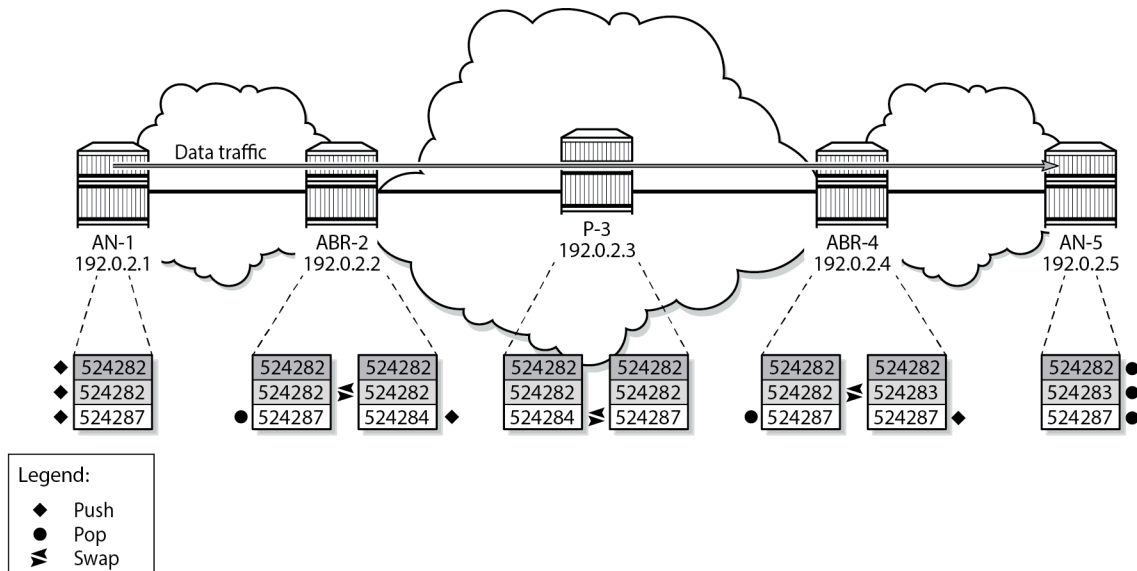
```
64 bytes from 192.0.1.1: icmp_seq=3 ttl=64 time=1.92ms.
. 64 bytes from 192.0.1.1: icmp_seq=4 ttl=64 time=1.90ms.
. 64 bytes from 192.0.1.1: icmp_seq=5 ttl=64 time=2.40ms.

---- 192.0.1.1 PING Statistics ----
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min = 1.82ms, avg = 2.02ms, max = 2.40ms, stddev = 0.206ms
```

## Data plane overview

Figure 116: Label stacks for traffic from AN-1 to AN-5 shows the label stacks used for traffic from AN-1 to AN-5. As an example, an Epipe service is used.

Figure 116: Label stacks for traffic from AN-1 to AN-5



38583

1. The service label added for Epipe 1 on AN-1 for egress traffic to AN-5 is 524282. Ingress traffic on AN-1 has service label 524282. This can be shown as follows:

```
[/]
A:admin@AN-1# show service id 1 labels

=====
Martini Service Labels
=====
Svc Id   Sdp Binding   Type  I.Lbl   E.Lbl
-----
1        15:1         Spok  524282  524282
-----
Number of Bound SDPs : 1
=====
```

This service label remains unchanged end-to-end.

On AN-1, the (middle) BGP label for traffic with destination AN-5 is 524282, as follows:

```
[/]
A:admin@AN-1# show router bgp routes 192.0.2.5 label-ipv4
=====
BGP Router ID:192.0.2.1      AS:64496      Local AS:64496
=====
Legend -
Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                l - leaked, x - stale, > - best, b - backup, p - purge
Origin codes  : i - IGP, e - EGP, ? - incomplete
=====
BGP LABEL-IPV4 Routes
=====
Flag  Network                LocalPref  MED
      Nexthop (Router)      Path-Id    IGP Cost
      As-Path                Label
-----
u*>i 192.0.2.5/32              100        None
      192.0.2.2              None        10
      No As-Path              524282
-----
Routes : 1
=====
```

The next hop is ABR-2. AN-1 pushes the LDP label 524287 to reach ABR-2, as follows:

```
[/]
A:admin@AN-1# show router ldp bindings active prefixes prefix 192.0.2.2/32
=====
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
-----
192.0.2.2/32          Push
--                    524287
192.168.12.2         1/1/c1/1
-----
No. of IPv4 Prefix Active Bindings: 1
=====
```

2. At ABR-2, the service label 524282 remains unchanged. The LDP label 524287 is popped, as follows:

```
[/]
A:admin@ABR-2# show router ldp bindings active prefixes prefix 192.0.2.2/32

=====
LDP Bindings (IPv4 LSR ID 192.0.2.2)
(IPv6 LSR ID ::)
=====
---snip---
=====
LDP IPv4 Prefix Bindings (Active)
=====
Prefix                               Op
IngLbl                               EgrLbl
EgrNextHop                           EgrIf/LspId
-----
192.0.2.2/32                         Pop
524287                                --
--                                     --
-----
No. of IPv4 Prefix Active Bindings: 1
=====
```

- On ABR-2, the BGP next hop is ABR-4 for prefix 192.0.2.5, as follows:

```
[/]
A:admin@ABR-2# show router bgp routes 192.0.2.5 label-ipv4

=====
BGP Router ID:192.0.2.2      AS:64496      Local AS:64496
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
               l - leaked, x - stale, > - best, b - backup, p - purge
Origin codes  : i - IGP, e - EGP, ? - incomplete
=====
BGP LABEL-IPV4 Routes
=====
Flag  Network                               LocalPref  MED
      Nexthop (Router)                     Path-Id    IGP Cost
      As-Path                               Label
-----
u*>i 192.0.2.5/32                           100        None
      192.0.2.4                             None        20
      No As-Path                             524282
-----
Routes : 1
=====
```

- On ABR-2, the BGP (middle) label 524282 is swapped with (in this case, the same label) 524282 for BGP next hop ABR-4, as follows:

```
[/]
A:admin@ABR-2# show router bgp inter-as-label

=====
BGP Inter-AS labels
Flags: B - entry has backup, P - entry is promoted
=====
NextHop                               Received   Advertised   Label
-----
```

	Label	Label	Origin
192.0.2.1	524283	524283	Internal
<b>192.0.2.4</b>	<b>524282</b>	<b>524282</b>	<b>Internal</b>
-----			
Total Labels allocated: 2			
=====			

ABR-2 pushes a new LDP label (524284) to reach the BGP next hop (ABR-4), as follows:

```
[/]
A:admin@ABR-2# show router ldp bindings active prefixes prefix 192.0.2.4/32

=====
LDP Bindings (IPv4 LSR ID 192.0.2.2)
(IPv6 LSR ID ::)
=====
---snip---
=====
LDP IPv4 Prefix Bindings (Active)
=====
Prefix                               Op
IngLbl                               EgrLbl
EgrNextHop                           EgrIf/LspId
-----
192.0.2.4/32                         Push
--                                   524284
192.168.23.2                         1/1/c1/1

192.0.2.4/32                         Swap
524284                               524284
192.168.23.2                         1/1/c1/1

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

3. At LSR P-3, only an LDP label swap occurs. P-3 swaps LDP label 524284 with 524287, as follows:

```
[/]
A:admin@P-3# show router ldp bindings active prefixes prefix 192.0.2.4/32

=====
LDP Bindings (IPv4 LSR ID 192.0.2.3)
(IPv6 LSR ID ::)
=====
---snip---
=====
LDP IPv4 Prefix Bindings (Active)
=====
Prefix                               Op
IngLbl                               EgrLbl
EgrNextHop                           EgrIf/LspId
-----
192.0.2.4/32                         Push
--                                   524287
192.168.34.2                         1/1/c1/1

192.0.2.4/32                         Swap
524284                               524287
192.168.34.2                         1/1/c1/1

-----
```

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

4. At ABR-4, LDP label 524287 is popped and BGP label 524282 is swapped to label 524283, as follows:

```
[/]
A:admin@ABR-4# show router bgp inter-as-label

=====
BGP Inter-AS labels
Flags: B - entry has backup, P - entry is promoted
=====
NextHop                Received      Advertised    Label
                        Label         Label         Origin
-----
192.0.2.2              524283       524283        Internal
192.0.2.5            524283       524282        Internal
-----
Total Labels allocated:  2
=====
```

- ABR-4 pushes a new LDP label 524287 to reach AN-5, as follows:

```
[/]
A:admin@ABR-4# show router ldp bindings active prefixes prefix 192.0.2.5/32

=====
LDP Bindings (IPv4 LSR ID 192.0.2.4)
(IPv6 LSR ID ::)
=====
---snip---
=====
LDP IPv4 Prefix Bindings (Active)
=====
Prefix                Op
IngLbl                EgrLbl
EgrNextHop            EgrIf/LspId
-----
192.0.2.5/32          Push
--                    524287
192.168.45.2          1/1/c1/1

192.0.2.5/32          Swap
524284                524287
192.168.45.2          1/1/c1/1

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

5. Finally, at AN-5, all labels in the stack are popped. The LDP label 524287 is popped as follows:

```
[/]
A:admin@AN-5# show router ldp bindings active prefixes prefix 192.0.2.5/32

=====
LDP Bindings (IPv4 LSR ID 192.0.2.5)
(IPv6 LSR ID ::)
=====
---snip---
=====
LDP IPv4 Prefix Bindings (Active)
```



```

=====
Prefix                               Op
IngLbl                               EgrLbl
EgrNextHop                           EgrIf/LspId
-----
192.0.2.5/32                         Pop
524287                                --
--                                    --
-----
No. of IPv4 Prefix Active Bindings: 1
=====

```

The BGP (middle) label 524283 is popped.

```

[/]
A:admin@AN-5# show router bgp inter-as-label

=====
BGP Inter-AS labels
Flags: B - entry has backup, P - entry is promoted
=====
NextHop                Received      Advertised    Label
                        Label         Label         Origin
-----
0.0.0.0                0            524283        Edge
-----
Total Labels allocated: 1
=====

```

The ingress service label 524282 is popped, as follows:

```

[/]
A:admin@AN-5# show service id 1 labels

=====
Martini Service Labels
=====
Svc Id   Sdp Binding   Type  I.Lbl           E.Lbl
-----
1        51:1         Spok  524282          524282
-----
Number of Bound SDPs : 1
=====

```

## OAM

The following Operations, Administration, and Maintenance (OAM) commands can be launched to validate reachability between regions using BGP labeled IPv4 routes.

```

[/]
A:admin@AN-1# oam lsp-ping bgp-label prefix 192.0.2.5/32
LSP-PING 192.0.2.5/32: 80 bytes MPLS payload
Seq=1, send from intf int-AN-1-ABR-2, reply from 192.0.2.5
      udp-data-len=32 ttl=255 rtt=2.63ms rc=3 (EgressRtr)

---- LSP 192.0.2.5/32 PING Statistics ----
1 packet sent, 1 packet received, 0.00% packet loss

```

```
round-trip min = 2.63ms, avg = 2.63ms, max = 2.63ms, stddev = 0.000ms
```

```
[/]
A:admin@AN-5# oam lsp-ping bgp-label prefix 192.0.2.1/32
LSP-PING 192.0.2.1/32: 80 bytes MPLS payload
Seq=1, send from intf int-AN-5-ABR-4, reply from 192.0.2.1
    udp-data-len=32 ttl=255 rtt=3.84ms rc=3 (EgressRtr)

---- LSP 192.0.2.1/32 PING Statistics ----
1 packet sent, 1 packet received, 0.00% packet loss
round-trip min = 3.84ms, avg = 3.84ms, max = 3.84ms, stddev = 0.000ms
```

In a similar way, LSP trace can validate the reachability of the remote AN, as follows:

```
[/]
A:admin@AN-1# oam lsp-trace bgp-label prefix 192.0.2.5/32 output-format detail
lsp-trace to 192.0.2.5/32: 1 hops min, 30 hops max, 104 byte packets
0 192.0.2.1
    DS 1: ipaddr=192.168.12.2 ifaddr=192.168.12.2 iftype=ipv4Numbered MRU=1560
        label[1]=524282 protocol=2(BGP)
1 192.0.2.2 rtt=1.18ms rc=8(DSRtrMatchLabel) rsc=1
    DS 1: ipaddr=192.168.23.2 ifaddr=192.168.23.2 iftype=ipv4Numbered MRU=1560
        label[1]=524282 protocol=2(BGP).
2 192.0.2.4 rtt=2.03ms rc=8(DSRtrMatchLabel).
3 192.0.2.5 rtt=2.26ms rc=3(EgressRtr) rsc=1
```

```
[/]
A:admin@AN-5# oam lsp-trace bgp-label prefix 192.0.2.1/32 output-format detail
lsp-trace to 192.0.2.1/32: 1 hops min, 30 hops max, 104 byte packets
0 192.0.2.5
    DS 1: ipaddr=192.168.45.1 ifaddr=192.168.45.1 iftype=ipv4Numbered MRU=1560
        label[1]=524283 protocol=2(BGP)
1 192.0.2.4 rtt=1.06ms rc=8(DSRtrMatchLabel) rsc=1
    DS 1: ipaddr=192.168.34.1 ifaddr=192.168.34.1 iftype=ipv4Numbered MRU=1560
        label[1]=524283 protocol=2(BGP).
2 192.0.2.2 rtt=1.54ms rc=8(DSRtrMatchLabel).
3 192.0.2.1 rtt=1.88ms rc=3(EgressRtr) rsc=1
```

## Conclusion

Seamless MPLS helps to solve the scalability problems of large networks. Seamless MPLS partitions the core, aggregation, and access networks into isolated IGP/LDP domains, which helps to maintain IGP databases small and controlled. Labeled BGP allows the establishment of hierarchical LSPs for end-to-end service set up.

# Shared Risk Link Groups for RSVP-Based LSPs

This chapter provides information about Shared Risk Link Groups for RSVP-Based LSPs.

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

## Applicability

This chapter was initially written for SR OS Release 7.0.R5, but the MD-CLI in the current edition corresponds to SR OS Release 21.2.R1. There are no prerequisites.

## Overview

### Introduction

Shared Risk Link Group (SRLG) is a feature which allows the user to establish a backup secondary label switched path (LSP) or a fast-reroute (FRR) LSP which is disjoint from the primary LSP. Links which are members of the same SRLG represent resources which share the same risk. For example, fiber links sharing the same conduit or multiple wavelengths sharing the same fiber.

A typical application of the SRLG feature is to provide an automatic placement of secondary backup LSPs or FRR bypass/detour LSPs that minimizes the probability of fate sharing with the primary LSP.

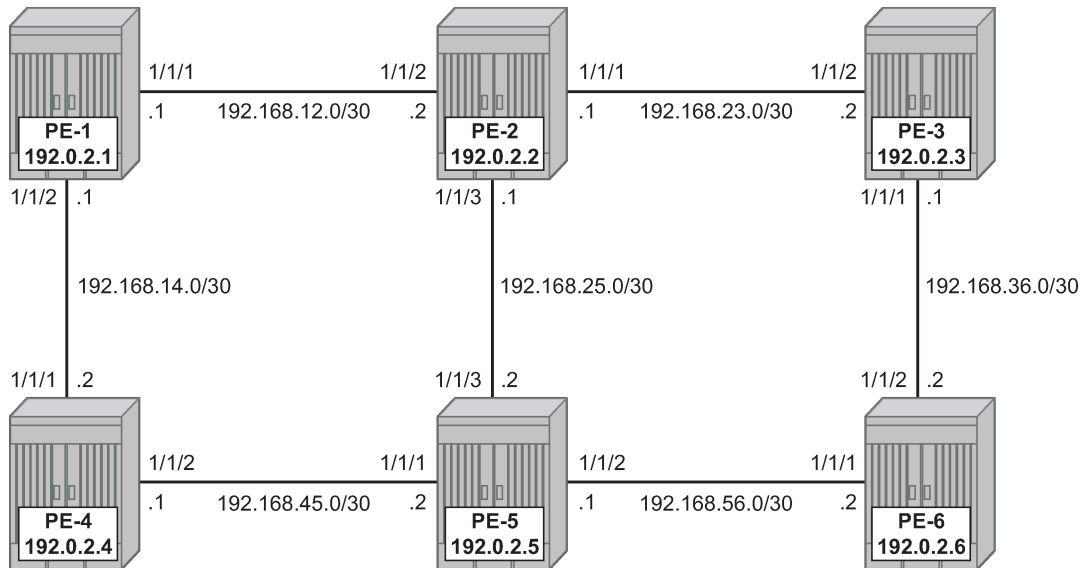
SRLG groups are used to determine which links belong to the same SRLG. The mechanism is similar to Multi-Protocol Label Switching (MPLS) admin groups. To advertise SRLG, the information is part of the IGP TE parameters in an opaque link state advertisement (LSA). In IS-IS (RFC 4205, *Intermediate System to Intermediate System (IS-IS) Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)*), the SRLG is advertised in a new Shared Risk Link Group TLV (type 138). In OSPF (RFC 4203, *OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)*), the SRLG is advertised in a new SRLG sub-TLV (type 16) of the existing Link TLV.

For FRR, a choice can be made on what to do when no FRR tunnel can be found with the SRLG constraints. No FRR tunnel may be signaled or a FRR tunnel might be signaled not taking the SRLG constraints into account.

### SRLG

[Figure 117: Example topology](#) shows the example topology for this chapter.

Figure 117: Example topology



OSSG413

A single IGP area (IS-IS in this case) with traffic engineering (TE) enabled is required for the SRLG feature to work properly.

When OSPF is used as the IGP, the functionality is similar.

## Configuration

### Configuring the IP/MPLS network

IS-IS, MPLS, and RSVP are configured on all interfaces. TE is enabled in IS-IS. Optionally, admin groups "green" and "red" are configured on all nodes. The "green" links are the following: the link between PE-1 and PE-2, the link between PE-2 and PE-3, and the link between PE-3 and PE-6. The "red" links are: the link between PE-1 and PE-4, the link between PE-4 and PE-5, and the link between PE-5 and PE-6. The remaining link is the link between PE-2 and PE-5, which does not belong to an admin group. For more information about admin groups, see chapter [RSVP Point-to-Point LSPs](#).

In addition, ECMP is set to 2, instead of the default value 1 in order to highlight the application of SRLG in the final example: SRLG database.

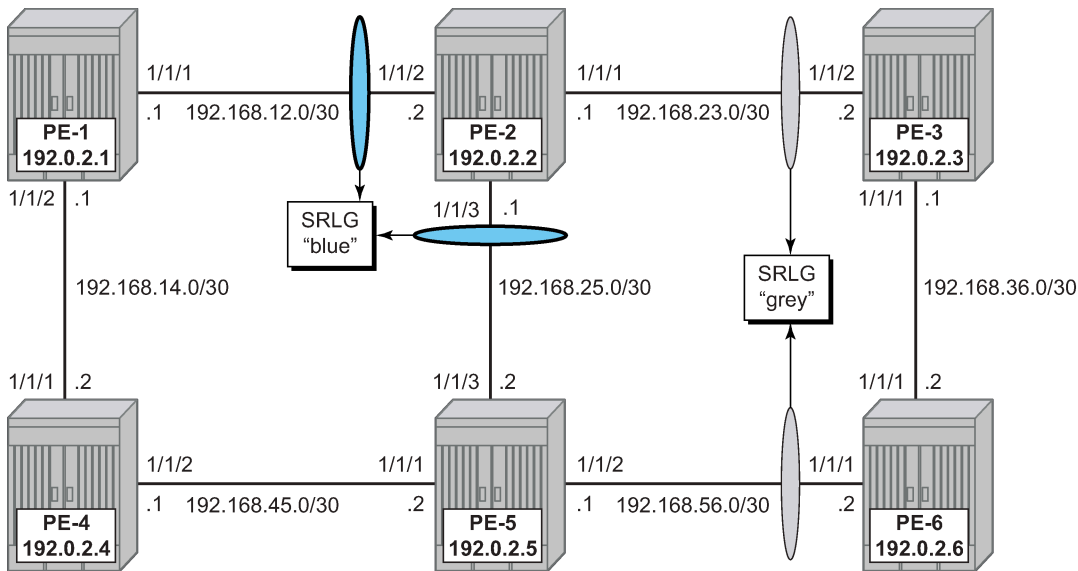
```
# on PE-1:
configure {
  router "Base" {
    ecmp 2
```

## Define SRLG groups

Define the SRLG groups, and link them to the related MPLS interfaces.

Two SRLG groups are defined, named blue and gray, as shown in [Figure 118: SRLG topology](#).

Figure 118: SRLG topology



OSSG415

The configuration of the blue SRLG group is only mandatory on PE-1, PE-2, and PE-5, while the gray SRLG group is only mandatory on PE-2, PE-3, PE-5, and PE-6. However, it is good practice to configure both SRLG groups on all nodes, as follows:

```
# on all nodes:
configure {
  routing-options {
    if-attribute {
      srlg-group "blue" {
        value 1
      }
      srlg-group "gray" {
        value 2
      }
    }
  }
}
```

The IP/MPLS interfaces need to be linked to the related SRLG group, which is a unidirectional indicator, applying only to the egress direction; therefore, it needs to be configured on both sides of the IP/MPLS interface. For example on PE-1, the interface to PE-2 is part of srlg-group "blue". An interface can be part of multiple SRLG groups similar to the admin-group functionality.

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

```

        admin-group ["green"]
        srlg-group "blue" { }
    }
    interface "int-PE-1-PE-4" {
        admin-group ["red"]
    }

```

The same must be done on PE-2, PE-3, PE-5, and PE-6. Afterward, verify the MPLS configuration for example on PE-2, where the SRLG groups are linked to the interfaces. Admin groups are configured in parallel to indicate that both can be configured and will work independently.

```

# on PE-2:
configure {
    router "Base" {
        mpls {
            admin-state enable
            interface "int-PE-2-PE-1" {
                admin-group ["green"]
                srlg-group "blue" { }
            }
            interface "int-PE-2-PE-3" {
                admin-group ["green"]
                srlg-group "gray" { }
            }
            interface "int-PE-2-PE-5" {
                srlg-group "blue" { }
            }
        }
    }
}

```

The SRLG configuration can be verified using the following show commands.

The following shows all SRLG groups on the node:

```

[/]
A:admin@PE-2# show router if-attribute srlg-group
=====
Interface Srlg Groups
=====
Group Name                Group Value    Penalty Weight
-----
blue                       1              0
gray                       2              0
-----
No. of Groups: 2
=====

```

In the following list of MPLS interfaces, admin groups and SRLG groups are indicated.

```

[/]
A:admin@PE-2# show router mpls interface
=====
MPLS Interfaces
=====
Interface                Port-id        Adm  Opr(V4/V6)  TE-
                        metric
-----
system                    system         Up   Up/Down     None
  Admin Groups            None
  SRLG Groups             None
int-PE-2-PE-1            1/1/2         Up   Up/Down     None
  Admin Groups            green
  SRLG Groups             blue

```

```

int-PE-2-PE-3          1/1/1          Up   Up/Down   None
  Admin Groups         green
  SRLG Groups          gray
int-PE-2-PE-5          1/1/3          Up   Up/Down   None
  Admin Groups         None
  SRLG Groups          blue
-----
Interfaces : 4
=====

```

To verify the SRLG groups in the IGP TE database, the following command can be used. The output can be extensive, but searching on the SRLG group name will lead to the correct interfaces.

The following output shows the link-state advertisements of PE-2 on PE-1 in this case. The SRLG information is linked to the IP interfaces in a dedicated TE-TLV.

```

[/]
A:admin@PE-2# show router isis database PE-2.00-00 detail

=====
Rtr Base ISIS Instance 0 Database (detail)
=====

Displaying Level 1 database
-----
LSP ID   : PE-2.00-00          Level   : L1
Sequence : 0x34                Checksum : 0xdfef          Lifetime : 1051
Version  : 1                   Pkt Type : 18              Pkt Ver  : 1
Attributes: L1                 Max Area : 3               Alloc Len : 508
SYS ID   : 1920.0000.2002      SysID Len : 6              Used Len  : 508

TLVs :

---snip---

TE SRLGs :
  SRLGs : PE-1.00
  Lcl Addr : 192.168.12.2
  Rem Addr : 192.168.12.1
  Num SRLGs : 1
          1

---snip---

TE SRLGs :
  SRLGs : PE-3.00
  Lcl Addr : 192.168.23.1
  Rem Addr : 192.168.23.2
  Num SRLGs : 1
          2

---snip---

TE SRLGs :
  SRLGs : PE-5.00
  Lcl Addr : 192.168.25.1
  Rem Addr : 192.168.25.2
  Num SRLGs : 1
          1

---snip---

```

## On-line verification

An on-line verification can be done by a **tools perform** command. This will trigger a Constrained Shortest Path First (CSPF) call to the Interior Gateway Protocol (IGP) TE database, and the result will be an Explicit Route Object (ERO) object which can potentially be used to set up a CSPF-based LSP.

The following shows the command syntax.

```
[/]
A:admin@PE-1# tools perform router mpls cspf to 192.0.2.3 ?

cspf to <IP address> [from <IP address>]
                    [bandwidth <number>]
                    [include-bitmap <string>]
                    [exclude-bitmap <string>]
                    [hop-limit <number>]
                    [exclude-address <ipv4 address>]
                    [metric-type-te]
                    [strict-srlg]
                    [srlg-group <number>]
                    [exclude-node <ipv4 address>]
                    [skip-interface <string>]
                    [ds-class-type <number>]
                    [cspf-reqtype <keyword>]
                    [least-fill-min-thd <number>]
                    [setup-priority <number>]
                    [hold-priority <number>]

bandwidth           - rate-in-mbps
cspf-reqtype        - all|random|least-fill : keywords
ds-class-type       - <number> - <0..7>
exclude-address     - a.b.c.d (outbound interface)
exclude-bitmap      - string '<1..40 characters>'
exclude-node        - [a.b.c.d] (outbound interface)
from                - ipv4 address '<d.d.d.d>' or ipv6 address
                    '(<x:x:x:x:x:x:x>|<x:x:x:x:x:d.d.d.d>)'
hold-priority       - <number> - <0..7>
hop-limit           - <number> - <2..255>
include-bitmap      - [0..4294967295] - accepted in decimal, hex(0x) or binary(0b)
least-fill-min-thd  - <number> - <1..100>
metric-type-te      - <keyword>
setup-priority      - <number> - <0..7>
skip-interface      - [max 32 chars]
srlg-group          - <number> - <0..4294967295>
strict-srlg         - <keyword>
```

Where the relevant parameters are:

- **to** — Defines the far-end address of the LSP. This is the system-address of the destination LER
- **srlg-group** — Specifies which SRLG groups should be avoided while building the path to the destination (ERO object)
- **strict-srlg** — Indicates whether the SRLG group is a strict requirement or not. When this parameter is given, only paths without traversing the SRLG will be displayed.

Example:

On PE-1, a CSPF calculation is made with PE-3 as destination, without any SRLG restrictions, as follows:

```
[/]
```



```
A:admin@PE-1# tools perform router mpls cspf to 192.0.2.3
Req C SPF for all ECMP paths
  from: this node to: 192.0.2.3 w/(no Diffserv) class: 0 , setup Priority 7,
                                         Hold Priority 0 TE Class: 7

CSPF Path
To      : 192.0.2.3
Path 1  : (cost 20)
  Src:   192.0.2.1   (= Rtr)
  Egr:   192.168.12.1   -> Ingr: 192.168.12.2      Rtr: 192.0.2.2      (met 10)
  Egr:   192.168.23.1   -> Ingr: 192.168.23.2      Rtr: 192.0.2.3      (met 10)
  Dst:   192.0.2.3   (= Rtr)
```

With a restriction on **srlg-group "blue"** (grp-id =1), the CSPF calculation is as follows:

```
[/]
A:admin@PE-1# tools perform router mpls cspf to 192.0.2.3 srlg-group 1
Req C SPF for all ECMP paths
  from: this node to: 192.0.2.3 w/(no Diffserv) class: 0 , setup Priority 7,
                                         Hold Priority 0 TE Class: 7

CSPF Path
To      : 192.0.2.3
Path 1  : (cost 40)
  Src:   192.0.2.1   (= Rtr)
  Egr:   192.168.14.1   -> Ingr: 192.168.14.2      Rtr: 192.0.2.4      (met 10)
  Egr:   192.168.45.1   -> Ingr: 192.168.45.2      Rtr: 192.0.2.5      (met 10)
  Egr:   192.168.56.1   -> Ingr: 192.168.56.2      Rtr: 192.0.2.6      (met 10)
  1 SRLGs: 2
  Egr:   192.168.36.2   -> Ingr: 192.168.36.1      Rtr: 192.0.2.3      (met 10)
  Dst:   192.0.2.3   (= Rtr)
```

The path will be through PE-4, PE-5, and PE-6.

When a strict restriction is requested on **srlg-group "gray"**, no valid CSPF path toward the destination can be found.

```
[/]
A:admin@PE-1# tools perform router mpls cspf to 192.0.2.3 srlg-group 2 strict-srlg
Req C SPF for all ECMP paths
  from: this node to: 192.0.2.3 w/(no Diffserv) class: 0 , setup Priority 7,
                                         Hold Priority 0 TE Class: 7

MINOR: CLI #2005: Error while processing command - No CSPF path to "192.0.2.3" with specified
constraints.
```

Removing the **strict** restriction results in a successful return of CSPF, indicating that the CSPF path is not SRLG disjoint.

```
*A:PE-1# tools perform router mpls cspf to 192.0.2.3 srlg-group 2
Req C SPF for all ECMP paths
  from: this node to: 192.0.2.3 w/(no Diffserv) class: 0 , setup Priority 7,
                                         Hold Priority 0 TE Class: 7

CSPF Path
To      : 192.0.2.3 (NOT SRLG DISJOINT)
Path 1  : (cost 20)
  Src:   192.0.2.1   (= Rtr)
  Egr:   192.168.12.1   -> Ingr: 192.168.12.2      Rtr: 192.0.2.2      (met 10)
  1 SRLGs: 1
  Egr:   192.168.23.1   -> Ingr: 192.168.23.2      Rtr: 192.0.2.3      (met 10)
  1 SRLGs: 2
```

```
Dst: 192.0.2.3 (= Rtr)
```

The best practice for debugging is to enable debug-tracing on the CSPF process, with following command:

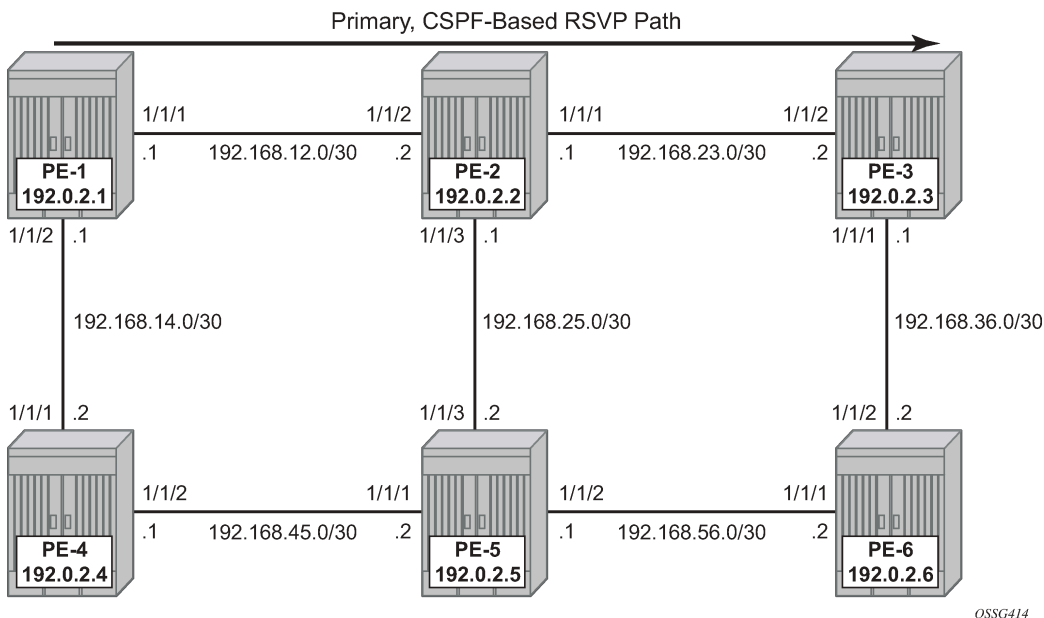
```
# on PE-1 (in classic CLI):
debug
router Base
isis 0
cspf
```

## SRLG for FRR

The fast-reroute mechanism used here is facility link protection (**frr-method facility node-protect false**). The SRLG feature is independent of the FRR type and works for all combinations (facility versus one-to-one, link versus node protection).

Configure an LSP from PE-1 to PE-3, and enable CSPF.

Figure 119: Path primary RSVP-TE LSP



The configuration of the LSP "LSP-PE-1-PE-3\_FRR\_facility-link" is based on an empty path, with FRR facility link protection enabled.

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      path "dyn" {
        admin-state enable
      }
      lsp "LSP-PE-1-PE-3_FRR_facility-link" {
        admin-state enable
        type p2p-rsvp
        to 192.0.2.3
      }
    }
  }
}
```

```

path-computation-method local-cspf
fast-reroute {
    frr-method facility
    node-protect false
}
primary "dyn" {
}
}

```

To verify the primary path, the classic-CLI command **oam lsp-trace** can be used, checking the intermediate nodes.

```

[/]
A:admin@PE-1# //
INFO: CLI #2051: Switching to the classic CLI engine
A:PE-1# oam lsp-trace "LSP-PE-1-PE-3_FRR_facility-link" detail
lsp-trace to LSP-PE-1-PE-3_FRR_facility-link: 0 hops min, 0 hops max, 116 byte packets
1 192.0.2.2 rtt=3.00ms rc=8(DSRtrMatchLabel) rsc=1
   DS 1: ipaddr=192.168.23.2 ifaddr=192.168.23.2 iftype=ipv4Numbered MRU=1564
       label[1]=524287 protocol=4(RSVP-TE)
2 192.0.2.3 rtt=3.47ms rc=3(EgressRtr) rsc=1

```

To verify if the bypass tunnels are up and running, an indication (@) can be found in the detail output of **show router mpls ls <x> path detail** as seen in the following output.

```

*A:PE-1# show router mpls lsp "LSP-PE-1-PE-3_FRR_facility-link" path detail
=====
MPLS LSP LSP-PE-1-PE-3_FRR_facility-link Path (Detail)
=====
Legend :
  @ - Detour Available          # - Detour In Use
  b - Bandwidth Protected      n - Node Protected
  s - Soft Preemption
  S - Strict                    L - Loose
  A - ABR                       + - Inherited
=====
-----
LSP LSP-PE-1-PE-3_FRR_facility-link
Path dyn
-----
LSP Name      : LSP-PE-1-PE-3_FRR_facility-link
From          : 192.0.2.1
To            : 192.0.2.3
Admin State   : Up              Oper State    : Up
Path Name     : dyn
Path LSP ID   : 25600           Path Type     : Primary
Path Admin    : Up              Path Oper     : Up
Out Interface : 1/1/1           Out Label     : 524287

---snip---

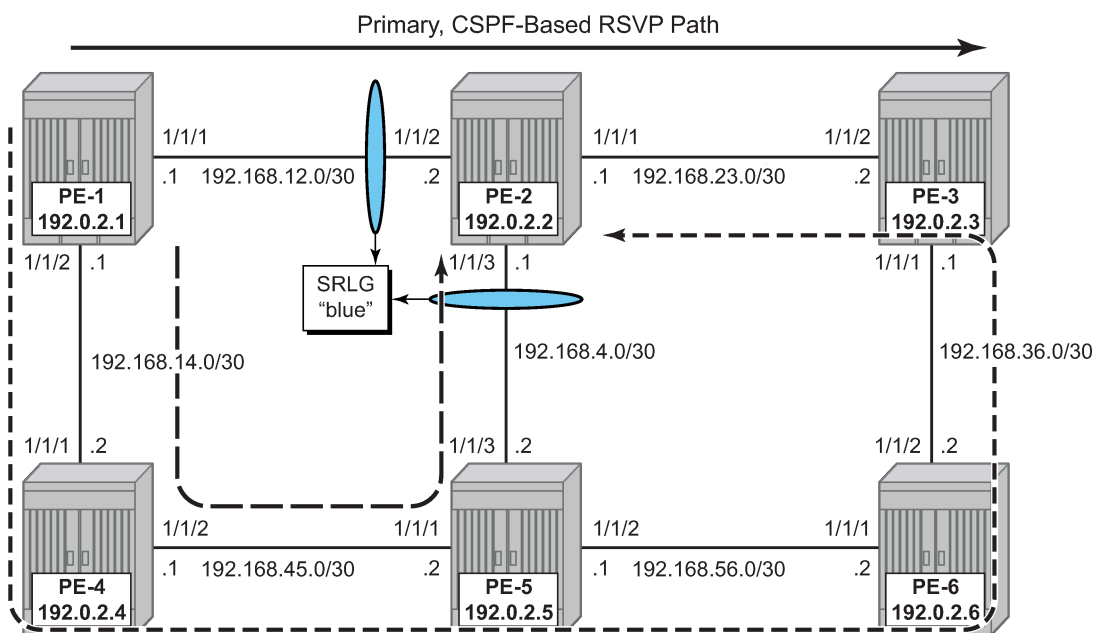
Explicit Hops :
  No Hops Specified
Actual Hops   :
  192.168.12.1(192.0.2.1) @      Record Label  : N/A
-> 192.168.12.2(192.0.2.2) @      Record Label  : 524287
-> 192.168.23.2(192.0.2.3) @      Record Label  : 524287
Computed Hops :
  192.168.12.1(S)
-> 192.168.12.2(S)
-> 192.168.23.2(S)

```

```
Resignal Eligible: False
Last Resignal    : n/a
CSPF Metric      : 20
=====
* indicates that the corresponding row element may have been truncated.
```

Two links are protected: one bypass tunnel originates in PE-1 protecting the link between PE-1 and PE-2. Another bypass tunnel originates in PE-2 protecting the link between PE-2 and PE-3. The focus is on the bypass tunnel originating in PE-1. When SRLG is enabled, the bypass tunnel originating in PE-1 will have different hops. The expected paths followed by the bypass tunnels originating in PE-1 with and without SRLG are shown in [Figure 120: FRR bypass tunnels originating in PE-1 with and without SRLG](#).

Figure 120: FRR bypass tunnels originating in PE-1 with and without SRLG



OSSG417

To verify the bypass data path on the point of local repair (PLR) PE-1, the following command can be used.

```
[/]
A:admin@PE-1# show router mpls bypass-tunnel detail

=====
MPLS Bypass Tunnels (Detail)
=====
-----
bypass-link192.168.12.2-61441
-----
To          : 192.168.25.1      State       : Up
Out I/F     : 1/1/2            Out Label   : 524287
Up Time     : 0d 00:05:20      Active Time : n/a
Reserved BW : 0 Kbps          Protected LSP Count : 1
Type        : Dynamic          Bypass Path Cost : 30
Setup Priority : 7              Hold Priority  : 0
Class Type  : 0
Exclude Node : None            Inter-Area    : False
Computed Hops :
    192.168.14.1(S)           Egress Admin Groups :
                                red
```

```

-> 192.168.14.2(S)      Egress Admin Groups :
                        red
-> 192.168.45.2(S)      Egress Admin Groups : None
-> 192.168.25.1(S)      Egress Admin Groups : None
Actual Hops      :
  192.168.14.1(192.0.2.1) Record Label      : N/A
-> 192.168.14.2(192.0.2.4) Record Label      : 524287
-> 192.168.45.2(192.0.2.5) Record Label      : 524286
-> 192.168.25.1(192.0.2.2) Record Label      : 524286
Last Resignal    :
Attempted At     : n/a      Resignal Reason    : n/a
Resignal Status  : n/a      Reason              : n/a
=====

```

The SRLG restriction is not taken into account at this moment at PLR PE-1. The actual hops are PE-4, PE-5, and PE-2 visualized by the path with the long dashes in [Figure 120: FRR bypass tunnels originating in PE-1 with and without SRLG](#).

To take the SRLG restrictions into account, the following additional configuration is needed for MPLS on PE-1.

```

[ex:configure router "Base" mpls]
A:admin@PE-1# srlg-frr ?

srlg-frr <keyword>
<keyword> - (loose|strict)

Use of Shared Risk Link Group (SRLG) constraint for FRR path computation

```

```

# on PE-1:
configure {
  router "Base" {
    mpls {
      srlg-frr strict
    }
  }
}

```

The option strict should only be used if the logical topology allows this. In other words, one must be sure that an alternative path is possible which avoids SRLG-groups.



**Note:**

Enabling or disabling SRLG for FRR is a system-wide configuration and requires the MPLS routing instance to be manually disabled and then re-enabled to activate the change. This may cause service outage. Nokia recommends that the operator incorporates the SRLG into the initial network design and implementation to minimize the traffic loss.

```

# on all nodes:
configure exclusive
  router "Base" {
    rsvp {
      admin-state disable
      commit
      admin-state enable
      commit
    }
  }

```

The bypass tunnel originating in PLR PE-1 can be verified with a previously used command.

```

[/]
A:admin@PE-1# show router mpls bypass-tunnel detail

```

```

=====
MPLS Bypass Tunnels (Detail)
=====
-----
bypass-link192.168.12.2-61442
-----
To           : 192.168.23.1      State          : Up
Out I/F      : 1/1/2            Out Label     : 524287
Up Time     : 0d 00:00:06      Active Time   : n/a
Reserved BW  : 0 Kbps          Protected LSP Count : 1
Type        : Dynamic          Bypass Path Cost : 50
Setup Priority : 7              Hold Priority   : 0
Class Type   : 0
Exclude Node : None            Inter-Area     : False
Computed Hops :
    192.168.14.1(S)           Egress Admin Groups :
    -> 192.168.14.2(S)         red
    -> 192.168.45.2(S)        Egress Admin Groups :
    -> 192.168.56.2(S)        red
    -> 192.168.36.1(S)        Egress Admin Groups :
    -> 192.168.23.1(S)        green
    Actual Hops :              Egress Admin Groups :
    192.168.14.1(192.0.2.1)   Record Label       : N/A
    -> 192.168.14.2(192.0.2.4) Record Label       : 524287
    -> 192.168.45.2(192.0.2.5) Record Label       : 524286
    -> 192.168.56.2(192.0.2.6) Record Label       : 524286
    -> 192.168.36.1(192.0.2.3) Record Label       : 524284
    -> 192.168.23.1(192.0.2.2) Record Label       : 524285
Last Resignal :
Attempted At  : n/a            Resignal Reason   : n/a
Resignal Status: n/a          Reason            : n/a
=====

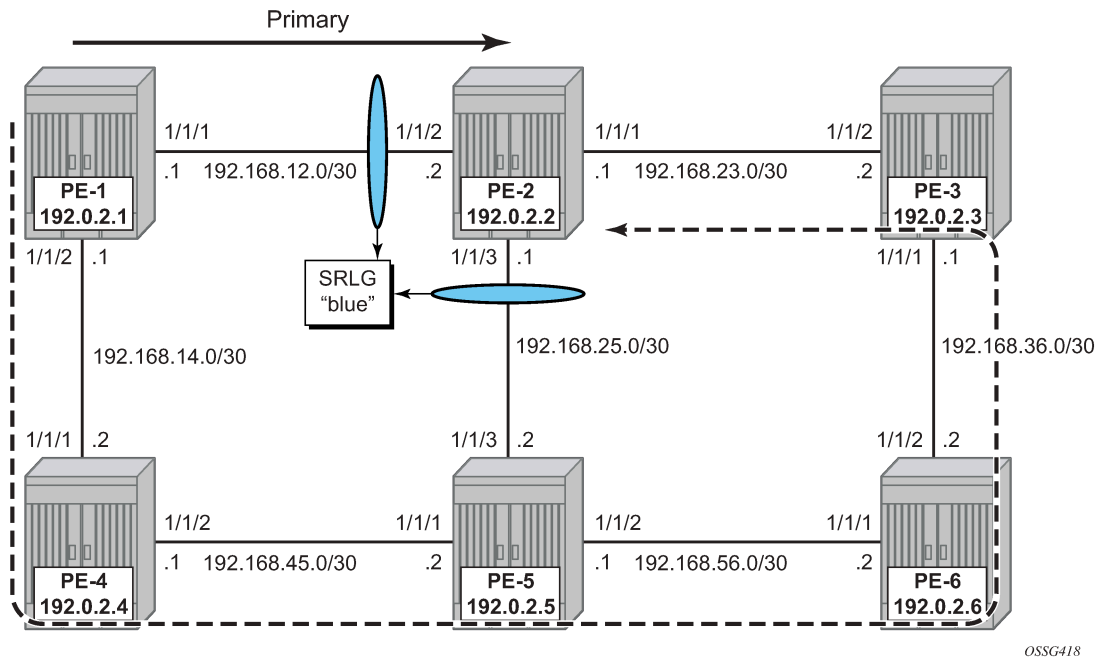
```

This path taking the SRLG constraints into account is represented by the line with the short dashes in [Figure 120: FRR bypass tunnels originating in PE-1 with and without SRLG](#).

### SRLG for standby path

Where SRLG groups can be constraints for bypass tunnels, they can also be a constraint to set up a secondary path. [Figure 121: SRLG for secondary path](#) shows that the secondary path is expected to follow the dashed line instead of passing over the direct link between PE-5 and PE-2.

Figure 121: SRLG for secondary path



OSSG418

An LSP is configured with a primary and a secondary path, which have no hops defined. The configuration of the LSP will need a specific indication at the level of the secondary path to enable the restriction on the srlg-groups.

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      path "prim" {
        admin-state enable
      }
      path "secon" {
        admin-state enable
      }
    }
    lsp "LSP-PE-1-PE-2-srlg" {
      admin-state enable
      type p2p-rsvp
      to 192.0.2.2
      path-computation-method local-cspf
      primary "prim" {
      }
      secondary "secon" {
        srlg true
        standby true
      }
    }
  }
}
```

Where both paths are empty paths, the ERO object creation solely relies on CPSF without any specific hop.

To verify the data path, the detailed output of the **show router mpls lsp <.> path** command can be used, as well as the **lsp-trace** OAM command. This output shows both ERO objects of the primary and secondary path.

```
*A:PE-1# show router mpls lsp "LSP-PE-1-PE-2-srlg" path detail

=====
MPLS LSP LSP-PE-1-PE-2-srlg Path (Detail)
=====
Legend :
  @ - Detour Available          # - Detour In Use
  b - Bandwidth Protected      n - Node Protected
  s - Soft Preemption
  S - Strict                    L - Loose
  A - ABR                      + - Inherited
=====
-----
LSP LSP-PE-1-PE-2-srlg
Path prim
-----
---snip---

Explicit Hops      :
  No Hops Specified
Actual Hops        :
  192.168.12.1(192.0.2.1)      Record Label      : N/A
-> 192.168.12.2(192.0.2.2)      Record Label      : 524287
Computed Hops      :
  192.168.12.1(S)
-> 192.168.12.2(S)
Resignal Eligible: False
Last Resignal     : n/a          CSPF Metric       : 10
-----
LSP LSP-PE-1-PE-2-srlg
Path secon
-----
---snip---

Explicit Hops      :
  No Hops Specified
Actual Hops        :
  192.168.14.1(192.0.2.1)      Record Label      : N/A
-> 192.168.14.2(192.0.2.4)      Record Label      : 524286
-> 192.168.45.2(192.0.2.5)      Record Label      : 524285
-> 192.168.56.2(192.0.2.6)      Record Label      : 524285
-> 192.168.36.1(192.0.2.3)      Record Label      : 524287
-> 192.168.23.1(192.0.2.2)      Record Label      : 524284
Computed Hops      :
  192.168.14.1(S)
-> 192.168.14.2(S)
-> 192.168.45.2(S)
-> 192.168.56.2(S)
-> 192.168.36.1(S)
-> 192.168.23.1(S)
Srlg               : Enabled          Srlg Disjoint     : True
Resignal Eligible: False
Last Resignal     : n/a          CSPF Metric       : 50
=====
```



The **lsp-trace** command can be used for secondary path as well. The intermediate LSRs and the MPLS labels used can be clearly seen.

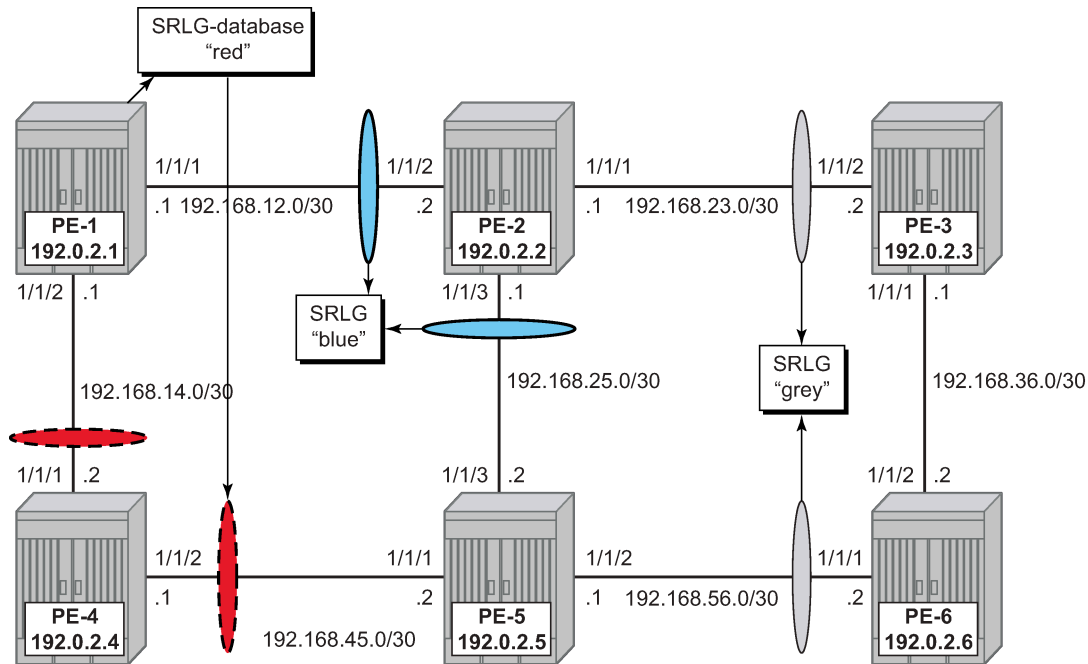
```
[/]
A:admin@PE-1# //
INFO: CLI #2051: Switching to the classic CLI engine
A:PE-1# oam lsp-trace LSP-PE-1-PE-2-srlg"LSP-PE-1-PE-2-srlg" path "secon" detail
lsp-trace to LSP-PE-1-PE-2-srlg: 0 hops min, 0 hops max, 116 byte packets
1 192.0.2.4 rtt=2.61ms rc=8(DSRtrMatchLabel) rsc=1
   DS 1: ipaddr=192.168.45.2 ifaddr=192.168.45.2 iftype=ipv4Numbered MRU=1564
        label[1]=524285 protocol=4(RSVP-TE)
2 192.0.2.5 rtt=3.97ms rc=8(DSRtrMatchLabel) rsc=1
   DS 1: ipaddr=192.168.56.2 ifaddr=192.168.56.2 iftype=ipv4Numbered MRU=1564
        label[1]=524285 protocol=4(RSVP-TE)
3 192.0.2.6 rtt=4.78ms rc=8(DSRtrMatchLabel) rsc=1
   DS 1: ipaddr=192.168.36.1 ifaddr=192.168.36.1 iftype=ipv4Numbered MRU=1564
        label[1]=524287 protocol=4(RSVP-TE)
4 192.0.2.3 rtt=4.43ms rc=8(DSRtrMatchLabel) rsc=1
   DS 1: ipaddr=192.168.23.1 ifaddr=192.168.23.1 iftype=ipv4Numbered MRU=1564
        label[1]=524284 protocol=4(RSVP-TE)
5 192.0.2.2 rtt=4.38ms rc=3(EgressRtr) rsc=1
A:PE-1# //
INFO: CLI #2052: Switching to the MD-CLI engine
```

## SRLG database

In case not all IP/MPLS routers in the area support SRLG, a static SRLG database can be created on the systems which will be used as an additional constraint when performing the CSPF calculation to define the path.

[Figure 122: SRLG database example](#) shows an example where an additional SRLG group "red" is defined on PE-1, with information related to the interface between PE-4 and PE-5.

Figure 122: SRLG database example



OSSG419

```
# on PE-1:
configure {
  routing-options {
    if-attribute {
      srlg-group "red" {
        value 3
      }
    }
  }
}
```

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      admin-state enable
      srlg-database {
        router-id 192.0.2.4 {
          admin-state enable
          interface 192.168.45.1 srlg-group "red" { }
        }
        router-id 192.0.2.5 {
          admin-state enable
          interface 192.168.45.2 srlg-group "red" { }
        }
      }
    }
  }
  interface "int-PE-1-PE-4" {
    srlg-group "red" { }
  }
}
```

This information is local to PE-1 and will only have effect on CSPF calculations on PE-1, not on the other nodes.

When a CSPF calculation is done for a path from PE-1 to PE-5, the result will be two equal-cost paths, because ECMP equals 2. When adding the **srlg-group "red"** as a restriction, only a single path will be found, passing PE-2.

## Conclusion

Interpreting the SRLG information into the TE database makes it possible to protect an LSP even when multiple IP/MPLS interfaces fail as a result of an underlying transmission failure. Transmission failures can occur quite often because not all transmission links are one to one protected.

SRLG groups in MPLS provide a very dynamic and simple way to assure LSP FRR path protection on every PLR throughout the followed LSP. The SRLG groups are also taken into account when defining the ERO for secondary paths, at least if the configured secondary path is empty.

For interoperability reasons, the SRLG-database is available, because systems can link interfaces to an SRLG with interconnecting systems that do not support the SRLG feature; so they cannot advertise the SRLG information through the IGP.

The creation and maintenance of an SRLG database requires operational effort and systems that do not support SRLG will never take any SRLG information into account during CSPF calculation for the creation of FRR bypass or detour tunnels.

# Static Point-to-Point LSPs

This chapter provides information about static point-to-point label switched paths (LSPs).

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

## Applicability

This chapter is applicable to SR OS and was originally written for SR OS Release 7.0.R5. The MD-CLI in the current edition corresponds to SR OS Release 21.2.R1. There are no prerequisites or conditions on the hardware for this configuration.

## Overview

Due to the connectionless nature of the network layer protocol IP, packets travel through the network on a hop-by-hop basis with routing decisions made at each node. As a result, hyperaggregation of data on certain links may occur and it may impact the provider's ability to provide guaranteed service levels across the network end-to-end. To address these shortcomings, multiprotocol label switching (MPLS) was developed. MPLS provides the capability to establish connection-oriented paths, called label switched paths (LSPs), over a connectionless (IP) network.

The LSP offers a mechanism to engineer network traffic independently from the underlying network routing protocol (mostly IP) to improve the network resiliency and recovery options and to permit delivery of new services that are not readily supported by conventional IP routing techniques, such as Layer 2 IP Virtual Private Networks (VPNs). These benefits are essential for today's communication network explaining the wide deployment base of the MPLS technology.

RFC 3031, *Multiprotocol Label Switching Architecture*, specifies the MPLS architecture while this document describes the configuration and troubleshooting of static point-to-point LSPs on SR OS. Point-to-point LSPs can also be dynamically established using a label signaling protocol, such as label distribution protocol (LDP)—as described in chapter [LDP Point-to-Point LSPs](#)—or resource reservation protocol (RSVP)—as described in chapter [RSVP Point-to-Point LSPs](#).

## Packet forwarding

As a packet of a connectionless network layer protocol travels from one router to the next, each router in the network makes an independent forwarding decision by performing the following basic tasks: first analyzing the packet header, then referencing the local routing table to find the longest match based on the destination address in the IP header, and finally sending out the packet on the selected interface.

In other terms, the first function partitions the entire set of possible packets into a set of forwarding equivalence classes (FECs). All packets associated with a particular FEC will be forwarded along the same logical path to the same destination. The second function maps each FEC to a next hop destination router. Each router along the path performs these actions.

In MPLS, the assignment of a particular packet to a particular FEC is done just once, as the packet enters the network. In turn, the FEC is mapped to an LSP, which is established prior to any data flowing.

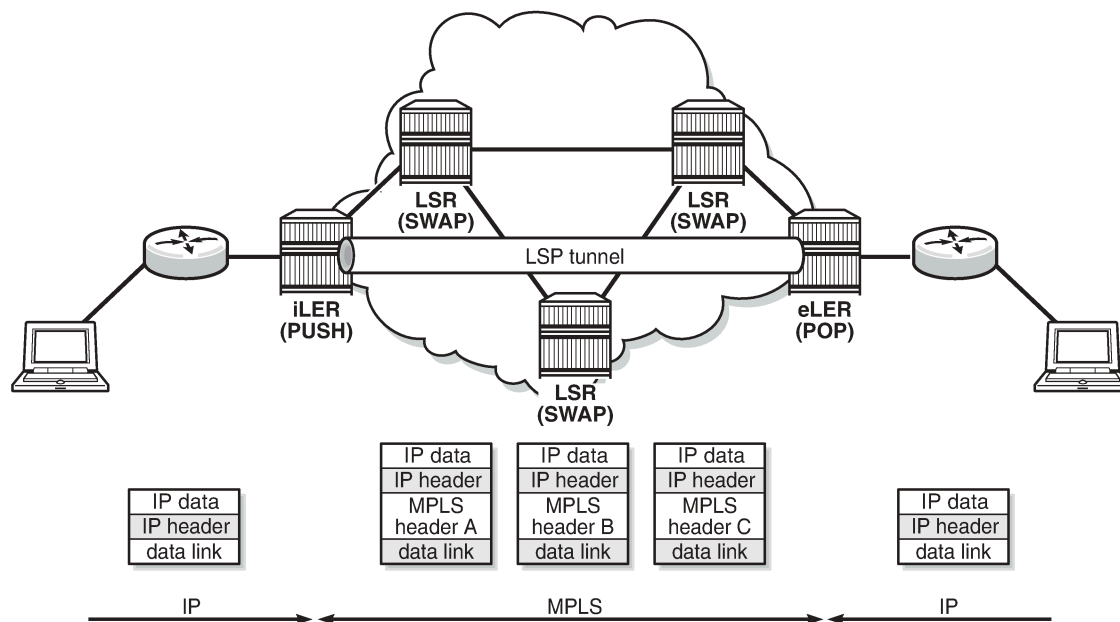
An MPLS label, representing the FEC to which the packet is assigned, is attached to the packet (push operation) and once labeled, the packet is forwarded to the next hop router along that LSP path.

At subsequent hops, there is no further analysis of the network layer header of the packet. Instead, the label is used as an index into a table which specifies the next hop and a new label. The old label is replaced with the new label (swap operation), and the packet is forwarded to its next hop.

At the MPLS network egress, the label is removed from the packet (pop operation). If this router is the final destination (based on the remaining packet), the packet is handed to the receiving application, such as a virtual private LAN service (VPLS). If this router is not the final destination of the packet, the packet will be sent into a new MPLS tunnel or forwarded by conventional IP forwarding toward the layer 3 destination.

## Terminology

Figure 123: Generic MPLS network, MPLS label operations



25762

Figure 123: Generic MPLS network, MPLS label operations shows a general network topology clarifying the MPLS-related terms. A Label Edge Router (LER) is a device at the edge of an MPLS network, with at least one interface outside the MPLS domain. A router is usually defined as an LER based on its position relative to a particular LSP. The MPLS router at the head-end of an LSP is called the ingress label edge router (iLER). The MPLS router at the tail-end of an LSP is called the egress label edge router (eLER).

The iLER receives unlabeled packets from outside the MPLS domain, then applies MPLS labels to the packets, and forwards the labeled packets into the MPLS domain. The eLER receives labeled packets from the MPLS domain, then removes the labels, and forwards unlabeled packets outside the MPLS domain.

The last LSR before the eLER can be configured with an implicit-null label (numeric value 3). This LSR will pop the outer label and send MPLS packets without an outer label to the eLER. This is known as Penultimate Hop Popping (PHP). A Label Switching Router (LSR) is a device internal to an MPLS network, with all interfaces inside the MPLS domain. These devices switch labeled packets inside the MPLS domain. In the core of the network, LSRs ignore the network layer (IP) header of the packet and simply forward the packet using the MPLS label swapping mechanism.

A single LSP is unidirectional. In common practice, because the bidirectional nature of most traffic flows is implied, the term LSP often is used to define the pair of LSPs that enable the bidirectional flow. For ease of terminology and discussion however, the LSP in this chapter is referred to as a single entity.

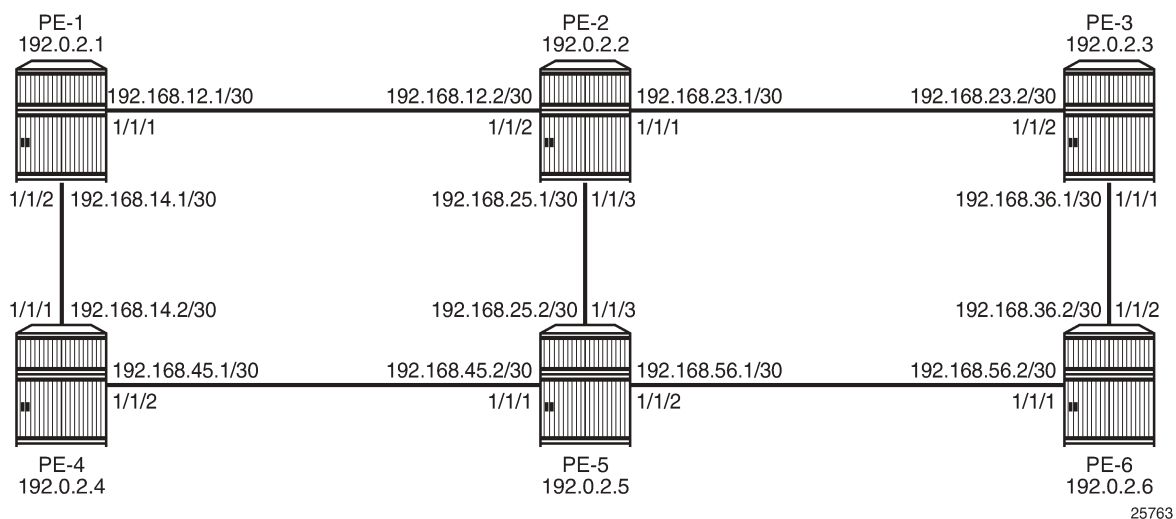
## LSP establishment

Prior to packet forwarding, the LSP must be established. In order to do so, labels need to be distributed for the path. For static LSPs, the label distribution is done manually by the network administrator. Although a high control level of the labels in use is achieved, the LSP cannot enjoy the resilience and recovery functionality the dynamic label signaling protocols can offer.

## Example topology

Figure 124: MPLS example topology shows the example topology consisting of six SR OS nodes located in a single autonomous system.

Figure 124: MPLS example topology



## Configuration

For static LSPs, there is no need for an IGP.

For LSPs that are set up manually, the first step is to enable MPLS and RSVP on all network interfaces that will be used to carry LSPs. On PE-1, the following interfaces are added to the MPLS and RSVP protocol instances:

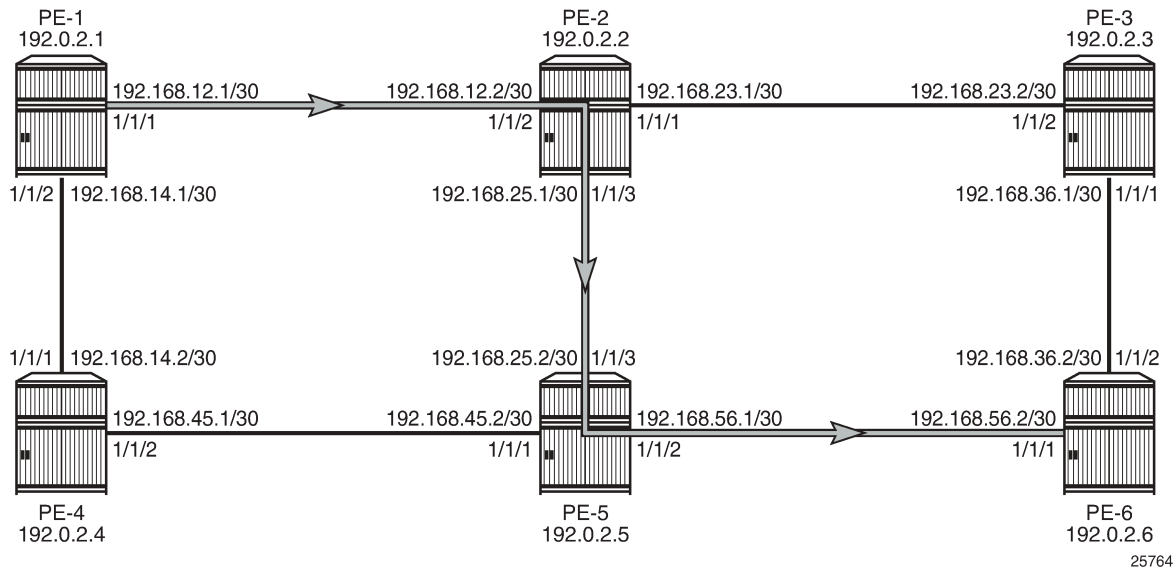
```
# on PE-1:configure {
  router "Base" {
    mpls {
      admin-state enable
      interface "int-PE-1-PE-2" {
      }
      interface "int-PE-1-PE-4" {
      }
    }
    rsvp {
      admin-state enable
      interface "int-PE-1-PE-2" {
      }
      interface "int-PE-1-PE-4" {
      }
    }
  }
}
```

For manually configured LSPs, any interface used by the static LSP must be added into the MPLS and RSVP protocol instances, even though RSVP is not actually used to signal labels. The following error messages are raised on PE-1 when attempting to add the interfaces into the MPLS protocol instance without configuring RSVP:

```
*[ex:configure router "Base"]
A:admin@PE-1# commit
MINOR: MGMT_CORE #4001: configure router "Base" mpls - RSVP needs to be present along with the
MPLS - configure router "Base" rsvp
MINOR: MGMT_CORE #4001: configure router "Base" mpls interface "system" - RSVP interface needs
to be present along with the MPLS interface - configure router "Base" rsvp interface "system"
MINOR: MGMT_CORE #4001: configure router "Base" mpls interface "int-PE-1-PE-2" - RSVP interface
needs to be present along with the MPLS interface - configure router "Base" rsvp interface
"int-PE-1-PE-2"
MINOR: MGMT_CORE #4001: configure router "Base" mpls interface "int-PE-1-PE-4" - RSVP interface
needs to be present along with the MPLS interface - configure router "Base" rsvp interface
"int-PE-1-PE-4"
```

As an example, a static LSP will be created starting from PE-1, running over PE-2 and PE-5, then terminating on PE-6 as shown in [Figure 125: Static LSP running over PE-1, PE-2, PE-5, PE-6](#).

Figure 125: Static LSP running over PE-1, PE-2, PE-5, PE-6



Verify the acceptable label range for use with static configurations for each node; as follows:

```
[/]
A:admin@PE-1# show router mpls-labels label-range

=====
Label Ranges
=====
Label Type      Start Label End Label  Aging    Available  Total
-----
Static          32          18431    -         18400     18400
Dynamic        18432       524287    0         505856    505856
  Seg-Route     0            0         -            0         0
=====
```

The label range for static LSPs extends from the value 32 to 18431. To ensure the labels have not yet been allocated to another configuration, use the command:

```
[/]
A:admin@PE-2# show router mpls-labels label 32 18431 in-use

=====
MPLS Labels from 32 to 18431 (In-use)
=====
Label           Label Type      Label Owner
-----
In-use labels (Owner: All) in specified range : 0
In-use labels in entire range                 : 0
=====
```

This command shows the number of incoming labels in use. At the iLER, the number of labels in use will remain 0 even after the static LSP has been configured where the iLER has a push operation for a label. The reason is that the labels shown are relevant to the labels that the router is generating, as for label



swap or pop operations. There is no information shown about labels that other routers are advertising. For the push operation, any label can be used, even if it is not within the label range of the router pushing the label. For the originating router PE-1, the label 100 will be used for the push operation on the interface toward PE-2.

Static LSPs are configured within the MPLS configuration context, but do not rely on dynamic label signaling.

The configuration of the MPLS static LSP head-end PE-1 contains:

- The system IP address of the destination router PE-6 (to).
- A push operation of the out-label 100.
- The interface address facing the current node of the next-hop along the static path, which is PE-2:

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      static-lsp "LSP-PE-1-PE-6-static" {
        admin-state enable
        to 192.0.2.6
        push {
          out-label 100
          next-hop 192.168.12.2
        }
      }
    }
  }
}
```

The transit LSRs PE-2 and PE-5 perform swap operations and forward the packet to the manually defined next-hop. On the LSR under the context of the interface on which the incoming LSP arrives, the correct label is selected (label-map) and in this context a swap operation with a new label and the new next-hop is entered. The MPLS configuration on LSR PE-2 is as follows:

```
# on PE-2:
configure {
  router "Base" {
    mpls {
      interface "int-PE-2-PE-1" {
        label-map 100 {
          admin-state enable
          swap {
            out-label 150
            next-hop 192.168.25.2
          }
        }
      }
    }
  }
}
```

The MPLS configuration on LSR PE-5 is as follows:

```
# on PE-5:
configure {
  router "Base" {
    mpls {
      interface "int-PE-5-PE-2" {
        label-map 150 {
          admin-state enable
          swap {
            out-label 200
          }
        }
      }
    }
  }
}
```



On LSR PE-5:

```
[/]
A:admin@PE-5# show router mpls static-lsp transit-terminate transit

=====
MPLS Static LSPs (Transit)
=====
In Label   In Port   Out Label  Out Port   Next Hop           Adm  Opr
-----
150        1/1/3    200        1/1/2      192.168.56.2      Up   Up
-----
LSPs : 1
=====
```

On the terminating router (eLER), the keyword transit-terminate terminate is added, as follows:

```
[/]
A:admin@PE-6# show router mpls static-lsp transit-terminate terminate

=====
MPLS Static LSPs (Terminate)
=====
In Label   In Port   Out Label  Out Port   Next Hop           Adm  Opr
-----
200        1/1/1    n/a        n/a        n/a                Up   Up
-----
LSPs : 1
=====
```

To track the label action associated with the static LSP configuration, the **show router mpls interface label-map** command can be used on all LSRs and eLERs, but not on the iLER.

```
[/]
A:admin@PE-2# show router mpls interface label-map

=====
MPLS Interfaces (Label-Map)
=====
In Label   In I/F    Out Label  Out I/F    Next Hop           Type  Adm  Opr
-----
100        1/1/2    150        1/1/3      192.168.25.2      Static Up   Up
-----
Interfaces : 1
=====
```

```
[/]
A:admin@PE-6# show router mpls interface label-map

=====
MPLS Interfaces (Label-Map)
=====
In Label   In I/F    Out Label  Out I/F    Next Hop           Type  Adm  Opr
-----
200        1/1/1    n/a        n/a        n/a                Static Up   Up
-----
Interfaces : 1
=====
```

The **show router mpls status** command is used to verify each of the LSP types, the number of configured LSPs and whether they originate on, transit through or terminate on the router.

```
[/]
A:admin@PE-1# show router mpls status
=====
MPLS Status
=====
Admin Status           : Up
Oper(V4) State         : Up           Oper(V6) State           : Down
IPv4 Oper Down Reason  : n/a
IPv6 Oper Down Reason  : ipv6TeRtrDown
FRR Object             : Enabled   Resignal Timer           : Disabled
Hold Timer             : 1 seconds Next Resignal            : N/A
Srlg Frr               : Disabled  Srlg Frr Strict          : Disabled
Admin Group Frr       : Disabled
Dynamic Bypass         : Enabled   User Srlg Database       : Disabled
BypassResignalTimer   : Disabled BypassNextResignal      : N/A
LeastFill Min Thd     : 5 percent LeastFill Reopti Thd    : 10 percent
Local TTL Prop        : Enabled   Transit TTL Prop        : Enabled
AB Sample Multiplier  : 1        AB Adjust Multiplier     : 288
Exp Backoff Retry     : Disabled  CSPF On Loose Hop       : Disabled
Lsp Init RetryTimeout : 30 seconds MBB Pref Current Hops   : Disabled
Logger Event Bundling : Disabled
Retry on IGP Overload : Disabled  Resignal on IGP Overload : Disabled
P2mp Resignal Timer   : Disabled  P2mp Next Resignal      : N/A
Sec FastRetryTimer    : Disabled  Static LSP FR Timer     : 30 seconds
P2P Max Bypass Association: 1000
Max Bypass PLR Association: 16
P2PActPathFastRetry   : Disabled  P2MP S2L Fast Retry     : Disabled
In Maintenance Mode   : No
MplsTp                : Disabled
Next Available Lsp Index : 2
Entropy Label RSVP-TE : Enabled   Entropy Label SR-TE     : Enabled
PCE Report RSVP-TE    : Disabled  PCE Report SR-TE       : Disabled
PCE Init LSP          : Disabled
SR-TE Resignal Timer  : Disabled  SR-TE Next Resignal     : N/A
SR-TE Resig on IGP Event : Disabled
LSP Self Ping Timeout : 300 seconds LSP Self Ping Interval  : 1 seconds
RSVP-TE LSP Self Ping : Disabled  Self Ping Timeout Action : retry
=====
MPLS LSP Count
=====
-----

```

	Originate	Transit	Terminate
Static LSPs	1	0	0
Dynamic LSPs	0	0	0
P2P LSPs	0	N/A	N/A
Detour LSPs	0	0	0
P2MP S2Ls	0	0	0
MPLS-TP LSPs	0	0	0
Mesh-P2P LSPs	0	N/A	N/A
One Hop-P2P LSPs	0	N/A	N/A
SR-TE LSPs	0	N/A	N/A
Mesh-P2P SR-TE LSPs	0	N/A	N/A
One Hop-P2P SR-TE LSPs	0	N/A	N/A
PCE Init SR-TE LSPs	0	N/A	N/A

```
=====
```

Penultimate Hop Popping (PHP) can be used with static LSPs. This is achieved by configuring the last LSR PE-5 before the eLER PE-6 to swap the incoming label to implicit-null instead of a specific label value, as follows:

```
# on PE-5:
configure {
  router "Base" {
    mpls {
      interface "int-PE-5-PE-2" {
        label-map 150 {
          swap {
            out-label implicit-null-label
            next-hop 192.168.56.2
          }
        }
      }
    }
  }
}
```

The previous configuration will cause PE-5 to pop the top label from the incoming labeled frame received from PE-2 and send it to PE-6 without adding another outer label. The following command shows out label 3, but label 3 is never actually pushed onto a frame.

```
[/]
A:admin@PE-5# show router mpls static-lsp transit-terminate transit
=====
MPLS Static LSPs (Transit)
=====
In Label   In Port   Out Label  Out Port   Next Hop           Adm  Opr
-----
150        1/1/3    3          1/1/2     192.168.56.2      Up   Up
-----
LSPs : 1
=====
```

If the traffic arriving at PE-5 were IP with a single label, then it would arrive at PE-6 as unlabeled IP traffic.

If the static LSP spans a single hop, for example, from PE-1 to PE-2, the ingress LER PE-1 pushes the implicit-null instead of a label. The configuration on PE-1 is as follows:

```
# on PE-1:
configure {
  router "Base" {
    mpls {
      static-lsp "LSP-PE-1-PE-2-static" {
        admin-state enable
        to 192.0.2.2
        push {
          out-label implicit-null-label
          next-hop 192.168.12.2
        }
      }
    }
  }
}
```

In this case, no MPLS action (swap or pop) is required for this LSP on PE-2.

## Conclusion

MPLS provides the capability to establish connection-oriented paths over a connectionless network. The static LSP offers a mechanism to engineer network traffic. In this chapter, the configuration of static LSPs is given together with the associated show output which can be used to verify and troubleshoot.

# Topology-Independent Loop-Free Alternate for Link Protection

This chapter describes the Topology-Independent Loop-Free Alternate for Link Protection.

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

## Applicability

This chapter was initially written based on SR OS Release 16.0.R5, but the MD-CLI in the current edition corresponds to SR OS Release 21.2.R1. Topology-Independent Loop-Free Alternate (TI-LFA) is supported from SR OS Release 15.0.R1 for IS-IS and 15.0.R4 for OSPF.

## Overview

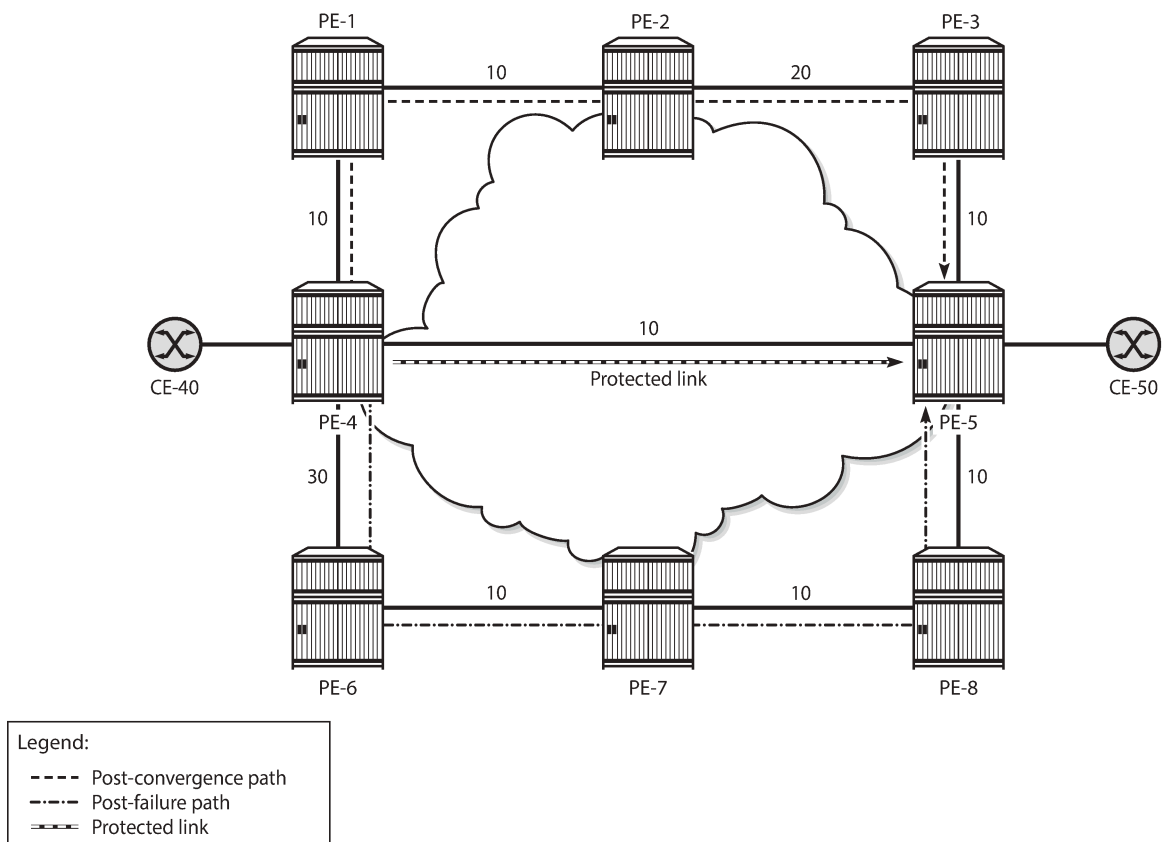
For IP Fast Reroute (FRR), the routers use a precomputed Loop-Free Alternate (LFA) next-hop installed in the FIB until the Shortest Path First (SPF) algorithm runs and the network converges again. The following LFA modes can be applied:

- Regular LFA installs an alternate next-hop in the FIB. Regular LFA provides protection for native IP traffic as well as for Segment Routing (SR) and LDP traffic.
- Remote LFA uses a repair tunnel to a PQ node, which is a node where traffic is not looped back toward the computing node. Remote LFA provides protection for SR and LDP traffic, not for native IP traffic.
- If a computing router has multiple backup next-hop routers, TI-LFA creates a repair tunnel on the post-convergence path so that the post-failure next-hop is avoided, if different from the post-convergence next-hop. In this case, traffic will not be dropped after SPF converges. TI-LFA extends the remote LFA algorithm by computing a backup tunnel where the P and Q nodes do not coincide. TI-LFA uses a repair tunnel to the closest Q node on the post-convergence path. This repair tunnel uses the shortest path to the P node and a source-routed path from the P node to the Q node. TI-LFA provides protection for SR and LDP traffic, not for native IP traffic.

Regular LFA is described in chapter MPLS LDP FRR using ISIS as IGP. Remote LFA and TI-LFA use segment routing to create repair tunnels in cases where there is no regular LFA backup.

[Figure 126: Post-failure LFA path does not match post-convergence path](#) shows the example topology where traffic flows from CE-40 toward CE-50, and a post-failure LFA path that does not match the post-convergence path.

Figure 126: Post-failure LFA path does not match post-convergence path



29352

During normal operation, traffic goes from CE-40 to PE-4 and straight on to PE-5 and CE-50. This is the shortest path between CE-40 and CE-50. Consider the failure of the link between PE-4 and PE-5. This is the protected link. If a failure occurs on the protected link between PE-4 and PE-5, there are two possible backup next-hops from computing node PE-4: PE-1 or PE-6.

When enabling regular LFA on PE-4, two consecutive failovers will occur: the first one, nearly instantaneously, from the preferred path (optimum distance) to the precomputed post-failure path via next-hop PE-6 and the second one, after SPF has run again, from the post-failure path to the post-convergence path via PE-1. When enabling TI-LFA, a single failover will occur, so the computed post-failure path must match the post-convergence path.

The post-convergence path will be from PE-4 to PE-1, PE-2, PE-3, and PE-5, with a path cost of  $10 + 10 + 20 + 10 = 50$ . With regular LFA, the post-failure path should not use PE-1 as next-hop, because PE-1 would loop back traffic to reach PE-5 via PE-4, through the protected link (which is not allowed).

As described in RFC 5286, the following inequality 1 for link protection must be true for a neighbor next-hop (NH) to provide an LFA. The cost is the optimum distance between the nodes:

$$\text{cost}(\text{NH}, \text{Destination}) < \text{cost}(\text{NH}, \text{Source}) + \text{cost}(\text{Source}, \text{Destination})$$

For next-hop PE-1, the following LFA inequality 1 is false on the calculating node PE-4, indicating that no regular LFA path is possible via PE-1:



$$\text{cost}(PE-1,PE-5) < \text{cost}(PE-1,PE-4) + \text{cost}(PE-4,PE-5)$$

$$(10 + 10) < 10 + 10 \rightarrow \text{False}$$

For next-hop PE-6, the following LFA inequality 1 is true on the calculating node PE-4, indicating that a regular LFA path is possible via PE-6:

$$\text{cost}(PE-6,PE-5) < \text{cost}(PE-6,PE-4) + \text{cost}(PE-4,PE-5)$$

$$(10 + 10 + 10) < 30 + 10 \rightarrow \text{True}$$

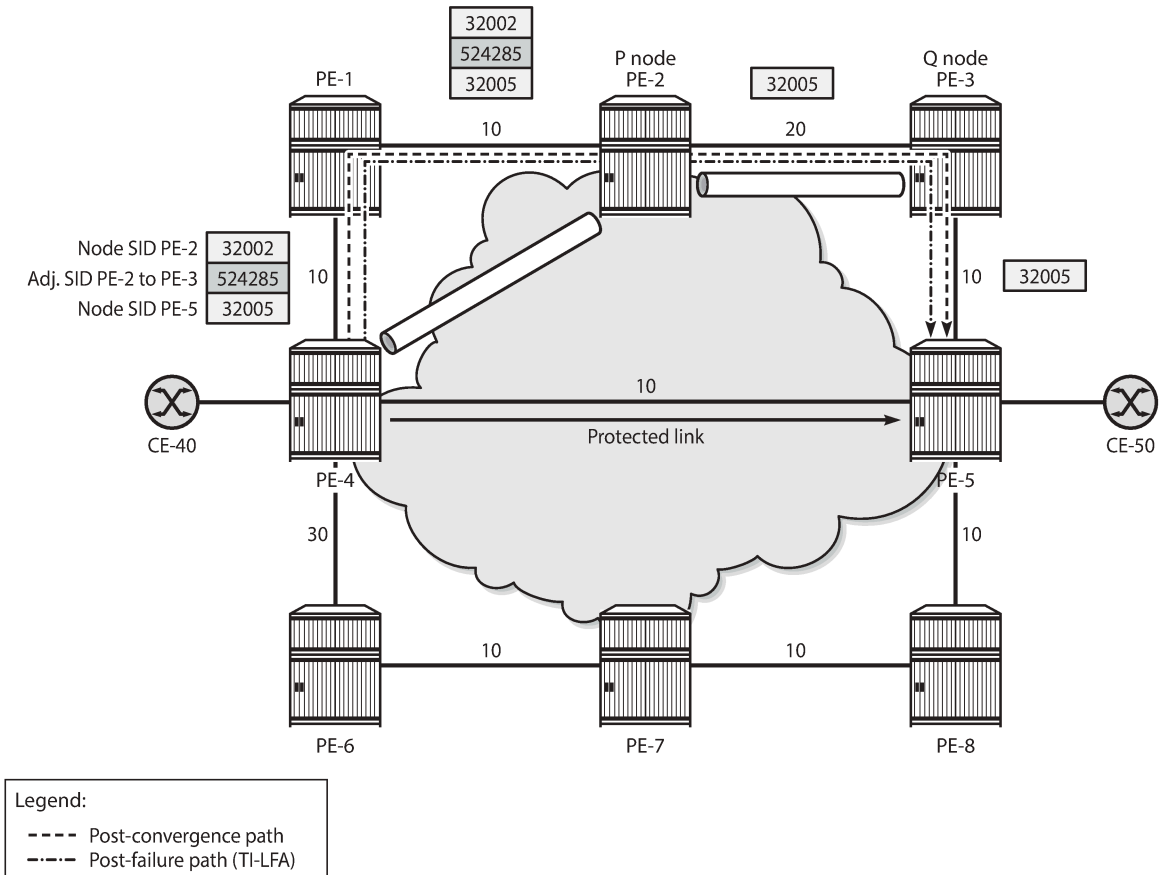
Because of the higher metric between PE-4 and PE-6 (30), PE-6 will not loop back traffic via PE-4: the path cost from PE-6 to PE-5 via PE-4 = 30 + 10 = 40, while the path cost from PE-6 to PE-5 via PE-7 and PE-8 = 10 + 10 + 10 = 30. So, PE-6 will forward the traffic to PE-7, PE-8, and PE-5.

For these reasons, the post-failure path uses PE-6 as regular LFA next-hop.

TI-LFA ensures that traffic is forwarded in a tunnel to the closest Q node, where it will not be looped back to PE-4. In this example, PE-3 is the Q node and it is one hop away from P node PE-2.

With TI-LFA enabled, additional labels are pushed to ensure that the post-failure next-hop matches the post-convergence next-hop. When the protected link between PE-4 and PE-5 fails, PE-4 pushes the node SID of PE-2 as top label plus the adjacency SID of the PE-2 to PE-3 link as an extra label. The bottom label is the node SID of the destination PE-5, which is present in any packet to PE-5 (located on the primary path); see [Figure 127: Post-failure TI-LFA path matches post-convergence path](#).

Figure 127: Post-failure TI-LFA path matches post-convergence path



29353

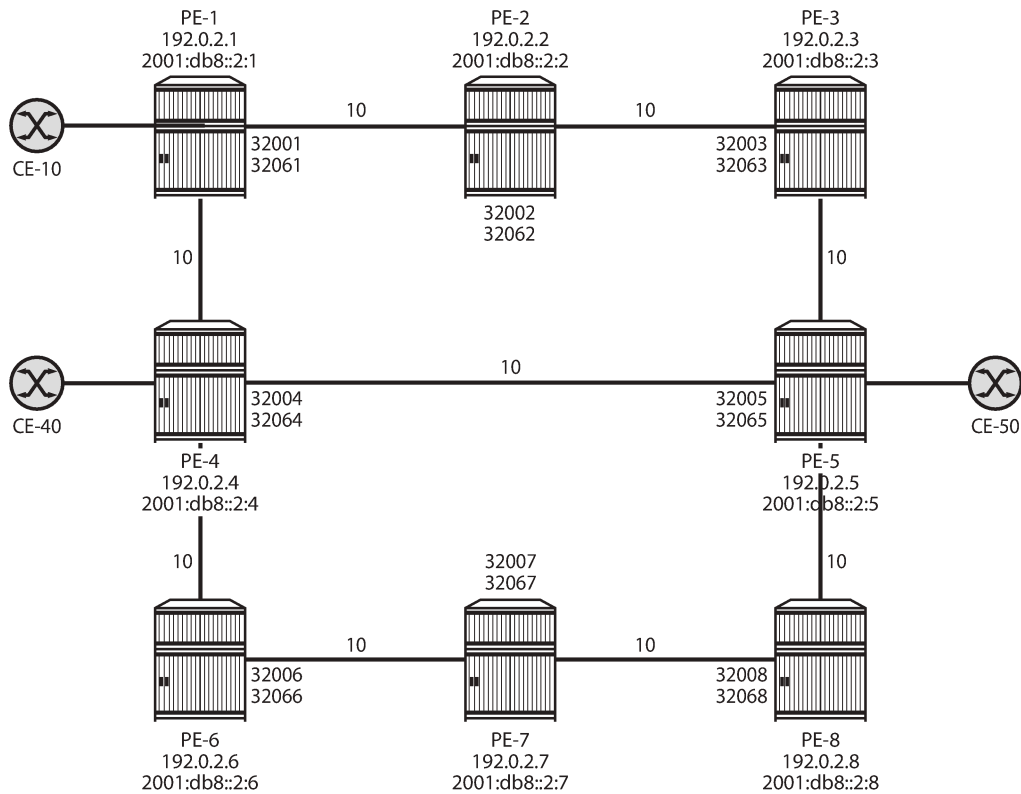
In this chapter, the following LFA modes are described and configured:

- Regular LFA
- Remote LFA
- TI-LFA

## Configuration

Figure 128: Example topology shows the example topology, but that will be reduced in the first two scenarios. The default metric of all links is 10, but that may be configured with a different value afterward.

Figure 128: Example topology



29354

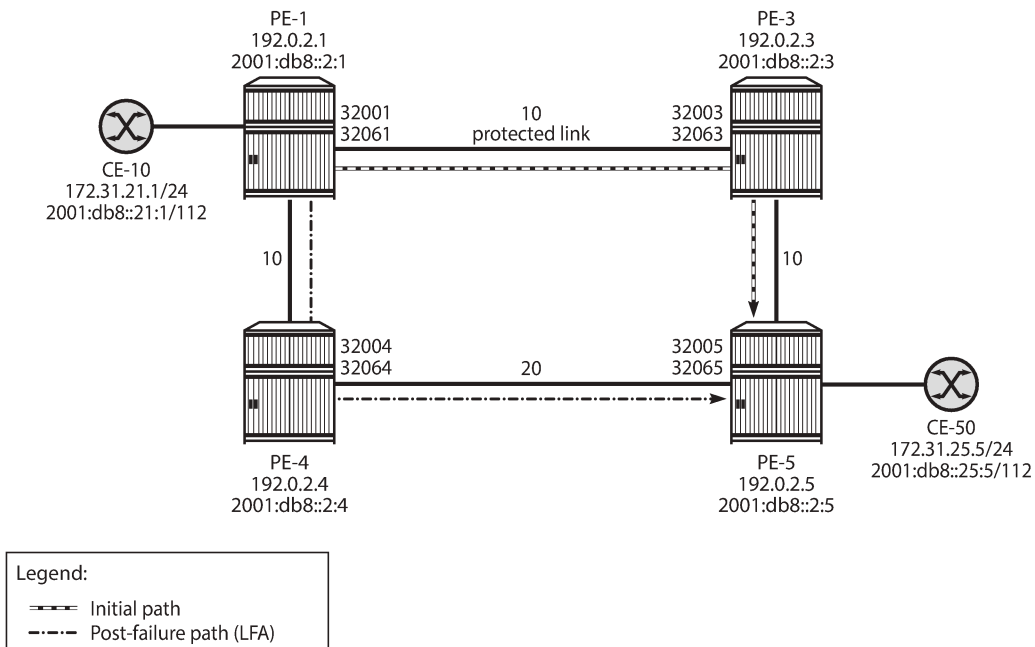
The initial configuration includes the following:

- Cards, MDAs, ports
- Dual-stack router interfaces (IPv4/IPv6)
- IS-IS as IGP on the router interfaces. The metric is 10, but that may be configured otherwise.
- Segment routing (SR-ISIS) with node SIDs 3200x for IPv4 and 3206x for IPv6 system addresses.

## Regular LFA

Figure 129: Example topology with regular LFA configured on PE-4 shows the example topology reduced to four PEs. Without a failure of the protected link, traffic from CE-10 to CE-50 is sent via PE-3. The protected link is the link between PE-1 and PE-3 and the LFA path after failure goes via next-hop PE-4.

Figure 129: Example topology with regular LFA configured on PE-4



29355

The IGP metric on the interface between PE-4 and PE-5 is 20, as follows:

```
# on PE-4:
configure {
  router "Base" {
    isis 0 {
      interface "int-PE-4-PE-5" {
        level 1 {
          metric 20
        }
        level 2 {
          metric 20
        }
      }
    }
  }
}
```

```
#on PE-5:
configure {
  router "Base" {
    isis 0 {
      interface "int-PE-5-PE-4" {
        level 1 {
          metric 20
        }
        level 2 {
          metric 20
        }
      }
    }
  }
}
```

Regular LFA is configured on the nodes, as follows:

```
# on PE-1, PE-3, PE-4, PE-5:
configure {
  router "Base" {
```

```
isis 0 {
    loopfree-alternate {
    }
}
```

In the normal situation, without failures, the preferred traffic path from CE-10 to CE-50 is via PE-1, PE-3, and PE-5 with a cost (optimum distance) of 10 + 10 = 20. When the link between PE-1 and PE-3 fails, the post-failure LFA path is via PE-1, PE-4, and PE-5 with a cost of 10 + 20 = 30. The following LFA inequality 1 is true, so PE-4 is a valid LFA next-hop:

$$\text{cost}(\text{newNH}, \text{Destination}) < \text{cost}(\text{newNH}, \text{Source}) + \text{cost}(\text{Source}, \text{Destination})$$

$$\text{cost}(\text{PE-4}, \text{PE-5}) < \text{cost}(\text{PE-4}, \text{PE-1}) + \text{cost}(\text{PE-1}, \text{PE-5})$$

$$20 < 10 + (10 + 10) \rightarrow \text{True}$$

The route table on PE-1 for prefix 192.0.2.5 shows that the next-hop is 192.168.13.2 on PE-3 for the preferred path with metric 20; the LFA next-hop is 192.168.14.2 on PE-4 for the post-failure path with metric 30, as follows:

```
[/]
A:admin@PE-1# show router route-table 192.0.2.5 alternative
=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]                Type  Proto  Age           Pref
  Next Hop[Interface Name]
  Alt-NextHop                      Metric
                                   Alt-
                                   Metric
-----
192.0.2.5/32                      Remote ISIS  00h33m04s  15
  192.168.13.2                    20
  192.168.14.2 (LFA)              30
-----
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 FP tunnel table on PE-1 shows the SR-ISIS label 32005, which is the node SID of PE-5 for prefix 192.0.2.5/32. The same label 32005 is used for the LFA post-failure path indicated with (B) for FRR backup.

```
[/]
A:admin@PE-1# show router fp-tunnel-table 1 192.0.2.5/32
=====
IPv4 Tunnel Table Display
Legend:
label stack is ordered from bottom-most to top-most
B - FRR Backup
=====
Destination                Protocol  Tunnel-ID
  Lbl
  NextHop                   Intf/Tunnel
  Lbl      (backup)
```

```

      NextHop   (backup)
-----
192.0.2.5/32          SR-ISIS-0          524301
  32005
    192.168.13.2          1/1/3:1000
  32005
    192.168.14.2(B)      1/1/2:1000
-----
Total Entries : 1
=====
    
```

The following FP tunnel table on PE-1 shows the SR-ISIS label 32065, which is the node SID of PE-5 for prefix 2001:db8::2:5/128. The same label 32065 is used for the LFA post-failure path.

```

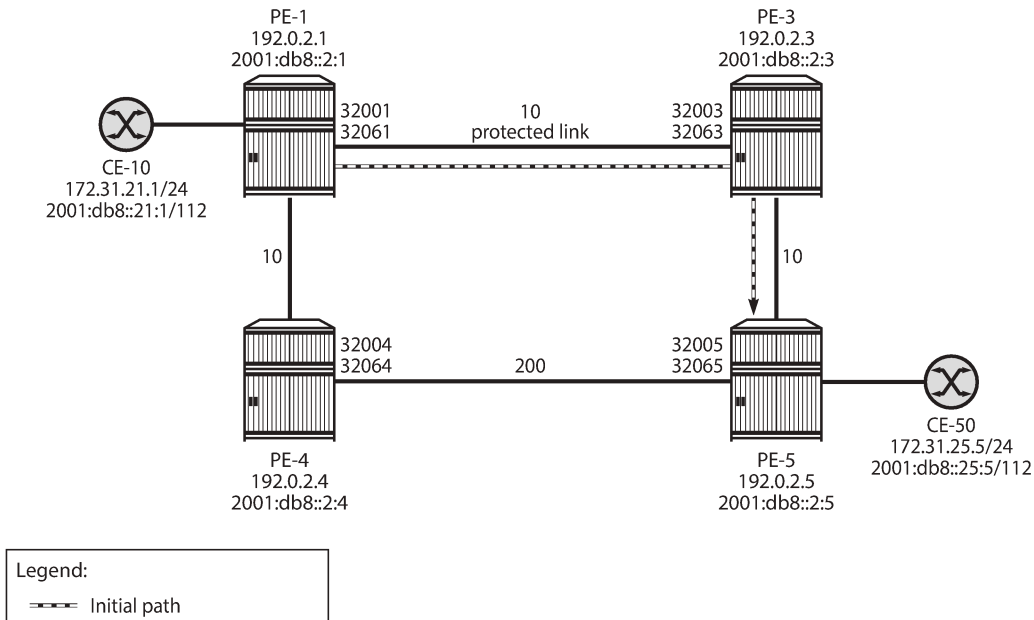
[/]
A:admin@PE-1# show router fp-tunnel-table 1 2001:db8::2:5/128

=====
IPv6 Tunnel Table Display

Legend:
Label stack is ordered from bottom most to top-most
B - FRR Backup
=====
Destination          Protocol          Tunnel-ID
Lbl
  NextHop
Lbl   (backup)
  NextHop   (backup)
-----
2001:db8::2:5/128    SR-ISIS-0          524302
  32065
    fe80::618:1ff:fe01:3-"int-PE-1-PE-3"  1/1/3:1000
  32065
    fe80::61c:1ff:fe01:1-"int-PE-1-PE-4" (B)  1/1/2:1000
-----
Total Entries : 1
=====
    
```

Figure 130: No post-failure LFA path when PE-4 loops back traffic shows that no backup LFA next-hop exists when the metric on the interface between PE-4 and PE-5 is increased to 200.

Figure 130: No post-failure LFA path when PE-4 loops back traffic



29356

The following configures the metric on the interface between PE-4 and PE-5 to a value of 200:

```
# on PE-4:
configure {
  router "Base" {
    isis 0 {
      interface "int-PE-4-PE-5" {
        level 1 {
          metric 200
        }
        level 2 {
          metric 200
        }
      }
    }
  }
}
```

```
# on PE-5:
configure {
  router "Base" {
    isis 0 {
      interface "int-PE-5-PE-4" {
        level 1 {
          metric 200
        }
        level 2 {
          metric 200
        }
      }
    }
  }
}
```

When the metric on the interface between PE-4 and PE-5 is increased to a value that exceeds the sum of the metrics on the path from PE-4 to PE-1 and the path from PE-1 to PE-5 (via PE-3), the computing node PE-1 cannot calculate a regular LFA path to protect the PE-5 prefixes. The following LFA inequality 1 is false:

$$\text{cost}(PE-4,PE-5) < \text{cost}(PE-4,PE-1) + \text{cost}(PE-1,PE-5)$$

200 < 10 + (10 + 10) --> False

If the preferred path cannot be used because of a failure, such as a link failure between PE-1 and PE-3, a micro-loop is created between PE-4 and PE-5 until convergence is completed. The following output shows that no LFA next-hop is available on PE-1:

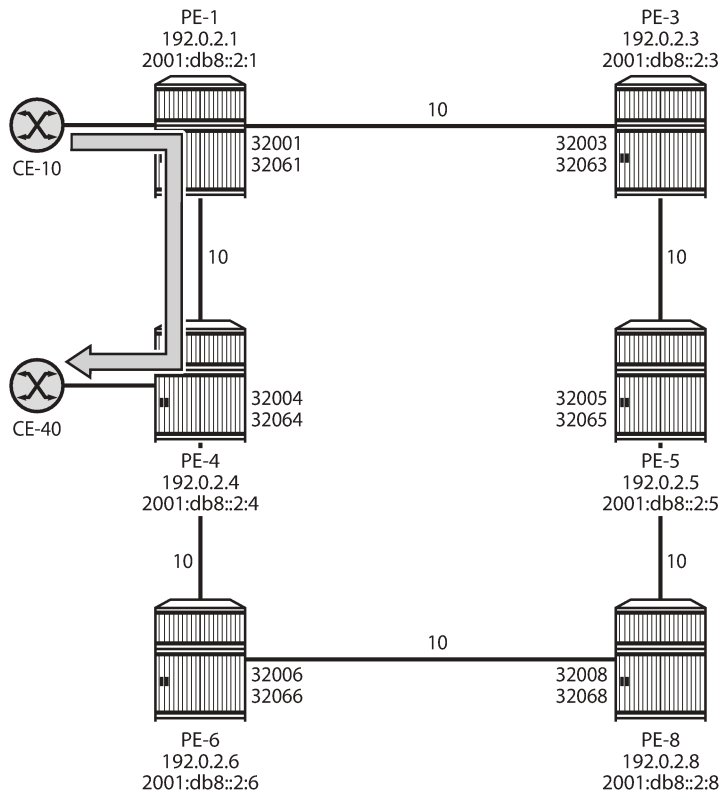
```
[/]
A:admin@PE-1# show router route-table 192.0.2.5 alternative
=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]                Type  Proto  Age      Pref
  Next Hop[Interface Name]          Alt-NextHop      Metric  Alt-
                                         Metric
-----
192.0.2.5/32                        Remote ISIS    00h05m26s 15
  192.168.13.2                       20
-----
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
=====
```

## Remote LFA

[Figure 131: Example topology for remote LFA](#) shows the example topology with six nodes in a ring. Traffic from CE-10 to CE-40 is preferably sent via PE-1 to PE-4.



Figure 131: Example topology for remote LFA



29357

The following command enables remote LFA on all nodes:

```
# on PE-1, PE-3, PE-4, PE-5, PE-6, PE-8:
configure {
  router "Base" {
    isis 0 {
      loopfree-alternate {
        remote-lfa {
        }
      }
    }
  }
}
```

If the link between PE-1 and PE-4 fails, the repair path on PE-1 can only use PE-3 as next-hop. Link-protection LFA inequality 1 is not valid, indicating that the backup path via next-hop PE-3 is not loop free:

$$\text{cost}(PE-3, PE-4) < \text{cost}(PE-3, PE-1) + \text{cost}(PE-1, PE-4)$$

$$(10 + 10) < 10 + 10 \rightarrow \text{False}$$

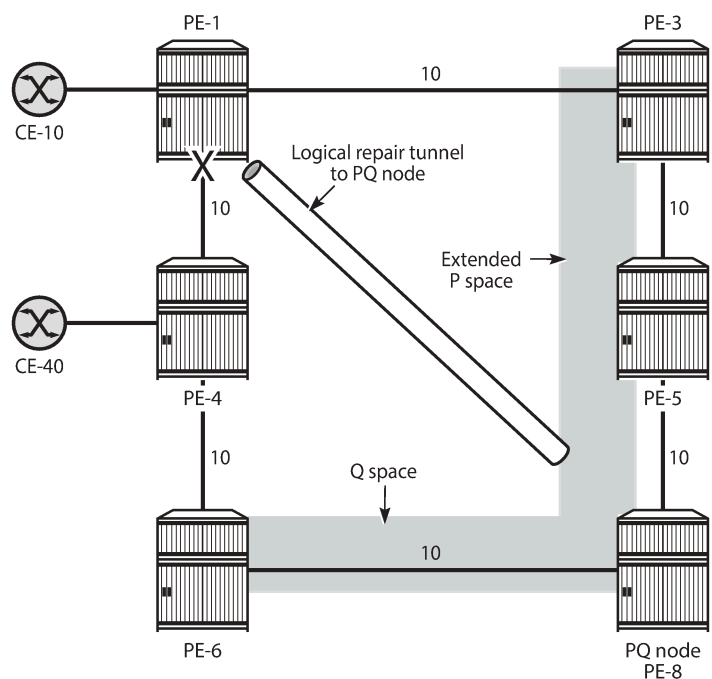
With this invalid LFA inequality 1, no coverage with regular LFA is possible. When remote LFA is enabled, a repair tunnel is computed from PE-1 toward a node (PE-8) where the traffic is not looped back toward the computing node PE-1. When traffic emerges from the repair tunnel on PE-8, it is forwarded to the destination PE-4, using node SID 32004 for IPv4 or node SID 32064 for IPv6.

The endpoint node of the repair tunnel for remote LFA (RLFA) is the PQ node, which is in the intersection of the extended P space of source PE-1 and the Q space of destination PE-4.

- The P space of PE-1 is the set of routers reachable on the shortest SPF path from the computing node PE-1, without using the protected link between PE-1 and PE-4; that is, SPF computed by PE-1 and rooted from PE-1. In this example, PE-3 and PE-5 are in the P space of PE-1.
- The extended P space of PE-1 is the set of routers, calculated by PE-1, in the P space of the next-hop router PE-3. An additional SPF computation by PE-1 and rooted from PE-3 results in P nodes PE-3, PE-5, and PE-8. The extended P space increases the repair coverage.
- The Q space of PE-4 is the set of routers that can reach the destination router PE-4 using the shortest path, without using the protected link; that is, reverse SPF computed by PE-1 and rooted from PE-4, resulting in Q nodes PE-6 and PE-8.
- PQ routers are in the intersection of the extended P space and the Q space; in this case, the only PQ node is PE-8.
- Repair tunnels are shortest path SR tunnels from the computing node PE-1 to the PQ router; in this case, from PE-1 to PE-8.

Figure 132: PQ node in remote LFA shows the extended P space of PE-1, comprising nodes PE-3, PE-5, and PE-8, and the Q space of PE-4, comprising nodes PE-6 and PE-8. In the event of a link failure, PE-1 will push the node SID of PE-8, along with the node SID of PE-4, and forward the packet toward the backup next-hop PE-8.

Figure 132: PQ node in remote LFA



The following shows the SR LFA coverage on PE-1; the five other node SIDS are all protected: one with regular LFA and the remaining four with remote LFA (in the column RLFA). Besides the node SIDs, the

adjacency SIDs toward the direct neighbors PE-3 and PE-4 are protected using RLFA. The LFA coverage is the same for IPv4 and IPv6.

```
[/]
A:admin@PE-1# show router isis sr-lfa-coverage

=====
Rtr Base ISIS Instance 0 SR LFA Coverage
=====
MT-ID  SidType      Level Proto LFA      RLFA      TILFA      Coverage
-----
0      node-sid     L1    ipv4  1(20%)  4(80%)  0(0%)      5/5(100%)
0      node-sid     L1    ipv6  1(20%)  4(80%)  0(0%)      5/5(100%)
---snip---
0      adj-sid     L1L2  ipv4  0(0%)   2(100%) 0(0%)      2/2(100%)
0      adj-sid     L1L2  ipv6  0(0%)   2(100%) 0(0%)      2/2(100%)
=====
```

The repair tunnel from PE-1 to PQ node PE-8 uses node SID 32008 for IPv4 and 32068 for IPv6.

The fifth entry in the following FP tunnel table shows that destination 192.0.2.8/32 of PE-8 is protected with regular LFA. The only label is 32008, which is the node SID of PE-8. All other destinations in the table are protected with remote LFA, having two node SID labels for the RLFA path, such as 32004/32008 for prefix 192.0.2.4 with next-hop 192.168.13.2 on PE-3. This means that the top label 32008 is pushed by PE-1 to match the repair-tunnel going via PE-3 to PQ-node PE-8. From PE-8 onward, the bottom label 32004 is used toward PE-4. Likewise, the other destinations in the list have top label 32008, so a tunnel is established to PE-8. The output is similar for IPv6.

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

=====
IPv4 Tunnel Table Display

Legend:
label stack is ordered from bottom-most to top-most
B - FRR Backup
=====
Destination                                Protocol      Tunnel-ID
Lbl
NextHop                                     Intf/Tunnel
Lbl      (backup)
NextHop  (backup)
-----
192.0.2.3/32                                SR-ISIS-0    524295
32003
192.168.13.2                                1/1/3:1000
32003/32008
192.168.14.2(B)                             1/1/2:1000
192.0.2.4/32                                SR-ISIS-0    524299
32004
192.168.14.2                                1/1/2:1000
32004/32008
192.168.13.2(B)                           1/1/3:1000
192.0.2.5/32                                SR-ISIS-0    524301
32005
192.168.13.2                                1/1/3:1000
32005/32008
192.168.14.2(B)                             1/1/2:1000
192.0.2.6/32                                SR-ISIS-0    524311
32006
```

```

192.168.14.2          1/1/2:1000
32006/32008
192.168.13.2(B)     1/1/3:1000
192.0.2.8/32        SR-ISIS-0  524312
32008
192.168.13.2          1/1/3:1000
32008
192.168.14.2(B)     1/1/2:1000
192.168.13.2/32     SR          524309
3
192.168.13.2          1/1/3:1000
32003/32008
192.168.14.2(B)     1/1/2:1000
192.168.14.2/32     SR          524297
3
192.168.14.2          1/1/2:1000
32004/32008
192.168.13.2(B)     1/1/3:1000
-----
Total Entries : 7
-----
=====
    
```

## TI-LFA

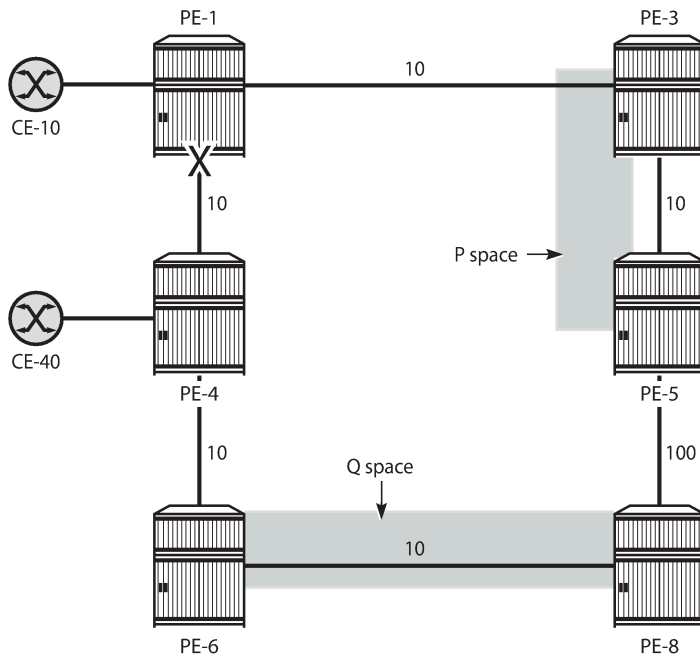
The following two use cases are described in this section:

- Directed LFA where the extended P space and the Q space do not overlap
- Extension of the RLFA algorithm to compute a repair path using directed LFA, but ensuring that the post-failure path matches the post-convergence path

## Directed LFA

Figure 133: [Extended P space of PE-1 and Q space of PE-4 are one hop apart](#) shows the example topology with increased metric between PE-5 and PE-8, reducing the extended P space to PE-3 and PE-5, so there is no PQ node.

Figure 133: Extended P space of PE-1 and Q space of PE-4 are one hop apart



29359

There is no remote LFA repair tunnel. No Q routers are on the shortest path from the computing router, and the P routers are not in the reverse SPF of the endpoint of the protected link. However, TI-LFA can calculate a repair tunnel in case the gap is only one or two hops. TI-LFA is enabled using the following command:

```
# on PE-1, PE-3, PE-4, PE-5, PE-6, PE-8:
configure {
  router "Base" {
    isis 0 {
      loopfree-alternate {
        ti-lfa {
          max-sr-frr-labels 2
        }
      }
    }
  }
}
```

Table 6: Values of the `max-sr-frr-labels` parameter in TI-LFA lists the possible values of the `max-sr-frr-labels` parameter. This parameter is used to specify the maximum number of labels that the TI-LFA backup next-hop can use.

Table 6: Values of the `max-sr-frr-labels` parameter in TI-LFA

Max. SR-FRR labels	LFA behavior
0	Regular LFA: TI-LFA backup restricted to next-hop that does not require a repair tunnel, so PQ node is a neighbor of the computing node.
1	Remote LFA: extended P space and Q space intersect and the repair tunnel requires 1 FRR label:

Max. SR-FRR labels	LFA behavior
	<ul style="list-style-type: none"> <li>Node SID to PQ router</li> </ul>
2 (default)	TI-LFA with extended P space and Q space one hop apart: <ul style="list-style-type: none"> <li>Node SID to P router</li> <li>Adjacency SID on P router to Q router</li> </ul>
3	TI-LFA with extended P space and Q space two hops apart: <ul style="list-style-type: none"> <li>Node SID to P router</li> <li>Two adjacency SIDs to Q router</li> </ul>

In this case, the extended P space and the Q space are one hop apart and TI-LFA calculates a post-failure path that consists of a repair tunnel to P router PE-5 (node SID 32005 for IPv4) and an adjacency SID toward Q router PE-8. For routes from PE-1 to PE-4, the LFA route has two additional labels combined with the bottom label that is the node SID of PE-4 (32004), which is also used for the primary path. The top label is the node SID of P router PE-5 (32005); the next label is the adjacency SID on PE-5 toward PE-8 (524285).

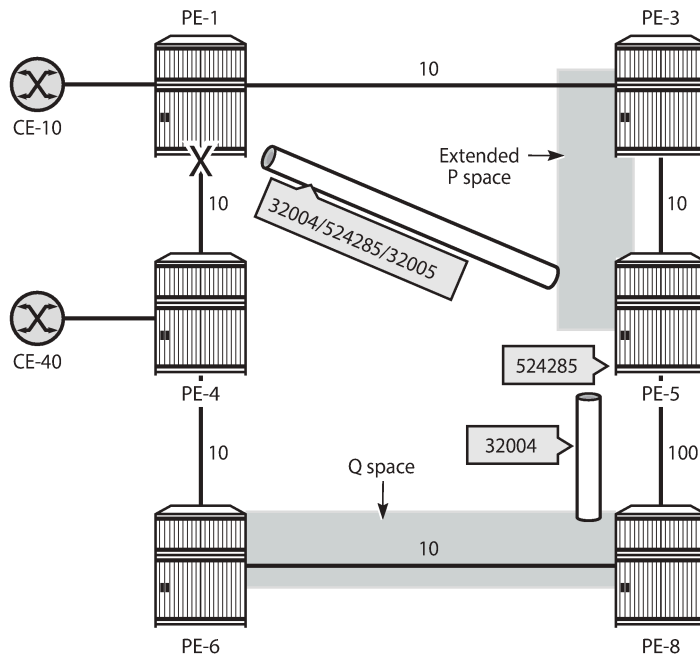
```
[/]
A:admin@PE-1# show router fp-tunnel-table 1 192.0.2.4/32

=====
IPv4 Tunnel Table Display

Legend:
label stack is ordered from bottom-most to top-most
B - FRR Backup
=====
Destination                                Protocol      Tunnel-ID
Lbl                                          Intf/Tunnel
  NextHop
Lbl (backup)
  NextHop (backup)
-----
192.0.2.4/32                                SR-ISIS-0    524299
32004
  192.168.14.2                               1/1/2:1000
32004/524285/32005
  192.168.13.2(B)                             1/1/3:1000
-----
Total Entries : 1
=====
```

Figure 134: Directed LFA with P router and Q router one hop apart shows the directed LFA path from source PE-1 to P router PE-5 (node SID), the adjacency SID from P router PE-5 to Q router PE-8, and the node SID of destination PE-4. P router PE-5 uses the adjacency SID for forwarding, but only sends the packets with the node SID of PE-4 (32004).

Figure 134: Directed LFA with P router and Q router one hop apart

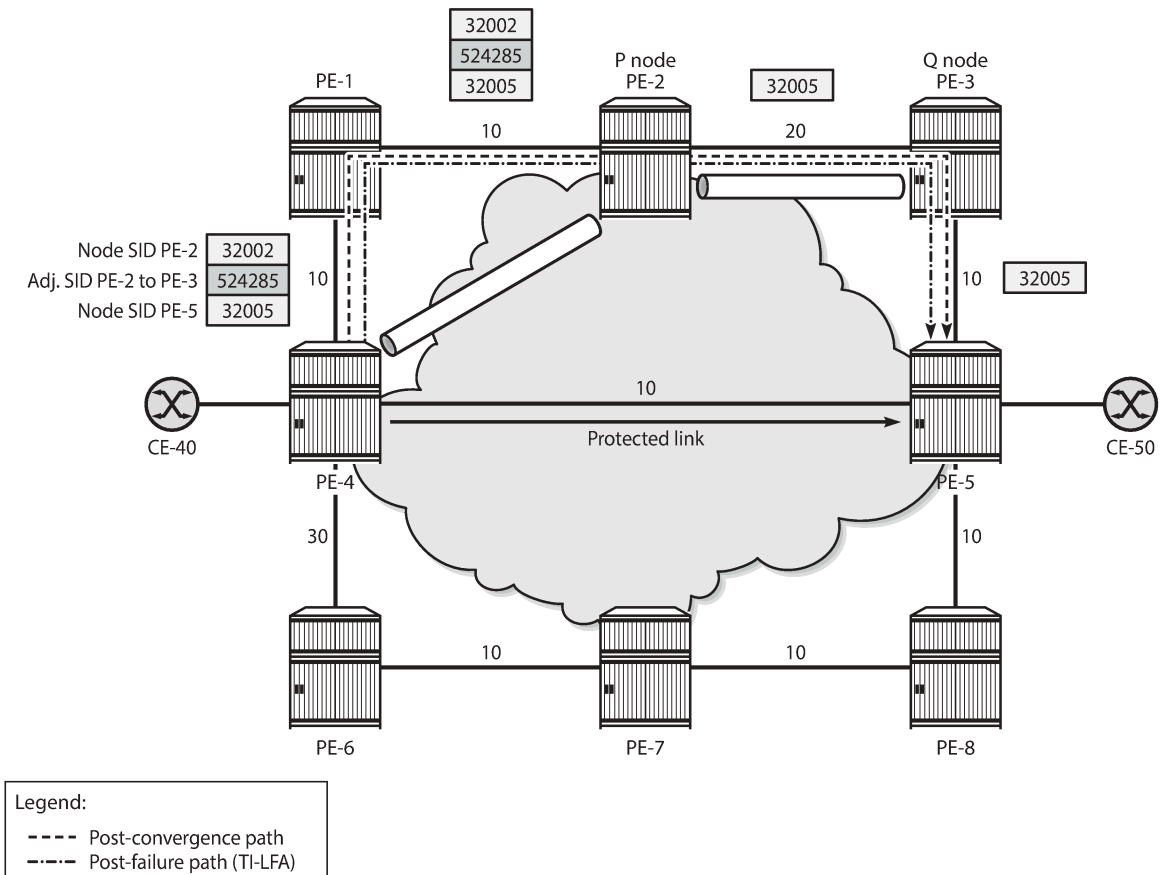


29360

## TI-LFA for coinciding post-failure and post-convergence paths

Figure 135: Post-failure TI-LFA path coincides with post-convergence path is the same as Figure 127: Post-failure TI-LFA path matches post-convergence path and is repeated here for readability. The router interfaces have IGP metric 10 by default, except for the interfaces between PE-2 and PE-3 that have metric 20, and the interfaces between PE-4 and PE-6 that have metric 30. As in Figure 134: Directed LFA with P router and Q router one hop apart, Figure 135: Post-failure TI-LFA path coincides with post-convergence path shows the different tunnels used for the TI-LFA path. TI-LFA ensures that the post-failure path coincides with the post-convergence path by adding additional labels: the node SID 32002 (or 32062 for IPv6) to P router PE-2, the adjacency SID on PE-2 for the interface toward Q router PE-3, and the node SID 32005 (or 32065 for IPv6) toward the destination PE-5.

Figure 135: Post-failure TI-LFA path coincides with post-convergence path



29353

## Regular LFA coverage

For a better comparison, the regular LFA coverage is calculated first. Without remote LFA and TI-LFA enabled, the LFA coverage is limited. The following command disables remote LFA and TI-LFA on all nodes, while regular LFA remains enabled:

```
# on all nodes:
configure {
  router "Base" {
    isis 0 {
      loopfree-alternate {
        delete remote-lfa
        delete ti-lfa
      }
    }
  }
}
```

The SR LFA coverage on PE-4 only protects node SIDs and adjacency SIDs that can be protected with regular LFA, as follows:

[/]



```
A:admin@PE-4# show router isis sr-lfa-coverage
```

```
=====
```

Rtr Base ISIS Instance 0 SR LFA Coverage							
MT-ID	SidType	Level	Proto	LFA	RLFA	TILFA	Coverage
0	node-sid	L1	ipv4	5(71%)	0(0%)	0(0%)	5/7(71%)
0	node-sid	L1	ipv6	5(71%)	0(0%)	0(0%)	5/7(71%)
---snip---							
0	adj-sid	L1L2	ipv4	2(66%)	0(0%)	0(0%)	2/3(66%)
0	adj-sid	L1L2	ipv6	2(66%)	0(0%)	0(0%)	2/3(66%)

```
=====
```

The following shows that no LFA paths exist on PE-4 for destinations 192.0.2.1 (PE-1), 192.0.2.2 (PE-2), and 192.168.14.1 (PE-1).

```
[/]
A:admin@PE-4# show router fp-tunnel-table 1
```

```
=====
```

IPv4 Tunnel Table Display

Legend:  
label stack is ordered from bottom-most to top-most  
B - FRR Backup

```
=====
```

Destination	Protocol	Tunnel-ID
Lbl		Intf/Tunnel
NextHop		
Lbl (backup)		
NextHop (backup)		
<b>192.0.2.1/32</b>	<b>SR-ISIS-0</b>	<b>524291</b>
32001		
192.168.14.1		1/1/1:1000
<b>192.0.2.2/32</b>	<b>SR-ISIS-0</b>	<b>524319</b>
32002		
192.168.14.1		1/1/1:1000
192.0.2.3/32	SR-ISIS-0	524293
32003		
192.168.45.2		1/1/3:1000
32003		
192.168.46.2(B)		1/1/2:1000
<b>192.0.2.5/32</b>	<b>SR-ISIS-0</b>	<b>524299</b>
32005		
192.168.45.2		1/1/3:1000
32005		
192.168.46.2(B)		1/1/2:1000
192.0.2.6/32	SR-ISIS-0	524311
32006		
192.168.46.2		1/1/2:1000
32006		
192.168.45.2(B)		1/1/3:1000
192.0.2.7/32	SR-ISIS-0	524323
32007		
192.168.45.2		1/1/3:1000
32007		
192.168.46.2(B)		1/1/2:1000
192.0.2.8/32	SR-ISIS-0	524313
32008		
192.168.45.2		1/1/3:1000
32008		

192.168.46.2(B)		1/1/2:1000
<b>192.168.14.1/32</b>	<b>SR</b>	<b>524317</b>
3		
192.168.14.1		1/1/1:1000
192.168.45.2/32	SR	524321
3		
192.168.45.2		1/1/3:1000
32005		
192.168.46.2(B)		1/1/2:1000
192.168.46.2/32	SR	524309
3		
192.168.46.2		1/1/2:1000
32006		
192.168.45.2(B)		1/1/3:1000
-----		
Total Entries : 10		
-----		
=====		

For destination 192.0.2.5, the post-failure path has next-hop 192.168.46.2 on PE-6, so the post-failure path does not coincide with the post-convergence path with next-hop 192.168.14.1 on PE-1. The path cost of the post-convergence path from PE-4 to PE-5 (via PE-1, PE-2, and PE-3) equals  $10 + 10 + 20 + 10 = 50$ ; the path cost of the post-failure path from PE-4 to PE-5 (via PE-6, PE-7, and PE-8) equals  $30 + 10 + 10 + 10 = 60$ .

## TI-LFA enabled

TI-LFA can be configured with remote LFA enabled or disabled. The following command configures remote LFA and TI-LFA (with default max-sr-frr-labels 2).

```
# on all nodes:
configure {
  router "Base" {
    isis 0 {
      loopfree-alternate {
        remote-lfa {
        }
      }
      ti-lfa {
        max-sr-frr-labels 2
      }
    }
  }
}
```

With TI-LFA enabled, the SR LFA coverage increases to 100%, as follows. For almost all destinations, the LFA protection is now using TI-LFA, even when regular LFA was possible before. The advantage is that TI-LFA ensures the post-failure path coincides with the post-convergence path.

If there is regular LFA protection via a path that does not coincide with the post-convergence path, that regular LFA protection will only change to TI-LFA protection when max-sr-frr-labels allows the needed number of labels (tunnels) to force the TI-LFA protection to the post-convergence path. The same applies for remote LFA protection.

```
[/]
A:admin@PE-4# show router isis sr-lfa-coverage

=====
Rtr Base ISIS Instance 0 SR LFA Coverage
=====
MT-ID  SidType      Level Proto LFA      RLFA     TILFA     Coverage
-----
```

```

0      node-sid    L1    ipv4  0(0%)   0(0%)   7(100%)  7/7(100%)
0      node-sid    L1    ipv6  0(0%)   0(0%)   7(100%)  7/7(100%)
---snip---
0      adj-sid    L1L2  ipv4  0(0%)   0(0%)   3(100%)  3/3(100%)
0      adj-sid    L1L2  ipv6  0(0%)   0(0%)   3(100%)  3/3(100%)
=====

```

The following FP tunnel table shows that prefixes 192.0.2.1 (PE-1), 192.0.2.2 (PE-2), and 192.168.14.1 (PE-1) are now protected too. For destination 192.0.2.5 (PE-5), the next-hop now is 192.168.14.1, which is also the next-hop on the post-convergence path to PE-5 via PE-1, PE-2, and PE-3. The top label 32002 is the node SID of PE-2, the label 524285 is the adjacency SID on PE-2 for the interface toward PE-3, and the bottom label 32005 is the node SID to reach the destination PE-5.

```

[/]
A:admin@PE-4# show router fp-tunnel-table 1

=====
IPv4 Tunnel Table Display

Legend:
Label stack is ordered from bottom-most to top-most
B - FRR Backup
=====
Destination                                Protocol          Tunnel-ID
Lbl
NextHop                                     Intf/Tunnel
Lbl      (backup)
NextHop  (backup)
-----
192.0.2.1/32                                SR-ISIS-0        524291
32001
  192.168.14.1                               1/1/1:1000
  32001/524285/32003
  192.168.45.2(B)                             1/1/3:1000
192.0.2.2/32                                SR-ISIS-0        524319
32002
  192.168.14.1                               1/1/1:1000
  32002/524285/32003
  192.168.45.2(B)                             1/1/3:1000
192.0.2.3/32                                SR-ISIS-0        524293
32003
  192.168.45.2                               1/1/3:1000
  32003/524285/32002
  192.168.14.1(B)                             1/1/1:1000
192.0.2.5/32                                SR-ISIS-0        524299
32005
  192.168.45.2                               1/1/3:1000
  32005/524285/32002
  192.168.14.1(B)                             1/1/1:1000
192.0.2.6/32                                SR-ISIS-0        524311
32006
  192.168.46.2                               1/1/2:1000
  32006
  192.168.45.2(B)                             1/1/3:1000
192.0.2.7/32                                SR-ISIS-0        524323
32007
  192.168.45.2                               1/1/3:1000
  32007
  192.168.46.2(B)                             1/1/2:1000
192.0.2.8/32                                SR-ISIS-0        524313
32008
  192.168.45.2                               1/1/3:1000

```

```

32008
  192.168.46.2(B)
192.168.14.1/32          SR          1/1/2:1000
  3                          524317
    192.168.14.1          1/1/1:1000
32001/524285/32003
  192.168.45.2(B)
192.168.45.2/32          SR          1/1/3:1000
  3                          524321
    192.168.45.2          1/1/3:1000
  32005/524285/32002
    192.168.14.1(B)      1/1/1:1000
192.168.46.2/32          SR          524309
  3
    192.168.46.2          1/1/2:1000
  32006
    192.168.45.2(B)      1/1/3:1000
-----
Total Entries : 10
-----
=====

```

The following **tools** command on PE-4 includes detailed information for the LFA protection for destination 192.0.2.5:

```

[/]
A:admin@PE-4# tools dump router isis sr-database prefix 192.0.2.5 detail
=====
Rtr Base ISIS Instance 0 SR Database

Legend:
label stack is ordered from bottom-most to top-most
=====
SID 5
-----
Label           : 32005          Adv System Id   : 1920.0000.2005
Prefix          : 192.0.2.5
Route Level     : 1            MT Id           : 0
Rtm Preference  : 15          Ttm Preference  : 11
Metric          : 10          Last Action     : LfaNhops
Num Ip NextHop  : 1            Num SR-Tnl NextHop : 1
Mtu             : 8970
Mtu Prim        : 8982          Mtu Backup      : 8982
Exclude from LFA : 0          LFA Type       : TI LFA
Duplicate Pending : 0          Tunnel Active State : Reported/Ack
SR Error        : SR_ERR_OK

LFA NextHop IP  : 192.168.14.1
LFA IsTunl     : N
LFA GIfId/TunlType : 1          LFA IfId/LspId   : 2
LFA PgId       : 0            LFA Adv Node      : False
LFA Labels    : 32005/524285/32002

NHOP: IP           IsTunl GIfId/  IfId/ PgId  IsAdv Label  IsLfaX
                TunlType LspId
-----
192.168.45.2      N      2      3      13      1      32005      0
-----

No. of Entries: 1
-----
LDP = LDP FEC is the SID NH for SR-LDP stitching

```

## TI-LFA enabled with max-sr-frr-labels lower than 2

When TI-LFA is configured with max-sr-frr-labels lower than 2, TI-LFA cannot substitute regular or remote LFA where more than 2 tunnel labels are needed for the substitution. Some destinations may remain protected then via regular or remote LFA, and only those destinations that can be protected with TI-LFA with less than 2 tunnel labels will have TI-LFA protection. The following configuration enables TI-LFA with max-sr-frr-labels equal to 1:

```
# on all nodes:
configure {
  router "Base" {
    isis 0 {
      loopfree-alternate {
        remote-lfa {
        }
        ti-lfa {
          max-sr-frr-labels 1
        }
      }
    }
  }
}
```

In the topology of [Figure 135: Post-failure TI-LFA path coincides with post-convergence path](#), for max-sr-frr-labels equal to 1, the SR LFA coverage drops below 100% again, as follows.

```
[/]
A:admin@PE-4# show router isis sr-lfa-coverage

=====
Rtr Base ISIS Instance 0 SR LFA Coverage
=====
MT-ID  SidType      Level Proto LFA      RLFA      TILFA      Coverage
-----
0      node-sid      L1   ipv4  2(28%)  0(0%)     3(42%)     5/7(71%)
0      node-sid      L1   ipv6  2(28%)  0(0%)     3(42%)     5/7(71%)
---snip---
0      adj-sid       L1L2 ipv4   1(33%)  0(0%)     1(33%)     2/3(66%)
0      adj-sid       L1L2 ipv6   1(33%)  0(0%)     1(33%)     2/3(66%)
=====
```

The preceding information can be derived from the FP tunnel table and the SR database as follows. For PE-4, the FP tunnel table shows that there are 10 destinations, 7 nodes and 3 next-hops. 5 out of 7 node destinations and 2 out of 3 next-hop destinations are protected with a backup (B). Node destination 192.0.2.1 (PE-1), and 192.0.2.2 (PE-2), and next-hop destination 192.168.14.1 are no longer protected.

```
[/]
A:admin@PE-4# show router fp-tunnel-table 1

=====
IPv4 Tunnel Table Display

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

NextHop (backup)		
192.0.2.1/32	SR-ISIS-0	524291
32001		
192.168.14.1		1/1/1:1000
192.0.2.2/32	SR-ISIS-0	524319
32002		
192.168.14.1		1/1/1:1000
192.0.2.3/32	SR-ISIS-0	524293
32003		
192.168.45.2		1/1/3:1000
32003		
192.168.46.2(B)		1/1/2:1000
192.0.2.5/32	SR-ISIS-0	524299
32005		
192.168.45.2		1/1/3:1000
32005		
192.168.46.2(B)		1/1/2:1000
192.0.2.6/32	SR-ISIS-0	524311
32006		
192.168.46.2		1/1/2:1000
32006		
192.168.45.2(B)		1/1/3:1000
192.0.2.7/32	SR-ISIS-0	524323
32007		
192.168.45.2		1/1/3:1000
32007		
192.168.46.2(B)		1/1/2:1000
192.0.2.8/32	SR-ISIS-0	524313
32008		
192.168.45.2		1/1/3:1000
32008		
192.168.46.2(B)		1/1/2:1000
192.168.14.1/32	SR	524317
3		
192.168.14.1		1/1/1:1000
192.168.45.2/32	SR	524325
3		
192.168.45.2		1/1/3:1000
32005		
192.168.46.2(B)		1/1/2:1000
192.168.46.2/32	SR	524309
3		
192.168.46.2		1/1/2:1000
32006		
192.168.45.2(B)		1/1/3:1000
-----		
Total Entries : 10		
-----		
=====		

The SR database indicates what type of protection corresponds with the (topmost) label of the destinations in the FP tunnel label. Destination 192.0.2.1 (PE-1), 192.0.2.2 (PE-2), and 192.168.14.1 have no backup. Their label indicates that there is no LFA protection (LT = -). Destination 192.0.2.3 (PE-3), 192.0.2.5 (PE-5), and 192.168.45.2 have a backup with a (topmost) label that indicates regular LFA protection (LT = L). So, destination 192.0.2.5 (PE-5) is no longer TI-LFA protected, because that would require 2 tunnel labels, which max-sr-frr-labels=1 prevents. Destination 192.0.2.6 (PE-6), 192.0.2.7 (PE-7), 192.0.2.8 (PE-8), and 192.168.46.2 have a backup with a (topmost) label that indicates TI-LFA protection (LT = T). As these destinations have no TI-LFA tunnel label, their TI-LFA protection does not need tunnels to ensure that the TI-LFA protection is via the post-convergence path.

The following **tools** command on PE-4 includes detailed information for the type of LFA protection that corresponds with a label:

```
[/]
A:admin@PE-4# tools dump router isis sr-database ipv4-unicast
=====
Rtr Base ISIS Instance 0 SR Database
=====
SID  Label  Prefix          Last-act  Lev MT RtmPref TtmPref Metric  IpNh SrNh
Mtu   MtuPrim MtuBk   D xL LT Act AdvSystemId  SrErr
-----
1    32001  192.0.2.1      RemLfaNh 1  0 15    11    10    1    1
8974  8982    -        0 0 - +R 1920.0000.2001 SR_ERR_OK
2    32002  192.0.2.2      RemLfaNh 1  0 15    11    20    1    1
8974  8982    -        0 0 - +R 1920.0000.2002 SR_ERR_OK
3    32003  192.0.2.3      LfaNhops 1  0 15    11    20    1    1
8974  8982    8982    0 0 L +R 1920.0000.2003 SR_ERR_OK
4    32004  192.0.2.4      Local     -  -  -    -    -    -    -
-      -      -        0 - - +R 1920.0000.2004 SR_ERR_OK
5    32005  192.0.2.5      LfaNhops 1  0 15    11    10    1    1
8974  8982    8982    0 0 L +R 1920.0000.2005 SR_ERR_OK
6    32006  192.0.2.6      TnlChange 1  0 15    11    30    1    1
8974  8982    8982    0 0 T +R 1920.0000.2006 SR_ERR_OK
7    32007  192.0.2.7      TnlChange 1  0 15    11    30    1    1
8974  8982    8982    0 0 T +R 1920.0000.2007 SR_ERR_OK
8    32008  192.0.2.8      TnlChange 1  0 15    11    20    1    1
8974  8982    8982    0 0 T +R 1920.0000.2008 SR_ERR_OK
-----
No. of Entries: 8
-----
Lev = route level
IpNh = number of IP next-hops
SrNh = number of SR-tunnel next-hops
D = duplicate pending
xL = exclude from LFA
LT = LFA type (L:LFA, R:RLFA, T:TILFA, n:nodeProtection)
Act = tunnel active state (R:reported, F:failed, +:SR-ack)
=====
```

Independent from the preceding ISIS Segment Routing LFA coverage (per Segment Routing LFA type and per ISIS Level), there is also the ISIS IP-routing LFA coverage (per IP version and per ISIS Level), as follows:

```
[/]
A:admin@PE-4# show router isis lfa-coverage
=====
Rtr Base ISIS Instance 0 LFA Coverage
=====
Topology          Level  Node          IPv4          IPv6
-----
IPV4 Unicast      L1     5/7(71%)      9/13(69%)     9/13(69%)
IPV6 Unicast      L1     0/0(0%)       0/0(0%)       0/0(0%)
IPV4 Multicast    L1     0/0(0%)       0/0(0%)       0/0(0%)
```

IPv6 Multicast	L1	0/0(0%)	0/0(0%)	0/0(0%)
<b>IPv4 Unicast</b>	<b>L2</b>	<b>5/7(71%)</b>	<b>9/13(69%)</b>	<b>9/13(69%)</b>
IPv6 Unicast	L2	0/0(0%)	0/0(0%)	0/0(0%)
IPv4 Multicast	L2	0/0(0%)	0/0(0%)	0/0(0%)
IPv6 Multicast	L2	0/0(0%)	0/0(0%)	0/0(0%)

The preceding information can be derived from the table of alternative ISIS routes as follows. For PE-4, there are 17 routes: 8 routes to nodes and 9 routes to networks. The node and the networks that have 0.0.0.0 as next-hop must not be considered. This leaves  $(8 - 1) = 7$  routes to nodes and  $(9 - 3) = 6$  routes to networks. 5 out of 7 node destinations, and 4 out of 6 network destinations have an LFA next-hop (L). This leads to  $(5 + 4) / (7 + 6) = 9/13$  IPv4 prefixes that have ISIS IP routing LFA coverage. A similar derivation applies for IPv6 prefixes.

```
[/]
A:admin@PE-4# show router isis routes alternative

=====
Rtr Base ISIS Instance 0 Route Table (alternative)
=====
Prefix[Flags]           Metric    Lvl/Typ    Ver.  SysID/Hostname
NextHop                MT        AdminTag/SID[F]
Alt-Nexthop            Alt-      Alt-Type
Metric
-----
192.0.2.1/32            10        1/Int.     57    PE-1
  192.168.14.1          0          0/1[NnP]
192.0.2.2/32            20        1/Int.     62    PE-1
  192.168.14.1          0          0/2[NnP]
192.0.2.3/32            20        1/Int.     82    PE-5
  192.168.45.2          0          0/3[NnP]
  192.168.46.2(L)      70        LP
192.0.2.4/32            0         1/Int.     3     PE-4
  0.0.0.0               0          0/4[NnP]
192.0.2.5/32            10        1/Int.     82    PE-5
  192.168.45.2          0          0/5[NnP]
  192.168.46.2(L)      60        LP
192.0.2.6/32            30        1/Int.     70    PE-6
  192.168.46.2          0          0/6[NnP]
  192.168.45.2(L)      40        LP
192.0.2.7/32            30        1/Int.     82    PE-5
  192.168.45.2          0          0/7[NnP]
  192.168.46.2(L)      40        NP
192.0.2.8/32            20        1/Int.     82    PE-5
  192.168.45.2          0          0/8[NnP]
  192.168.46.2(L)      50        NP
192.168.12.0/30         20        1/Int.     57    PE-1
  192.168.14.1          0          0
192.168.14.0/30         10        1/Int.     6     PE-4
  0.0.0.0               0          0
192.168.23.0/30         40        1/Int.     82    PE-1
  192.168.14.1          0          0
192.168.35.0/30         20        1/Int.     82    PE-5
  192.168.45.2          0          0
  192.168.46.2(L)      70        LP
192.168.45.0/30         10        1/Int.     82    PE-4
  0.0.0.0               0          0
192.168.46.0/30         30        1/Int.     66    PE-4
  0.0.0.0               0          0
192.168.58.0/30         20        1/Int.     82    PE-5
  192.168.45.2          0          0
  192.168.46.2(L)      70        LP
```



```

192.168.67.0/30          40      1/Int.    82    PE-5
  192.168.45.2          0
  192.168.46.2(L)      50      NP
192.168.78.0/30        30      1/Int.    82    PE-5
  192.168.45.2          0
  192.168.46.2(L)      60      NP
-----
No. of Routes: 17 (17 paths)
-----
Flags          : L = Loop-Free Alternate nexthop
Alt-Type       : LP = linkProtection, NP = nodeProtection
SID[F]        : R = Re-advertisement
                N = Node-SID
                nP = no penultimate hop POP
                E = Explicit-Null
                V = Prefix-SID carries a value
                L = value/index has local significance
=====
    
```

## Conclusion

TI-LFA extends the calculation of a backup path for cases where the extended P space and the Q space do not overlap. TI-LFA also ensures that the post-failure path coincides with the post-convergence path, which avoids a switchover after SPF convergence.

# Tunneling of ICMP Reply Packets over MPLS LSPs

This chapter provides information about tunneling of ICMP reply packets over MPLS LSPs.

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

## Applicability

This chapter is applicable to SR OS routers and was initially written for SR OS Release 13.0.R7. The CLI in the current edition corresponds to SR OS Release 23.3.R1. Internet Control Message Protocol (ICMP) tunneling over Multi-Protocol Label Switching (MPLS) Label Switched Paths (LSPs) is supported in SR OS Release 12.0.R4 or later.

## Overview

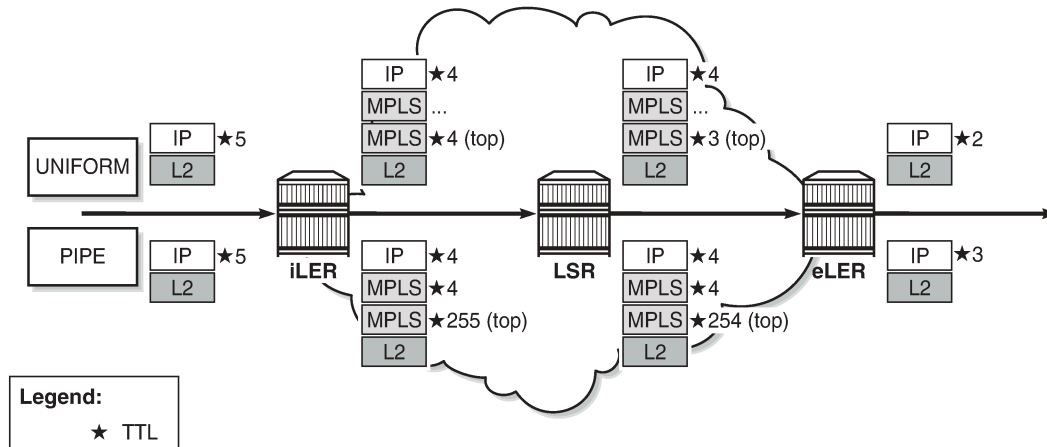
In IP forwarding, Time-To-Live (TTL) is a well-known mechanism to mitigate the damage in case of a loop. The TTL value in the IP header is decremented by one at each hop and the packet is discarded when the TTL equals 0. TTL is also used in traceroute, where the first batch of echo requests are sent with TTL equal to 1, the second batch of echo requests is sent with TTL equal to 2, and so on. Any intermediate node where the TTL expires (is decremented to 0) sends an ICMP reply of type "Time exceeded" (type 11) to the sender. From the replies, the sequence of hops can be determined.

If ICMP messages are sent in an MPLS tunnel, in pipe mode, the hops in the tunnels are invisible and the TTL is only decremented by the Label Edge Routers (LERs), not by the intermediate Label Switching Routers (LSRs). However, there are two modes for TTL handling, according to RFC 3443, *Time To Live Processing in MPLS Networks*:

1. Uniform mode: the MPLS network is visible from the outside. MPLS nodes use the TTL in the same way as any other IP node.
2. Pipe mode: the MPLS network is invisible from the outside. MPLS use of TTL is independent from IP TTL use. The network appears like a pipe between ingress Label Edge Router (iLER) and egress Label Edge Router (eLER).

Both TTL uses are shown in [Figure 136: Use of TTL: uniform versus pipe](#):

Figure 136: Use of TTL: uniform versus pipe



25696

Independent of the mode, the iLER decrements the TTL in the IP header by one. The iLER adds service and transport MPLS headers.



**Note:**

In an L2 Virtual Private Network (VPN), the TTL in the IP header is kept intact.

- In uniform mode, the iLER sets the TTL of every MPLS header to match the TTL in the IP header and every LSR decrements the MPLS TTLs. The IP header remains unchanged as long as the packet is in the MPLS tunnel. The eLER pops the MPLS labels and decrements the minimum TTL of the headers (which is the TTL in both MPLS headers) by one. This TTL is used in the IP header.
- In pipe mode, the iLER sets the TTL of the top MPLS header to 255 and every LSR decrements that TTL by one. The eLER pops the MPLS labels and decrements the minimum TTL of the headers (which is the IP TTL) by one. This TTL is used in the IP header. There can be uncounted hops in pipe mode, because the LSRs are not counted.

The LERs can be in uniform mode and the LSRs in pipe mode, and the other way around.

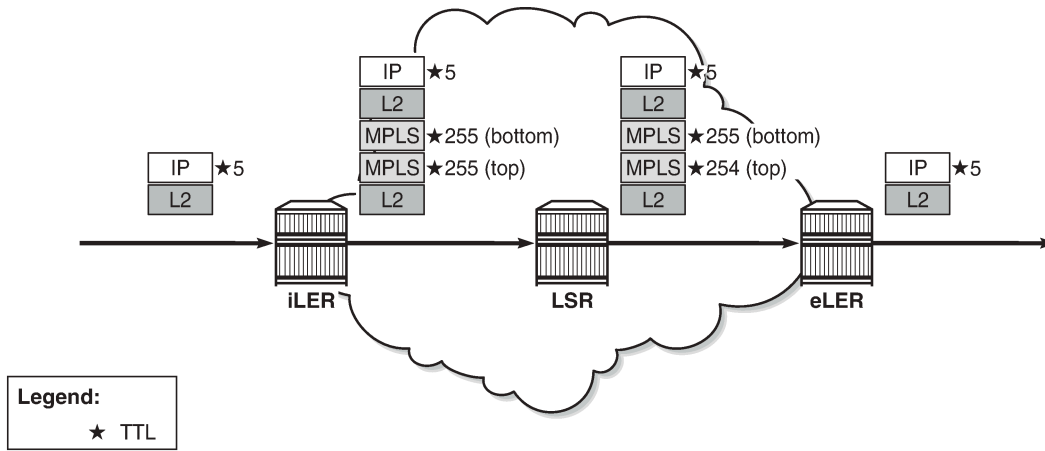
The default use of TTL in SR OS is as follows:

- Uniform mode for LSP shortcuts, Resource Reservation Protocol (RSVP) shortcuts, Label Distribution Protocol (LDP) shortcuts, and Border Gateway Protocol (BGP) shortcuts.
- Pipe mode for L2 and L3 VPN services, BGP labeled routes, IPv6 Provider Edge (6PE) router, and IPv6 on VPN to PE router (6VPE).

However, the use of TTL can be changed by configuration.

Figure 137: Use of TTL in an L2 VPN service in pipe mode shows the use of TTL for an L2 VPN service in pipe mode. The TTL in the IP header is preserved. There is no processing of the IP header for an L2 service. The TTL in the pushed MPLS headers is 255 and the TTL in the top MPLS header is decremented by one in the LSRs. The eLER pops the MPLS labels.

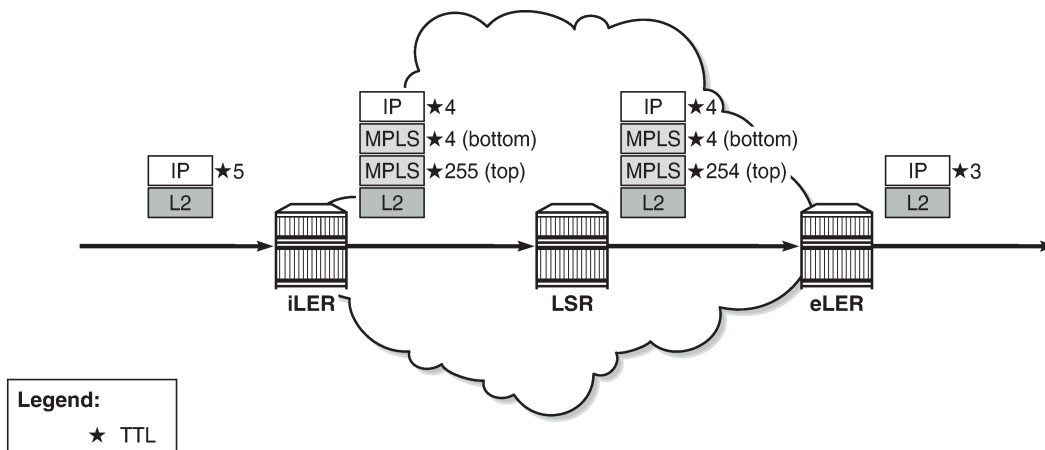
Figure 137: Use of TTL in an L2 VPN service in pipe mode



25697

Figure 138: Use of TTL in an L3 VPN service in pipe mode shows the use of TTL for an L3 VPN service in pipe mode. The TTL in the IP header is decremented by the iLER and the eLER, but not by the LSRs. In pipe mode, the bottom MPLS header inherits the IP TTL after it has been decremented by the iLER. The transport MPLS header gets TTL 255 and this TTL is decremented by one at each LSR. The eLER takes the minimum of the TTL of the MPLS headers and the IP TTL and decrements that by one. This will match the IP TTL in the forwarded packet. The MPLS labels are popped. There are uncounted hops, because the LSRs are invisible in pipe mode.

Figure 138: Use of TTL in an L3 VPN service in pipe mode



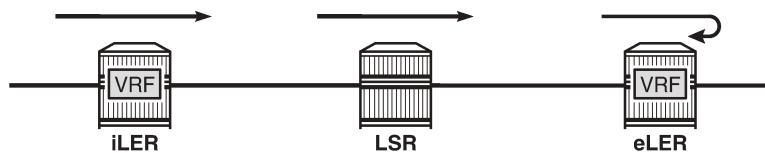
25698

Tunneling of ICMP reply packets over MPLS LSPs provides the ability for a network operator or customer to trace the MPLS network hops in the path, for Virtual Private Routed Network (VPRN), 6PE/6VPE, and BGP labeled routes.

## ICMP tunneling over an MPLS LSP

Figure 139: Tunneling of ICMP reply packets over an MPLS LSP shows the actions performed in iLER, LSR, and eLER, when tunneling ICMP messages over an MPLS LSP:

Figure 139: Tunneling of ICMP reply packets over an MPLS LSP



25699

- In the iLER, uniform mode is required within the VPRN service. The IP TTL is propagated in the MPLS TTL, both for in-transit and Control Processing Module (CPM) generated IP packets. In this example, it is assumed that a UDP traceroute message is forwarded with source IP address S1.
- In all LSRs, ICMP tunneling is enabled globally on the system, according to RFC 3032, *MPLS Label Stack Encoding*. When the MPLS TTL expires in an LSR, the LSR generates an ICMP reply with code "Time exceeded" and destination IP address S1. However, the CPM sends this ICMP reply packet in the forward direction of the MPLS LSP tunnel that the packet arrived on. The ICMP reply packet is sent to the eLER, not to the iLER.
- The eLER performs a lookup for the IP address S1 and sends the ICMP reply to S1 toward the iLER.
  - The lookup of the IP address S1 is in the Global Routing Table (GRT) for BGP shortcut, 6PE, and BGP labeled route prefixes.
  - The lookup of IP address S1 is in the Virtual Routing and Forwarding (VRF) table for VPRN and 6VPE prefixes.

## TTL propagation

The TTL propagation can be configured in LERs and LSRs.

### TTL propagation at the iLER

Different commands are used for TTL propagation in a VPRN versus BGP labeled routes. Pipe mode is enabled by default in either case.

### TTL propagation in iLER for VPRN

The TTL propagation of VPN-IPv4 or VPN-IPv6 packets in a VPRN service can be enabled globally as follows:

```
A:admin@PE-3# configure exclusive
  router "Base" {
    ttl-propagate {
      vprn-local ?
    }
  }
[ex:/configure router "Base" ttl-propagate]
```

```
A:admin@PE-3#          vprn-local ?

vprn-local <keyword>
<keyword> - (none|all|vc-only)
Default   - vc-only

    TTL propagation from IP header into label stack for VPRN local packets

A:admin@PE-3# configure exclusive
    router "Base" {
        ttl-propagate {
            vprn-transit ?
[ex:/configure router "Base" ttl-propagate]
A:admin@PE-3#          vprn-transit ?

vprn-transit <keyword>
<keyword> - (none|all|vc-only)
Default   - vc-only

    TTL propagation from IP header into label stack for VPRN transit packets
```

There are three options for the propagation of TTL in the iLER of a VPRN:

- none                      No IP TTL propagation to any MPLS header in the stack: the transport and the VC MPLS headers will have a value of 255. This is needed for correct operation of traceroute in an inter-AS option B VPRN.
- all                        Uniform mode: IP TTL is propagated to all MPLS headers in the stack.
- vc-only (default)        Pipe mode: IP TTL is propagated to the VC header, but not to the transport headers in the stack.

For more information about inter-AS option B, see chapter *Rosen MVPN Inter-AS Option B*.

In inter-AS option B, a traceroute for a VPN IP prefix issued from a Customer Edge (CE) router results in both ingress Autonomous System Boundary Router (ASBR) and egress ASBR not responding. The traceroute also misses a couple of hops if the target CE node is two or more hops away from the egress PE. The reason is that the VC label TTL inherits the decremented IP TTL at the ingress PE, but is decremented twice in the MPLS network whereas the IP TTL is only decremented at the ingress and the egress PE nodes. The option "none" for **ttl-propagate** makes the ASBRs transparent to the traceroute behavior and corrects the uncounted hop issue.

The global configuration can be overruled within each VPRN, as follows

```
A:admin@PE-3# configure exclusive
    service {
        vprn "1" {
            ttl-propagate {
                local ?
[ex:/configure service vprn "1" ttl-propagate]
A:admin@PE-3#          local ?

local <keyword>
<keyword> - (none|all|vc-only|use-base)
Default   - use-base

    Local TTL propagation control for the VPRN
```

```
A:admin@PE-3# configure exclusive
  service {
    vprn "1" {
      ttl-propagate {
        transit ?
      }
    }
  }
[ex:/configure service vprn "1" ttl-propagate]
A:admin@PE-3# transit ?

transit <keyword>
<keyword> - (none|all|vc-only|use-base)
Default   - use-base

Transit TTL propagation control for the VPRN
```

## TTL propagation in iLER for BGP labeled route

IPv4 and IPv6 packets are forwarded using BGP labeled routes in the GRT, as described in RFC 3107, *Carrying Label Information in BGP-4*. This also applies to 6PE. TTL propagation for RFC 3107 label routes can be configured as follows:

```
A:admin@PE-3# configure exclusive
  router "Base" {
    ttl-propagate {
      label-route-local ?
    }
  }
[ex:/configure router "Base" ttl-propagate]
A:admin@PE-3# label-route-local ?

label-route-local <keyword>
<keyword> - (none|all)
Default   - none

TTL propagation from IP header into label stack for local packets

A:admin@PE-3# configure exclusive
  router "Base" {
    ttl-propagate {
      label-route-transit ?
    }
  }
[ex:/configure router "Base" ttl-propagate]
A:admin@PE-3# label-route-transit ?

label-route-transit <keyword>
<keyword> - (none|all)
Default   - none

TTL propagation from IP header into label stack for transit packets
```

There are two options for TTL propagation in the iLER for BGP labeled routes:

- none (default) Pipe mode: No TTL is propagated from the IP header to the MPLS headers in the transport MPLS stack. However, the IP TTL is propagated to the bottom header: the virtual circuit (VC) header
- all Uniform mode: TTL is propagated to all headers in the transport MPLS stack.

If the BGP peer advertises the implicit-null label value for the BGP labeled route (in the case of a third-party implementation), the TTL propagation follows the configuration of the RSVP/LDP LSP shortcut that the BGP labeled route resolves to. This is not controlled by the preceding commands.

## TTL propagation at the LSR

In a VPRN service, there is no TTL propagation to be configured in the LSRs.

### TTL propagation in LSR for BGP labeled route

The IP TTL and VC TTL are not decremented by the LSRs. The TTL that is decremented is the minimum of the RSVP/LDP transport TTL and the BGP TTL.

1. The LSR determines the TTL using the following function:

**TTL = MIN {incoming transport label stack TTL, incoming swapped/stitched label TTL}**

2. The LSR decrements the TTL by one and writes it to the outgoing swapped/stitched BGP label.

This is always performed when an LSR is swapping or stitching a label at any stack depth.

The control plane indicates to the data plane whether a BGP labeled route is stitched or an LDP FEC is being stitched. The same node can perform stitching for one BGP labeled route and swapping for another one. See chapter [LDP FEC to BGP Label Route Stitching](#) for more information.

3. The LSR can propagate the decremented TTL to the outgoing transport label stack (if any) that is pushed on top of the BGP swapped/stitched label. This is configured as follows:

```
A:admin@PE-2# configure exclusive
  router "Base" {
    ttl-propagate lsr-label-route ?
  *[ex:/configure router "Base"]
A:admin@PE-2#          ttl-propagate lsr-label-route ?

  lsr-label-route <keyword>
  <keyword> - (none|all)
  Default   - none
```

TTL propagation from IP header into label stack

There are two options for TTL propagation in the LSR:

- |                   |  |
|-------------------|--|
| none<br>(default) | No TTL propagation of the decremented TTL to the MPLS transport label stack. |
| all               | TTL propagation of the decremented TTL to all LDP/RSVP transport labels.     |

It is safe to not propagate the TTL to the transport label stack for an ASBR/Area Border Router (ABR)/data path Route Reflector (RR)/BGP-LDP stitching node. Not propagating the TTL provides isolation of the network domains downstream of the LSRs. Operations, Administration, and Maintenance (OAM) packets, such as traceroute and ping, sent in the context of a BGP labeled route or VPRN will not expire in LSR nodes within these domains.

A node performing pseudowire (PW) switching terminates the transport label stack in pipe mode; the node ignores the TTL of the incoming transport label stack and propagates the TTL of the VC label. The TTL of the new pushed transport label stack is always 255.



## Some considerations on TTL propagation in LSR for BGP labeled routes

- When an LSR stitches an LDP label to a BGP label, the decremented TTL of the stitched label can be propagated to the LDP/RSVP transport labels with the preceding configuration.
- When an LSR stitches a BGP label to an LDP label, the decremented TTL of the stitched label is automatically propagated to the RSVP label if the outgoing LDP LSP is tunneled over RSVP.
- When the LSR pops a BGP label and forwards the packet using an IGP route (IGP route is preferred over BGP labeled route), the LSR pushes an LDP label on the packet and the TTL behavior is the same as when an LSR stitches a BGP label to an LDP label.
- In a Carrier Supporting Carrier (CSC) VPRN, the ingress CSC CE swaps an iBGP label for an eBGP label and the ingress CSC PE swaps the incoming eBGP label for a VPN-IPv4 label. The reverse operation is performed by the egress CSC PE and the egress CSC CE. In all cases, the decremented TTL of the swapped label is propagated to the LDP/RSVP transport labels.
- SR OS does not support ASBR or data path RR functionality for labeled IPv6 routes in the global routing instance (6PE).

## TTL propagation at the eLER

For packets received with a BGP labeled route and searched for in the GRT, the TTL of the forwarded IP packet is set to  $\text{MIN}\{\text{MPLS\_TTL}-1, \text{IP\_TTL}-1\}$ , where `MPLS_TTL` refers to the TTL in the outermost label in the popped stack. This is the same behavior as for LSP shortcuts.

For packets received in the context of VPRN, the TTL of the forwarded IP packet is set to  $\text{MIN}\{\text{MPLS\_TTL}-1, \text{VC\_TTL}-1, \text{IP\_TTL}-1\}$ , where `MPLS_TTL` refers to the TTL in the outermost label in the popped stack and `VC_TTL` refers to the TTL in the VC label in the popped stack.

## Some considerations on TTL propagation at the eLER

- When a packet is received in one VPRN instance and is redirected using policy-based routing to be forwarded in another VPRN instance, the TTL is governed by the configuration of the outgoing VPRN instance.
- When a packet is received in a `vprn` context but is searched for in the GRT (GRT leaking configured), the behavior of the TTL propagation is governed by:
  - the BGP labeled route configuration when the matching route is an RFC 3107 label route or a 6PE route
  - the LSP shortcut configuration when the matching route is an RSVP or LDP shortcut (default uniform mode)

For shortcuts, uniform mode is default. Pipe mode can be configured as follows:

```
# on eLER PE-6:
configure exclusive
  router "Base" {
    ldp {
      shortcut-local-ttl-propagate false
      shortcut-transit-ttl-propagate false
    }
    mpls {
      shortcut-local-ttl-propagate false
    }
  }
}
```

```

        shortcut-transit-ttl-propagate false
    }
    rsvp { }

```

## Enabling ICMP tunneling on LSRs

For all scenarios (VPRN and BGP labeled routes), ICMP tunneling needs to be enabled on all LSRs, as follows:

```

# on all LSRs:
configure exclusive
router "Base" {
    icmp-tunneling true
}

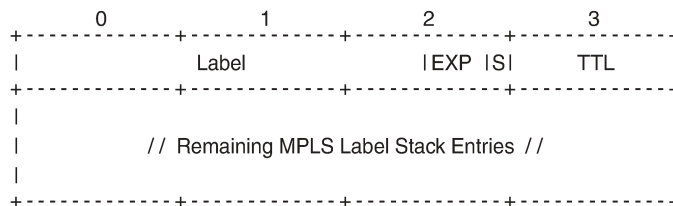
```

The LSR will generate the ICMP reply packet of type 11 - "Time exceeded", with source IP address set to a local address of the LSR node and appending the IP header and leading octets of the original datagram. The LSR does not perform a lookup for the destination IP address of the ICMP reply packet, which is the source IP address of the sender of the label TTL expiry packet. The CPM injects the ICMP reply packet in the forward direction toward the eLER. The TTL of pushed labels is 255.

There is no need to enable ICMP tunneling on the eLER. The eLER performs a user packet lookup in the data path in the VRF table or GRT and forwards the ICMP reply packet to the destination. If the eLER does not have a route to the destination, the packet is dropped.

RFC 4950, *ICMP Extensions for Multiprotocol Label Switching*, defines an extension object (MPLS label stack object) that allows LSRs to include label stack information to ICMP messages; see [Figure 140: MPLS label stack object](#):

Figure 140: MPLS label stack object

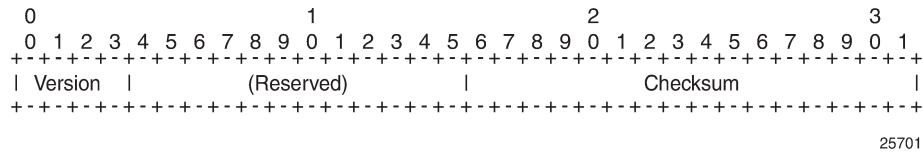


25700

The MPLS label stack object is applicable for ICMPv4 and ICMPv6. The MPLS label stack contains the MPLS shim header: label, experimental bits for Type of Service (ToS), S-bit indicating the bottom of the stack, and TTL. The object can be appended to the ICMP Time Exceeded and ICMP Destination Unreachable messages. The LSR that sends the ICMP reply message will not change the MPLS label stack.

RFC 4884, *Extended ICMP to Support Multi-Part Messages*, defines the ICMP extension header; see [Figure 141: ICMP extension header](#):

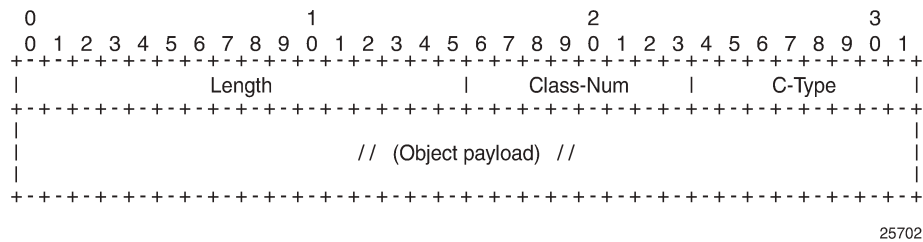
Figure 141: ICMP extension header



The version of the ICMP extension header is 2. The twelve reserved bits must be set to 0.

An extension object contains 32-bit words, representing an object header and payload, as defined in RFC 4884; see [Figure 142: ICMP extension object: object header and payload](#):

Figure 142: ICMP extension object: object header and payload



The length of the object is the length of the header (4 octets) plus the length of the object payload: 4 octets per LSR. The class number identifies the object class; in this case, object class 1 for MPLS label stack class. The C-type defines the object subtype; in this case, the subtype is 1 for an incoming MPLS label stack.

Backward compatibility is guaranteed between the ICMP message with extension header, the ICMP messages without extension header, and the ICMP message with a non-compliant extension header.

## Effect of ICMP tunneling on OAM

ICMP tunneling over MPLS LSPs affects the behavior of some CPM originated OAM packets that are forwarded within a **vprn** context.

- ICMP ping and UDP traceroute are sent according to the TTL propagation configured in the **vprn** context.
- VPRN ping and VPRN traceroute are not affected.

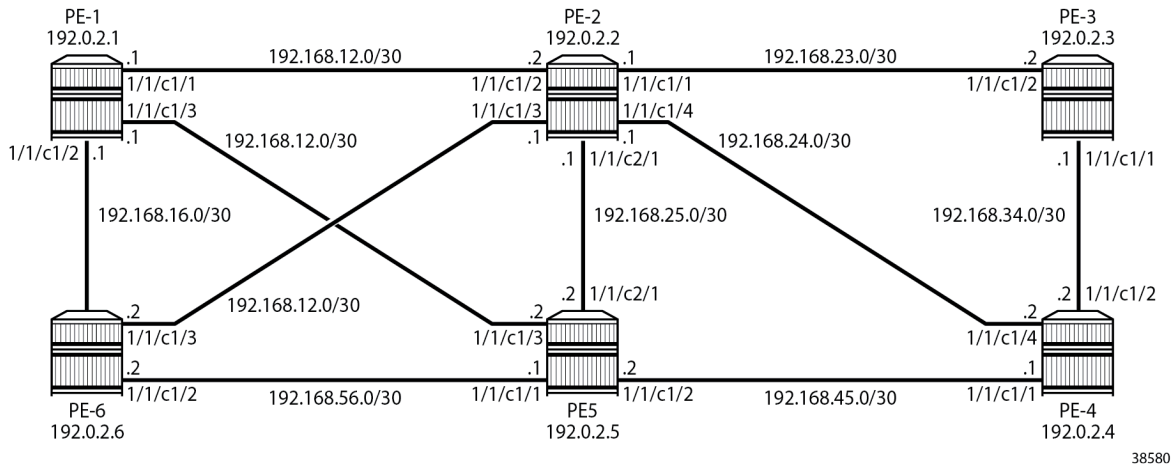
OAM packets forwarded over a BGP labeled route follow the TTL configuration of the iLER.

ICMP tunneling behavior at an LSR only applies to UDP traceroute packets. Other OAM packets expiring at the LSR, such as ICMP ping, VPRN ping, VPRN trace, LSP ping, and LSP trace, follow their specific procedures or are silently dropped.

## Configuration

[Figure 143: Example configuration](#) shows the example configuration, which has six 7750 SRs:

Figure 143: Example configuration



38580

## Initial configuration

The nodes have the following initial configuration:

- Cards, MDAs, ports
- Router interfaces.

```
# on PE-3:
configure exclusive
  router "Base" {
    interface "int-PE-3-PE-2" {
      port 1/1/c1/2
      ipv4 {
        primary {
          address 192.168.23.2
          prefix-length 30
        }
      }
    }
    interface "int-PE-3-PE-4" {
      port 1/1/c1/1
      ipv4 {
        primary {
          address 192.168.34.1
          prefix-length 30
        }
      }
    }
  }
  interface "system" {
    ipv4 {
      primary {
        address 192.0.2.3
        prefix-length 32
      }
    }
  }
}
```

- IGP: OSPF (alternatively, any IGP could have been used).

```
# on PE-3:
configure exclusive
router "Base" {
  ospf 0 {
    admin-state enable
    area 0.0.0.0 {
      interface "int-PE-3-PE-2" {
        interface-type point-to-point
      }
      interface "int-PE-3-PE-4" {
        interface-type point-to-point
      }
      interface "system" {
      }
    }
  }
}
```

- Link LDP.

```
# on PE-3:
configure exclusive
router "Base" {
  ldp {
    interface-parameters {
      interface "int-PE-3-PE-2" {
        ipv4 {
        }
      }
      interface "int-PE-3-PE-4" {
        ipv4 {
        }
      }
    }
  }
}
```

## Configure VPRN

A VPRN service is configured on PE-3 and PE-6. The routes are exchanged via BGP. BGP is configured on all nodes with PE-2 as route reflector (RR).

```
# on PE-1, PE-3, PE-4, PE-5, and PE-6:
configure exclusive
router "Base" {
  autonomous-system 64496
  bgp {
    ebgp-default-reject-policy {
      import false
      export false
    }
    group "internal" {
      peer-as 64496
      family {
        vpn-ipv4 true
      }
    }
  }
  neighbor "192.0.2.2" {
    group "internal"
  }
}
```

The configuration on RR PE-2 is as follows:

```
# on PE-2:
configure exclusive
router "Base" {
  autonomous-system 64496
  bgp {
    cluster {
      cluster-id 1.1.1.1
    }
    ebgp-default-reject-policy {
      import false
      export false
    }
    group "internal" {
      peer-as 64496
      family {
        vpn-ipv4 true
      }
    }
    neighbor "192.0.2.1" {
      group "internal"
    }
    neighbor "192.0.2.3" {
      group "internal"
    }
    neighbor "192.0.2.4" {
      group "internal"
    }
    neighbor "192.0.2.5" {
      group "internal"
    }
    neighbor "192.0.2.6" {
      group "internal"
    }
  }
}
```

Import and export policies are configured on PE-3 and PE-6, as follows:

```
# on PE-3 and PE-6:
configure exclusive
policy-options {
  community "VPN1" {
    member "target:64496:1" { }
  }
}
policy-statement "VPN1-export" {
  entry 10 {
    from {
      protocol {
        name [direct]
      }
    }
    to {
      protocol {
        name [bgp-vpn]
      }
    }
    action {
      action-type accept
      community {
        add ["VPN1"]
      }
    }
  }
}
```

```
    }  
  }  
  policy-statement "VPN1-import" {  
    entry 10 {  
      from {  
        community {  
          name "VPN1"  
        }  
        protocol {  
          name [bgp-vpn]  
        }  
      }  
      action {  
        action-type accept  
      }  
    }  
  }  
}
```

VPRN 1 is configured on PE-3 and PE-6, as follows:

```
# on PE-3:  
configure exclusive  
  service {  
    vprn "1" {  
      admin-state enable  
      customer "1"  
      service-id 1  
      bgp-ipvpn {  
        mpls {  
          admin-state enable  
          route-distinguisher "64496:13"  
          vrf-import {  
            policy ["VPN1-import"]  
          }  
          vrf-export {  
            policy ["VPN1-export"]  
          }  
          auto-bind-tunnel {  
            resolution filter  
            resolution-filter {  
              ldp true  
            }  
          }  
        }  
      }  
    }  
  }  
  interface "loopback1" {  
    loopback true  
    ipv4 {  
      primary {  
        address 192.0.1.3  
        prefix-length 32  
      }  
    }  
  }  
}
```

The configuration on PE-6 is similar, with route-distinguisher 64496:16 and loopback address 192.0.1.6/32.

## Default TTL handling in VPRN

The default configuration for TTL propagation on the iLER corresponds to pipe mode, as follows:

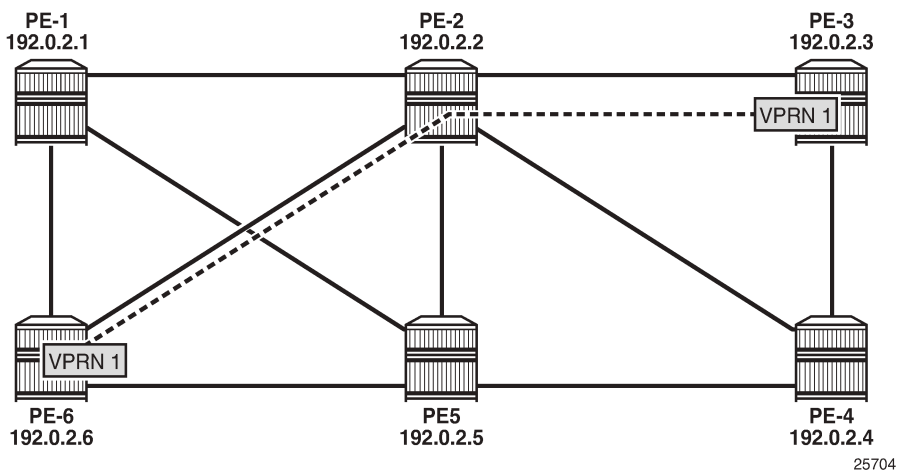
```
# on PE-3:
configure exclusive
  router "Base" {
    ttl-propagate {
      info detail
    }
  }
[ex:/configure router "Base" ttl-propagate]
A:admin@PE-3# info detail
## apply-groups
## apply-groups-exclude
label-route-local none
label-route-transit none
lsr-label-route none
vprn-local vc-only
vprn-transit vc-only
```

By default, no ICMP tunneling is enabled in the LSRs, which implies that no ICMP "Time exceeded" messages will be tunneled by the LSR to the eLER. A traceroute message sent in VPRN 1 from PE-3 to the loopback address in VPRN 1 on PE-6 shows that the loopback address is the next hop. There are no intermediate hops detected.

```
[/]
A:admin@PE-3# traceroute 192.0.1.6 router-instance "1"
traceroute to 192.0.1.6, 30 hops max, 40 byte packets
 1 192.0.1.6 (192.0.1.6) 1.23 ms 1.13 ms 1.16 ms
```

**Figure 144: Tunnel from iLER PE-3 to eLER PE-6 via LSR PE-2** shows the tunnel from iLER PE-3 to eLER PE-6:

*Figure 144: Tunnel from iLER PE-3 to eLER PE-6 via LSR PE-2*



For comparison, a traceroute in the base router toward the system address of PE-6 shows PE-2 as intermediate hop, as follows:

```
[/]
```



```
A:admin@PE-3# traceroute 192.0.2.6
traceroute to 192.0.2.6, 30 hops max, 40 byte packets
 1 192.168.23.1 (192.168.23.1)  0.856 ms  0.987 ms  0.854 ms
 2 192.0.2.6 (192.0.2.6)      1.16 ms  1.54 ms  1.10 ms
```

## Uniform mode in iLER and ICMP tunneling in LSR

In the iLER PE-3, uniform mode is enabled for local VPRNs, as follows:

```
# on iLER PE-3:
configure exclusive
  service {
    vprn "1" {
      ttl-propagate {
        local all
      }
    }
  }
```

This is a specific configuration for VPRN 1 that overrides the global configuration. By default, it is set to inherit the global configuration. By default, the global configuration is pipe mode.

This TTL propagation is only configured on PE-3, not on PE-6. This implies that traceroute messages from the VPRN in PE-3 will have TTL propagation to all MPLS labels (uniform mode), while traceroute messages from the VPRN in PE-6 will have pipe mode.

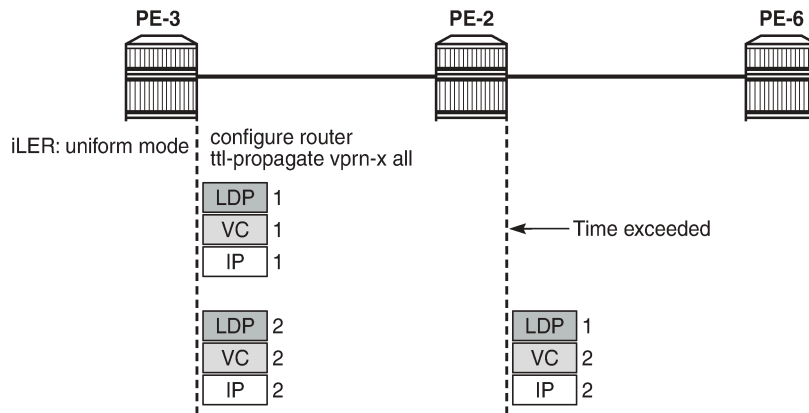
In the LSR PE-2, ICMP tunneling is enabled, as follows:

```
# on LSR PE-2:
configure exclusive
  router "Base" {
    icmp-tunneling true
  }
```

A UDP traceroute is sent from VPRN 1 on PE-3 to the loopback address in VPRN 1 on PE-6. A message with TTL 1 is sent first. The IP TTL and VC TTL are not decremented by the LSR PE-2. Only the LDP TTL is decremented, so this message times out on the LSR PE-2, which tunnels the ICMP Time Exceeded reply with destination address 192.0.1.3 toward eLER PE-6. The eLER looks up the prefix 192.0.1.3 in the VRF table and sends the ICMP Time Exceeded reply toward VPRN 1 in PE-3. Three UDP traceroute messages with TTL 1 are sent. Then, UDP traceroute messages with TTL 2 are sent. They reach the destination PE-6 before time-out.

**Figure 145: UDP traceroute in VPRN with iLER in uniform mode** shows the TTLs in the UDP traceroute messages:

Figure 145: UDP traceroute in VPRN with iLER in uniform mode



25705

In the following output, there is only an MPLS label stack object when TTL expires in the LSR, because ICMP tunneling is occurring. The MPLS label stack object is not used by the eLER. The LSR where ICMP tunneling occurs adds an MPLS label stack object to the ICMP reply message. The MPLS label stack object contains information about the MPLS labels (VC label and LDP transport label) in the stack: MPLS labels, experimental bits for ToS, and TTL. S indicates the bottom of the label stack. In the detailed output of the **traceroute** command, the MPLS label stack information is shown for the echo requests that timed out in the LSR:

```
[/]
A:admin@PE-3# traceroute 192.0.1.6 router-instance "1" detail
traceroute to 192.0.1.6, 30 hops max, 40 byte packets
 1  1  192.168.26.1 (192.168.26.1)  2.03 ms
      returned MPLS Label Stack Object
          entry 1: MPLS Label = 524282, Exp = 7, TTL =  1, S = 0
          entry 2: MPLS Label = 524280, Exp = 7, TTL =  1, S = 1
 1  2  192.168.26.1 (192.168.26.1)  1.70 ms
      returned MPLS Label Stack Object
          entry 1: MPLS Label = 524282, Exp = 7, TTL =  1, S = 0
          entry 2: MPLS Label = 524280, Exp = 7, TTL =  1, S = 1
 1  3  192.168.26.1 (192.168.26.1)  1.58 ms
      returned MPLS Label Stack Object
          entry 1: MPLS Label = 524282, Exp = 7, TTL =  1, S = 0
          entry 2: MPLS Label = 524280, Exp = 7, TTL =  1, S = 1
 2  1  192.0.1.6 (192.0.1.6)  1.70 ms
 2  2  192.0.1.6 (192.0.1.6)  1.72 ms
 2  3  192.0.1.6 (192.0.1.6)  1.76 ms
```

The top label or transport label 524282 is the LDP label pushed by PE-3:

```
[/]
A:admin@PE-3# show router ldp bindings active prefixes prefix 192.0.2.6/32

=====
LDP Bindings (IPv4 LSR ID 192.0.2.3)
(IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
```

```

FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
  (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
--                                  524282
192.168.23.1                         1/1/c1/2

192.0.2.6/32                         Swap
524282                               524282
192.168.23.1                         1/1/c1/2

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

The bottom label or service label 524280 is the BGP label, which remains the same end-to-end:

```

[/]
A:admin@PE-3# show router bgp routes vpn-ipv4
=====
BGP Router ID:192.0.2.3      AS:64496      Local AS:64496
=====
Legend -
Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                l - leaked, x - stale, > - best, b - backup, p - purge
Origin codes  : i - IGP, e - EGP, ? - incomplete
=====
BGP VPN-IPv4 Routes
=====
Flag  Network                               LocalPref  MED
      Nexthop (Router)                     Path-Id    IGP Cost
      As-Path                               Label
-----
i     64496:13:192.0.1.3/32                 100       None
      192.0.2.3                             None       0
      No As-Path                             524280
u*>i 64496:16:192.0.1.6/32                 100       None
      192.0.2.6                             None       20
      No As-Path                             524280
-----
Routes : 2
=====
  
```

When the iLER is configured in pipe mode (vc-only) or if there is no TTL propagation to any MPLS label (none), the output of the **traceroute detail** command does not contain the MPLS label stack object information. In pipe mode, the IP TTL is propagated to the VC header, but not to the LDP header. When the TTL propagation is none, the IP TTL is not propagated to VC or LDP. The LSRs are invisible and there will be missing hops.

```
# on PE-3:
```

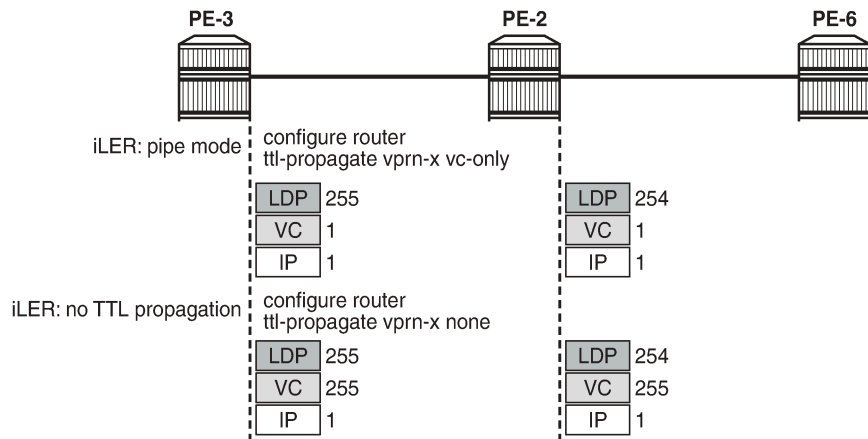
```

configure exclusive
  service {
    vprn "1" {
      ttl-propagate {
        local vc-only
      }
    }
  }

[/]
A:admin@PE-3# traceroute 192.0.1.6 router-instance "1" detail
traceroute to 192.0.1.6, 30 hops max, 40 byte packets
 1  1  192.0.1.6  (192.0.1.6)  2.02 ms
 1  2  192.0.1.6  (192.0.1.6)  2.05 ms
 1  3  192.0.1.6  (192.0.1.6)  2.03 ms
  
```

The reason is that the TTL of the LDP header is 255 at the iLER and it is decremented by one in every LSR. The UDP traceroute message will not time out on the LSR PE-2. The TTL of the VC header is not decremented in the LSR. When the traceroute message does not time out in PE-2, the hop PE-2 is invisible. In a similar way, the traceroute messages will not time out in the LSR when no TTLs are propagated to any MPLS header. [Figure 146: UDP traceroute in VPRN without TTL propagation to LDP](#) shows the TTLs in both cases:

Figure 146: UDP traceroute in VPRN without TTL propagation to LDP



25706

The TTL propagation is restored to uniform mode in the iLER as follows:

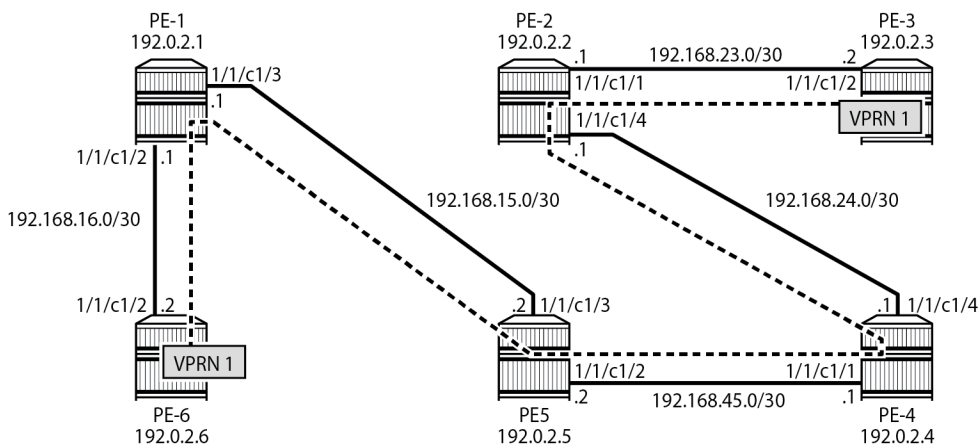
```

# on PE-3:
configure exclusive
  service {
    vprn "1" {
      ttl-propagate {
        local all
      }
    }
  }
  
```

## Uniform mode in iLER and ICMP tunneling in multiple LSRs

After disabling some ports in the nodes, the tunnel from iLER PE-3 to PE-6 has four intermediate hops (LSRs) instead of one, as shown in [Figure 147: Tunnel from iLER PE-3 to eLER PE-6 with multiple LSRs](#):

Figure 147: Tunnel from iLER PE-3 to eLER PE-6 with multiple LSRs



38581

```
# on PE-2:
configure exclusive
  port 1/1/c1/2 {
    admin-state disable
  }
  port 1/1/c1/3 {
    admin-state disable
  }
  port 1/1/c2/1 {
    admin-state disable
  }
}

# on PE-3:
configure exclusive
  port 1/1/c1/1 {
    admin-state disable
  }

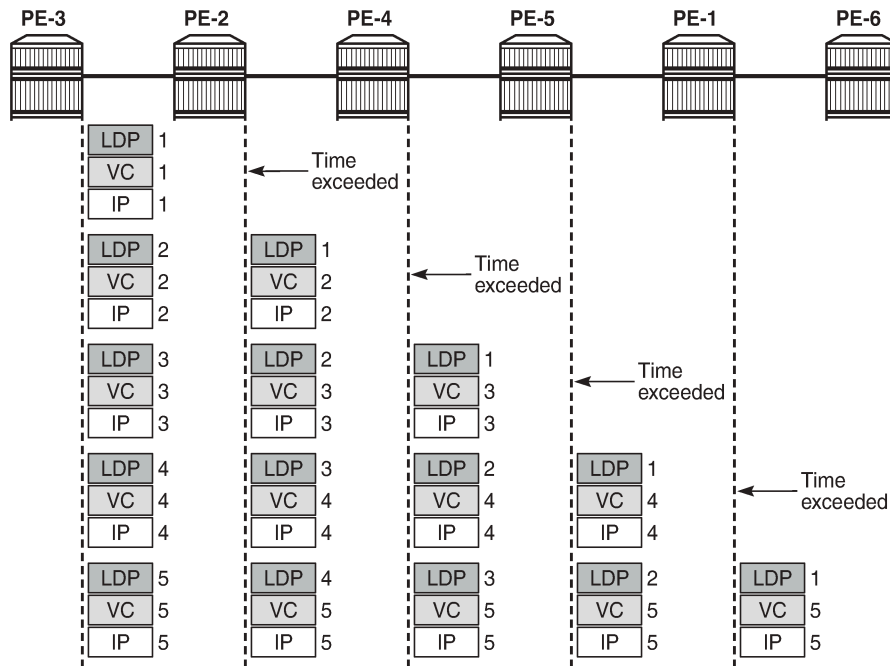
# on PE-5:
configure exclusive
  port 1/1/c1/1 {
    admin-state disable
  }
```

TTL propagation on the iLER PE-3 is in uniform mode. ICMP tunneling needs to be enabled on all LSRs, as follows:

```
# on all LSRs:
configure exclusive
  router "Base" {
    icmp-tunneling true
  }
```

UDP traceroute messages are sent from VPRN 1 on iLER PE-3 to VPRN 1 on PE-6, as shown in [Figure 148: UDP traceroute with iLER in uniform mode](#):

Figure 148: UDP traceroute with iLER in uniform mode



25708

The detailed output of the **traceroute** command shows the MPLS label stack object information as added by the LSR where the ICMP Time Exceeded message was tunneled. For brevity, only the first of the three messages from each intermediate node is shown.

```
[/]
A:admin@PE-3# traceroute 192.0.1.6 router-instance "1" detail
traceroute to 192.0.1.6, 30 hops max, 40 byte packets
 1  1  192.168.24.1 (192.168.24.1)  2.04 ms
    returned MPLS Label Stack Object
      entry 1: MPLS Label = 524286, Exp = 7, TTL =  1, S = 0
      entry 2: MPLS Label = 524280, Exp = 7, TTL =  1, S = 1
---snip---
 2  1  192.168.45.1 (192.168.45.1)  2.29 ms
    returned MPLS Label Stack Object
      entry 1: MPLS Label = 524281, Exp = 7, TTL =  1, S = 0
      entry 2: MPLS Label = 524280, Exp = 7, TTL =  2, S = 1
---snip---
 3  1  192.168.15.2 (192.168.15.2)  2.02 ms
    returned MPLS Label Stack Object
      entry 1: MPLS Label = 524282, Exp = 7, TTL =  1, S = 0
      entry 2: MPLS Label = 524280, Exp = 7, TTL =  3, S = 1
---snip---
 4  1  192.168.16.1 (192.168.16.1)  1.96 ms
    returned MPLS Label Stack Object
      entry 1: MPLS Label = 524282, Exp = 7, TTL =  1, S = 0
      entry 2: MPLS Label = 524280, Exp = 7, TTL =  4, S = 1
---snip---
 5  1  192.0.1.6 (192.0.1.6)  1.96 ms
---snip---
```

The TTL for the bottom MPLS header (BGP) is not decremented in each hop; only the TTL for the transport MPLS header (LDP) is decremented. The bottom label or BGP label of 524280 is not changed end-to-end. The LDP transport label for the different nodes is as follows.

For iLER PE-3, the LDP transport label for traffic toward PE-6 is 524286:

```
[/]
A:admin@PE-3# show router ldp bindings active prefixes prefix 192.0.2.6/32

=====
LDP Bindings (IPv4 LSR ID 192.0.2.3)
  (IPv6 LSR ID ::)
=====
---snip---
=====
LDP IPv4 Prefix Bindings (Active)
=====
Prefix                               Op
IngLbl                               EgrLbl
EgrNextHop                           EgrIf/LspId
-----
192.0.2.6/32                         Push
--                                   524286
192.168.23.1                         1/1/c1/2

192.0.2.6/32                         Swap
524285                                524286
192.168.23.1                         1/1/c1/2
-----
No. of IPv4 Prefix Active Bindings: 2
=====
```

For LSR PE-2, the LDP transport label toward PE-6 is 524281:

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

=====
LDP Bindings (IPv4 LSR ID 192.0.2.2)
  (IPv6 LSR ID ::)
=====
---snip---
=====
LDP IPv4 Prefix Bindings (Active)
=====
Prefix                               Op
IngLbl                               EgrLbl
EgrNextHop                           EgrIf/LspId
-----
192.0.2.6/32                         Push
--                                   524281
192.168.24.2                         1/1/c1/4

192.0.2.6/32                         Swap
524286                                524281
192.168.24.2                         1/1/c1/4
-----
No. of IPv4 Prefix Active Bindings: 2
=====
```

For LSR PE-4, the LDP transport label toward PE-6 is 524282:

```
[/]
A:admin@PE-4# show router ldp bindings active prefixes prefix 192.0.2.6/32

=====
LDP Bindings (IPv4 LSR ID 192.0.2.4)
(IPv6 LSR ID ::)
=====
---snip---
=====
LDP IPv4 Prefix Bindings (Active)
=====
Prefix                               Op
IngLbl                               EgrLbl
EgrNextHop                           EgrIf/LspId
-----
192.0.2.6/32                         Push
--                                   524282
192.168.45.2                         1/1/c1/1

192.0.2.6/32                         Swap
524281                               524282
192.168.45.2                         1/1/c1/1

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

For LSR PE-5, the LDP transport label toward PE-6 is 524282:

```
[/]
A:admin@PE-5# show router ldp bindings active prefixes prefix 192.0.2.6/32

=====
LDP Bindings (IPv4 LSR ID 192.0.2.5)
(IPv6 LSR ID ::)
=====
---snip---
=====
LDP IPv4 Prefix Bindings (Active)
=====
Prefix                               Op
IngLbl                               EgrLbl
EgrNextHop                           EgrIf/LspId
-----
192.0.2.6/32                         Push
--                                   524282
192.168.15.1                         1/1/c1/3

192.0.2.6/32                         Swap
524282                               524282
192.168.15.1                         1/1/c1/3

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



For LSR PE-1, the LDP transport label toward PE-6 is 524287, but this label will not be present in the traceroute detailed output, because this message cannot time out on an LSR where ICMP tunneling takes place:

```
[/]
A:admin@PE-1# show router ldp bindings active prefixes prefix 192.0.2.6/32

=====
LDP Bindings (IPv4 LSR ID 192.0.2.1)
      (IPv6 LSR ID ::)
=====
---snip---
=====
LDP IPv4 Prefix Bindings (Active)
=====
Prefix                               Op
IngLbl                               EgrLbl
EgrNextHop                           EgrIf/LspId
-----
192.0.2.6/32                         Push
--                                   524287
192.168.16.2                         1/1/c1/2

192.0.2.6/32                         Swap
524282                               524287
192.168.16.2                         1/1/c1/2

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

## Conclusion

Tunneling of ICMP reply messages over MPLS LSPs provides the ability to trace the hops in an MPLS tunnel. This mechanism applies to VPRN, 6PE/6VPE, and BGP labeled routes.

ICMP tunneling at an LSR applies to UDP traceroute packets that time out at the LSR. The ICMP Time Exceeded message is tunneled by the LSR toward the eLER and the eLER routes the packet to the sender of the traceroute message.

# Unnumbered Interfaces in RSVP-TE and LDP

This chapter provides information about unnumbered interfaces in RSVP-TE and LDP.

Topics in this chapter include:

- [Applicability](#)
- [Overview](#)
- [Configuration](#)
- [Conclusion](#)

## Applicability

This chapter is applicable to SR OS routers and was initially written for SR OS Release 13.0.R7. The CLI in this edition corresponds to SR OS Release 22.10.R3. SR OS supports unnumbered interfaces in RSVP-TE and LDP in SR OS Release 11.0.R1 or later.

## Overview

Unnumbered interfaces enable IP processing on a point-to-point (P2P) interface without an explicit IP address. Unnumbered interfaces are supported in ReSource reservation Protocol with Traffic Engineering (RSVP-TE) and Label Distribution Protocol (LDP).

An unnumbered interface is uniquely identified in the network by the tuple (Router ID, If Index), where the interface index (If Index) is unique on the router. The two endpoints of an unnumbered link exchange the If Index that they assigned to the link.

The (Router ID, If Index) tuple is used by the following:

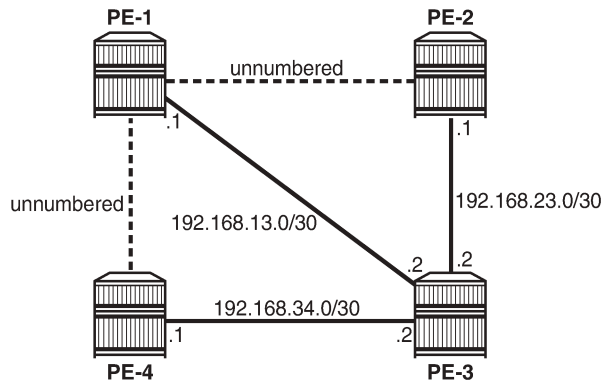
- Intermediate System to Intermediate System (IS-IS) or Open Shortest Path First (OSPF) to advertise link information
- RSVP to signal Label Switched Paths (LSPs) over this unnumbered interface
- LDP to establish hello adjacencies and resolve Forwarding Equivalence Classes (FECs)
- Operations, Administration, and Maintenance (OAM) to send or respond to a Multi-Protocol Label Switching (MPLS) echo request over an unnumbered interface

The unnumbered interface can "borrow" the IP address of another interface on the node.

The borrowed IP address is used exclusively as the source address for IP packets that are originated from the unnumbered interface. The borrowed IP address defaults to the system loopback interface address, but can be changed manually. The borrowed IP address corresponds to the Router ID in the tuple representing the unnumbered interface.

The configuration used in this chapter is shown in [Figure 149: Example topology for unnumbered interfaces in RSVP and LDP](#). There are two unnumbered links: one between PE-1 and PE-2, and another between PE-1 and PE-4. The remaining links are numbered.

Figure 149: Example topology for unnumbered interfaces in RSVP and LDP



25683

Configure unnumbered interfaces as follows:

```
configure {
  router <router-name> {
    interface <interface-name> {
      ipv4 {
        unnumbered {
          system|ip-address <ip-address>|ip-int-name <ip-int-name>
        }
      }
    }
  }
  exit all
}
```

To configure an unnumbered link with the system address as the borrowed IP address, no address needs to be configured:

```
# on PE-4: configure {
  router "Base" {
    interface "int-PE-4-PE-1" {
      port 1/1/c1/1
      ipv4 {
        unnumbered {
          system
        }
      }
    }
  }
  exit all
}
```

An unnumbered interface has to be a P2P link.

## Unnumbered interfaces in IS-IS

Unnumbered interfaces are identified in IS-IS by a combination of the system ID and an extended local circuit ID, as described in RFC 5307, *IS-IS Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)*.

Enable debugging on the unnumbered interface on PE-4 (int-PE-4-PE-1) as follows:

```
# on PE-4:
debug
  router
    isis
      packet "int-PE-4-PE-1" detail
```

```
exit all
```

The following IS-IS hello Protocol Data Unit (PDU) is received from PE-1. The I/F Address is the borrowed IP address; in this case, the system address of PE-1, because the interface is unnumbered.

```
# on PE-4:
2 2023/03/21 12:42:33.919 UTC MINOR: DEBUG #2001 Base ISIS PKT
"ISIS PKT:
RX ISIS PDU ifId 2 len 52:
  DMAC       : 09:00:2b:00:00:05
  Proto Disc : 131
  Header Len : 20
  Version PID : 1
  ID Length  : 0
  Version    : 1
  Reserved   : 0
  Max Area Addr : 3
  PDU Type   : (11) Point-2-Point IS-IS Hello Pdu
  Circuit Type : L1
  Source Id   : 19 20 00 00 20 01
  Hold Time   : 27
  Packet length : 52
  Circuit Id   : 0
  Area Addresses:
    Area Address : (3) 49.0001
  Supp Protocols:
    Protocols    : IPv4
  I/F Addresses :
    I/F Address : 192.0.2.1
  3Way Adjacency :
    State        : UP
    Ext ckt ID   : 4
    NbrSysID    : 19 20 00 00 20 04
    Nbr ext ckt ID : 2
"
```

The three-way adjacency contains the neighbor extended local circuit ID (**Nbr ext ckt ID: 2**). This is the local interface index of the unnumbered interface (int-PE-4-PE-1), which can be verified as follows:

```
[/]
A:admin@PE-4# show router interface "int-PE-4-PE-1" detail | match "If Index"
If Index      : 2                Virt. If Index : 2
Last Oper Chg : 03/21/2023 23:55:10 Global If Index : 1
```

On PE-1, the interface toward PE-4 has a different index, as follows:

```
[/]
A:admin@PE-1# show router interface "int-PE-1-PE-4" detail | match "If Index"
If Index      : 4                Virt. If Index : 4
Last Oper Chg : 03/21/2023 23:54:36 Global If Index : 3
```

For numbered interfaces, such as int-PE-4-PE-3, the I/F Address is the interface address; in that case, 192.168.34.1, for messages received from PE-3 instead of the Router ID.

When a Shared Risk Link Group (SRLG) is configured in combination with IS-IS and unnumbered interfaces, the least significant bit in the flags field of the SRLG Type-Length Value (TLV) indicates that the interface is unnumbered (0) or numbered (1).

## Unnumbered interfaces in OSPF

For unnumbered interfaces in OSPF, link local and remote identifiers are defined in RFC 4203, *OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)*. The OSPF link state advertisement (LSA) is defined in RFC 2328, *OSPF version 2*.

For numbered interfaces, the link data is the IP interface address; for unnumbered interfaces, the link data is the interface index value. The value starts from 1 in the format 0.0.0.1. SR OS recognizes an unnumbered interface when the first byte in the link data has a value of 0; SR OS then treats the link data as an interface index instead of an IP address.

## Unnumbered interfaces in RSVP-TE

Unnumbered IP interfaces can be used as Traffic Engineering (TE) links for the signaling of RSVP P2P LSPs and point-to-multipoint (P2MP) LSPs.

Fast Reroute (FRR) facility backup over unnumbered interfaces is supported, whereas FRR one-to-one only uses numbered interfaces in the detour path.

The unnumbered IP address is advertised by IS-IS or OSPF, and Constrained Shortest Path First (CSPF) can include them in the computation of a path.

Unnumbered interfaces of the remote router can be specified in the Explicit Route Object (ERO), and in the Record Route Object (RRO), by a combination of Router ID (borrowed IP address) and interface ID, as defined in RFC 3477, *Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE)*.

The choice of the data interface is indicated in the Path message by including the interface identifier of the data channel. In the Path message (**PATH Msg**), the IP address equals the local borrowed IP address; in the Resv message (**RESV Msg**), the IP address is the remote borrowed IP address. As well as the borrowed IP address, there is also a Logical Interface Handle (**LIH**). This interface identification is defined in RFC 3473, *Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions*.

To see the Path and Resv messages on PE-4, enter the following debug commands:

```
# on PE-4:
debug
  router
    rsvp
      packet
        path detail
        resv detail
      exit all
```

The Path message contains the **RSVPHop** object with the local interface identifier of the data channel, as follows:

```
# on PE-4:
19 2023/03/21 13:03:45.671 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: PATH Msg
Send PATH From:192.0.2.4, To:192.0.2.2
      TTL:255, Checksum:0xa648, Flags:0x0
Session   - EndPt:192.0.2.2, TunnId:1, ExtTunnId:192.0.2.4
SessAttr  - Name:LSP-PE-4-PE-2::dyn
          SetupPri:7, HoldPri:0, Flags:0x17
```

```

RSVPHop - Ctype:3, Addr:192.1.2.4, LIH:2
           RouterId :192.0.2.4, InterfaceId :2
TimeValue - RefreshPeriod:30
SendTempl - Sender:192.0.2.4, LspId:2576
SendTSpec - Ctype:QOS, CDR:0.000 bps, PBS:0.000 bps, PDR:infinity
           MPU:20, MTU:1564
LabelReq  - IfType:General, L3ProtID:2048
RRO     - Unnumbered: RouterId 192.0.2.4 InterfaceID 2, Flags:0x0
ERO     - Unnumbered RouterId 192.0.2.1, LinkId 4, Strict
           Unnumbered RouterId 192.0.2.2, LinkId 2, Strict
FRR0bj    - SetupPri:7, HoldPri:0, HopLimit:16, BW:0.000 bps, Flags:0x2
           ExcAny:0x0, IncAny:0x0, IncAll:0x0
"
    
```

The ERO and RRO objects are also shown. The unnumbered interfaces are defined by the combination of the Router ID (**RouterId**) and the interface ID (**InterfaceId**).

The Resv message also contains the **RSVPHop** object, but the address is now the remote address of PE-1 instead of the local address of PE-4, as follows:

```

# on PE-4:
23 2023/03/21 13:04:03.396 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: RESV Msg
Recv RESV From:192.1.2.1, To:192.0.2.4
           TTL:255, Checksum:0xbe23, Flags:0x0
Session   - EndPt:192.0.2.2, TunnId:1, ExtTunnId:192.0.2.4
RSVPHop - Ctype:3, Addr:192.1.2.1, LIH:2
           RouterId :192.0.2.1, InterfaceId :4
TimeValue - RefreshPeriod:30
Style     - SE
FlowSpec  - Ctype:QOS, CDR:0.000 bps, PBS:0.000 bps, PDR:infinity
           MPU:20, MTU:1560, RSpecRate:0, RSpecSlack:0
FilterSpec - Sender:192.0.2.4, LspId:2576, Label:524287
RRO     - Unnumbered: RouterId 192.0.2.1 InterfaceID 4, Flags:0x1
           Label:524287, Flags:0x1
           Unnumbered: RouterId 192.0.2.2 InterfaceID 2, Flags:0x0
           Label:524287, Flags:0x1
"
    
```

To see the Patherr (**PATHERR Msg**) and Resverr (**RESVERR Msg**) messages on PE-4, enter the following debug command:

```

# on PE-4:
debug
  router
    rsvp
      packet
        patherr detail
        resverr detail
      exit all
    
```

The Resverr message contains the following **ErrorSpec** object, as defined by RFC 3473. In this case, the error is caused by disabling TE on ingress Label Egress Router (iLER) PE-4. No route can be found to the destination because there is no lookup in the TE database. The LSP does not come up, even if CSPF is disabled on the LSP.

```

# on PE-4:
32 2023/03/21 13:05:11.398 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: RESVERR Msg
Send RESVERR From:192.1.2.4, To:192.0.2.1
           TTL:255, Checksum:0x68c1, Flags:0x0
    
```

```
Session      - EndPt:192.0.2.2, TunnId:1, ExtTunnId:192.0.2.4
RSVPHop     - Ctype:3, Addr:192.1.2.4, LIH:2
              RouterId :192.0.2.4, InterfaceId :2
ErrorSpec  - Ctype:3, ErrNode:192.1.2.4, Flags:0x0, ErrCode:3, ErrValue:0
              RouterId :192.0.2.4, InterfaceId :2
Style       - SE
FlowSpec    - Ctype:QOS, CDR:0.000 bps, PBS:0.000 bps, PDR:infinity
              MPU:20, MTU:1560, RSpecRate:0, RSpecSlack:0
FilterSpec  - Sender:192.0.2.4, LspId:2576
"
```

## Considerations for unnumbered interfaces in RSVP-TE

Consider the following for unnumbered interfaces in RSVP-TE:

- With RSVP, TE must be enabled in IS-IS or OSPF. The Router ID of the router that advertised an unnumbered interface index is obtained from the TE database. Therefore, if TE is disabled in IS-IS or OSPF, a non-CSPF LSP with the next hop for this path over an unnumbered interface does not come up. The Router ID of the neighbor that has the next hop of the Path message cannot be searched for.
  - The operational state of the LSP path remains down with reason **noRouteToDestination**.
  - If a Path message is received at the LSR in which TE is disabled and the next hop for the LSP path is over an unnumbered interface, a Patherr message is sent back to the iLER with error code 24: Routing problem; Error value 5: "No route available toward destination".
- Only FRR facility protection is supported; FRR one-to-one protection only works for numbered interfaces.
- There is no FRR facility protection if the point of local repair (PLR) is the iLER and the bypass tunnel egress interface is unnumbered.
- Bi-directional Forwarding Detection (BFD) can be enabled on an unnumbered router interface. Therefore, RSVP FRR procedures can be triggered via a BFD session timeout.
- Unnumbered interfaces cannot be configured as hops in a path. This is true for RSVP-TE LSPs, as well as for static LSPs.
- RSVP hello and hello-related capabilities, such as graceful restart helper, are not supported.
- SRLG is supported, but the user SRLG DB (**user-srlg-db**) feature at the iLER is not supported. Unnumbered interfaces cannot be added to the SRLG DB. When the user SRLG DB feature is enabled on the iLER, all unnumbered interfaces are considered as having no SRLG membership.

## Unnumbered interfaces in LDP

LDP can establish hello adjacencies and can resolve unicast and multicast FECs over unnumbered interfaces.

For link LDP, hello adjacencies are brought up using hello packets with source IP address set to the borrowed IP address and a destination IP address set to 224.0.0.2. The borrowed IP address is the system address, by default. Hello packets with the same source IP address are accepted when received over parallel unnumbered interfaces from the same peer LSR ID. The corresponding hello adjacencies are associated with a single LDP session.

The transport address for the TCP connection, which is encoded in the hello packet, is always set to the LSR ID of the node. The user can configure the **local-lsr-id** option on the interface and change the value

of the LSR ID to either the local interface or some other interface name: loopback or not, numbered or not. The transport address for the LDP session is updated with the new LSR ID.

For targeted LDP, the source and destination addresses of targeted hello packets are the LDP LSR IDs of the nodes. The user can configure the **local-lsr-id** option on the targeted session. The transport address for the LDP session and the source IP address of targeted hello messages are updated to the new LSR ID value.

LDP advertises/withdraws unnumbered interfaces using the address/address-withdraw messages. The borrowed IP address of the interface is used.

A FEC can be resolved to an unnumbered interface in the same way as it is resolved to a numbered interface. The outgoing interface and the next hop are searched for in the Routing Table Manager (RTM). The next hop consists of the Router ID and link identifier of the interface to the peer LSR. All LDP FEC types are supported. LDP FEC Equal Cost Multi-Path (ECMP) over a mix of unnumbered and numbered interfaces is supported.

RFC 5036, *LDP Specification*, describes the address list TLV that is used in the LDP address message, and the LDP address withdrawal message. For unnumbered interfaces, the borrowed IP address is used, which is typically the system address of the sender node.

On PE-1, enable debugging for LDP packets from peer 192.0.2.2 as follows:

```
# on PE-1:
debug
  router
    ldp
      peer 192.0.2.2
        packet
          init detail
          label detail
        exit all
```

The following LDP address packets are shown at PE-1:

```
6 2023/03/21 13:23:27.615 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Address packet (msgId 5) to 192.0.2.2:0
Protocol version = 1
Address Family = 1 Number of addresses = 3
Address 1 = 192.0.2.1
Address 2 = 192.1.2.1
Address 3 = 192.168.13.1
"
```

```
5 2023/03/21 13:23:27.436 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Recv Address packet (msgId 5) from 192.0.2.2:0
Protocol version = 1
Address Family = 1 Number of addresses = 3
Address 1 = 192.0.2.2
Address 2 = 192.1.2.2
Address 3 = 192.168.23.1
"
```

The received LDP address packet contains three IP addresses: the system IP address 192.0.2.2 for an unnumbered interface on the sending node, the loop back IP address 192.1.2.2 for an unnumbered loop back interface on the sending node, and the interface IP address 192.168.23.1 for a numbered interface on the sending node.



## Considerations for unnumbered interfaces in LDP

All LDP features are supported on unnumbered interfaces, except for the following:

- BFD can be enabled on an unnumbered router interface. The BFD parameters must be configured within the unnumbered **interface** context. If not, the BFD sessions are not established.
- Unnumbered interfaces cannot be added into LDP global and peer prefix policies.

## Unnumbered interfaces in OAM

The following applies to unnumbered interfaces in RSVP-TE or LDP.

The downstream mapping object is a TLV that can be included in an echo request, as described in RFC 4379, *Detecting Multi-Protocol Label Switched Data Plane Failures*.

Only one downstream mapping object may appear in an echo request. The presence of a downstream mapping object is a request that a downstream mapping object be included in the echo reply.

For unnumbered interfaces, the address type is 2 (**ipv4Unnumbered**), the downstream IP address is the borrowed IP address of the downstream LSR, and the downstream interface address is the index assigned by the upstream LSR to the interface.

The downstream detailed mapping object is a TLV that can be included in an echo request, as described in RFC 6424, *Mechanism for Performing Label Switched Path Ping (LSP Ping) over MPLS Tunnels*.

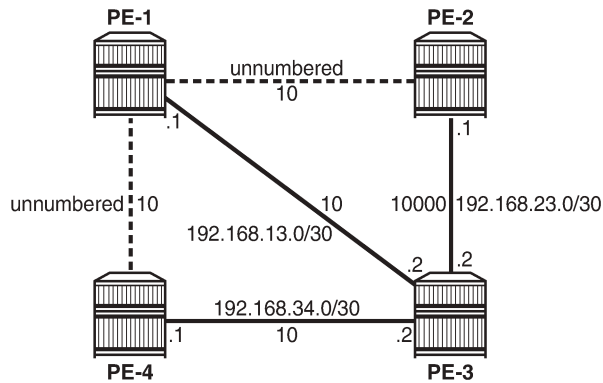
The following output shows the detailed LSP trace for an RSVP LSP from PE-4 to PE-2. Two unnumbered interfaces are used: the first between PE-4 and PE-1 and the second between PE-1 and PE-2. The interface type (**iftype**) is **ipv4Unnumbered**.

```
A:PE-4# oam lsp-trace "LSP-PE-4-PE-2" detail
lsp-trace to LSP-PE-4-PE-2: 0 hops min, 0 hops max, 116 byte packets
0 192.0.2.4
   DS 1: ipaddr=0.0.0.0 iftype=ipv4Unnumbered MRU=1564
       label[1]=524287 protocol=4(RSVP-TE)
1 192.0.2.1 rtt=4.09ms rc=8(DSRtrMatchLabel) rsc=1
   DS 1: ipaddr=0.0.0.0 ifaddr=2 iftype=ipv4Unnumbered MRU=1564
       label[1]=524287 protocol=4(RSVP-TE)
2 192.0.2.2 rtt=1.42ms rc=3(EgressRtr) rsc=1
```

## Configuration

The following configuration example is for unnumbered interfaces in RSVP and LDP; see [Figure 150: Configuration example for unnumbered Interfaces in RSVP and LDP](#). The nodes are 7750 SRs.

Figure 150: Configuration example for unnumbered Interfaces in RSVP and LDP



25684

All interfaces have a TE metric of 10, while the link between PE-2 and PE-3 has a TE metric of 10000. So, the preferred path from PE-4 to PE-2 is over the unnumbered interfaces between PE-4 and PE-1 and between PE-1 and PE-2.

## Unnumbered interfaces

Router interfaces are configured on all nodes, numbered and unnumbered. Initially, the unnumbered interfaces are configured with default settings. The following interfaces are configured on PE-1:

```
# on PE-1: configure {
  router "Base" {
    interface "int-PE-1-PE-2" {
      port 1/1/c1/1
      ipv4 {
        unnumbered {
          system
        }
      }
    }
    interface "int-PE-1-PE-3" {
      port 1/1/c1/3
      ipv4 {
        primary {
          address 192.168.13.1
          prefix-length 30
        }
      }
    }
    interface "int-PE-1-PE-4" {
      port 1/1/c1/2
      ipv4 {
        unnumbered {
          system
        }
      }
    }
  }
  interface "system" {
    ipv4 {
      primary {
        address 192.0.2.1
      }
    }
  }
}
```

```

        prefix-length 32
    }
}
exit all

```

There are two unnumbered interfaces: int-PE-1-PE-2 and int-PE-1-PE-4. There is no borrowed IP address configured for the unnumbered interfaces. So, the borrowed IP address is the system address of PE-1.

For the unnumbered interfaces, the borrowed IP address is indicated between square brackets in the following output:

```

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

=====
Interface Table (Router: Base)
=====
Interface-Name      Adm    Opr(v4/v6)  Mode    Port/SapId
IP-Address          PfxState
-----
int-PE-1-PE-2      Up     Up/Down     Network 1/1/c1/1
  Unnumbered If[system]          n/a
int-PE-1-PE-3      Up     Up/Down     Network 1/1/c1/3
  192.168.13.1/30                n/a
int-PE-1-PE-4      Up     Up/Down     Network 1/1/c1/2
  Unnumbered If[system]          n/a
system              Up     Up/Down     Network system
  192.0.2.1/32                  n/a
-----
Interfaces : 4
=====

```

Each interface, numbered or unnumbered, gets an interface index. This interface index can be retrieved as follows:

```

[/]
A:admin@PE-1# show router interface "int-PE-1-PE-2" detail | match "If Index"
If Index      : 2          Virt. If Index : 2
Last Oper Chg : 03/21/2023 23:54:36 Global If Index : 1

[/]
A:admin@PE-1# show router interface "int-PE-1-PE-3" detail | match "If Index"
If Index      : 3          Virt. If Index : 3
Last Oper Chg : 03/21/2023 23:54:36 Global If Index : 2

[/]
A:admin@PE-1# show router interface "int-PE-1-PE-4" detail | match "If Index"
If Index      : 4          Virt. If Index : 4
Last Oper Chg : 03/21/2023 23:54:36 Global If Index : 3

[/]
A:admin@PE-1# show router interface "system" detail | match "If Index"
If Index      : 1          Virt. If Index : 1
Last Oper Chg : 03/21/2023 23:54:36 Global If Index : 256

```

The unnumbered interface toward PE-2 has If Index 2, and the unnumbered interface toward PE-4 has If Index 4.

BFD can be enabled on unnumbered interfaces, as follows:

```
# on PE-4: configure {
  router "Base" {
    interface "int-PE-4-PE-1" {
      ipv4 {
        bfd {
          admin-state enable
          transmit-interval 100
          receive 100
          multiplier 3
        }
      }
    }
  }
  exit all
```

The BFD parameters must be configured within the unnumbered **interface** context. If not, the BFD sessions are not established.

An Interior Gateway Protocol (IGP) must be configured. In this case, IS-IS is chosen. OSPF could have been used equally well. TE must be enabled for unnumbered interfaces used in RSVP, even when CSPF is disabled. The IS-IS configuration on PE-1 is as follows:

```
# on PE-1: configure {
  router "Base" {
    isis 0 {
      admin-state enable
      level-capability 1
      traffic-engineering true
      area-address [49.0001]
      interface "int-PE-1-PE-2" {
        interface-type point-to-point
      }
      interface "int-PE-1-PE-3" {
        interface-type point-to-point
      }
      interface "int-PE-1-PE-4" {
        interface-type point-to-point
      }
      interface "system" {
      }
    }
  }
  exit all
```

An unnumbered interface has to be a P2P link.

The TE database contains the Router ID and the If Index for unnumbered interfaces, as follows:

```
[/]
A:admin@PE-1# show router isis database PE-2.00-00 detail

=====
Rtr Base ISIS Instance 0 Database (detail)
=====

Displaying Level 1 database
-----
LSP ID   : PE-2.00-00          Level    : L1
Sequence : 0x2                Checksum : 0x1c5a   Lifetime : 1049
Version  : 1                  Pkt Type : 18      Pkt Ver  : 1
Attributes: L1                Max Area : 3        Alloc Len: 126
SYS ID   : 1920.0000.2002     SysID Len: 6        Used Len  : 126

TLVs :
  Area Addresses:
    Area Address : (3) 49.0001
```

```

Supp Protocols:
  Protocols      : IPv4
  IS-Hostname    : PE-2
  Router ID     :
    Router ID    : 192.0.2.2
  I/F Addresses :
    I/F Address  : 192.168.23.1
    I/F Address  : 192.0.2.2
  TE IS Nbrs   :
    Nbr         : PE-1.00
    Default Metric : 10
    Sub TLV Len  : 10
    LcLId       : 2
    RmtId       : 2
  TE IS Nbrs   :
    Nbr         : PE-3.00
    Default Metric : 10000
    Sub TLV Len  : 12
    IF Addr     : 192.168.23.1
    Nbr IP      : 192.168.23.2
---snip---
=====

```

PE-2 has an unnumbered interface toward neighbor PE-1 (**Nbr: PE-1.00**), with local interface index 2 (**LcLId: 2**) and remote interface index 2 (**RmtId: 2**). For the numbered interface toward neighbor PE-3, the local and remote interface IP addresses are shown (**IF Addr** and **Nbr IP**), not the interface index.

## Unnumbered interfaces in RSVP

MPLS and RSVP must be enabled on the interfaces on the nodes. TE metrics are configured on the MPLS interfaces. For node PE-4, the configuration is as follows:

```

# on PE-4: configure {
  router "Base" {
    mpls {
      admin-state enable
      interface "int-PE-4-PE-1" {
        te-metric 10
      }
      interface "int-PE-4-PE-3" {
        te-metric 10
      }
    }
    rsvp {
      admin-state enable
      interface "int-PE-4-PE-1" {
      }
      interface "int-PE-4-PE-3" {
      }
    }
  }
}
exit all

```

An LSP is configured from PE-4 to PE-2 with CSPF enabled and using the TE metrics, not the IGP metrics. Unnumbered interfaces cannot be configured as hops in a path. A dynamic path **"dyn"**, without any hops, is configured to be used in an LSP from PE-4 to PE-2, as follows:

```

# on PE-4: configure {
  router "Base" {
    mpls {

```

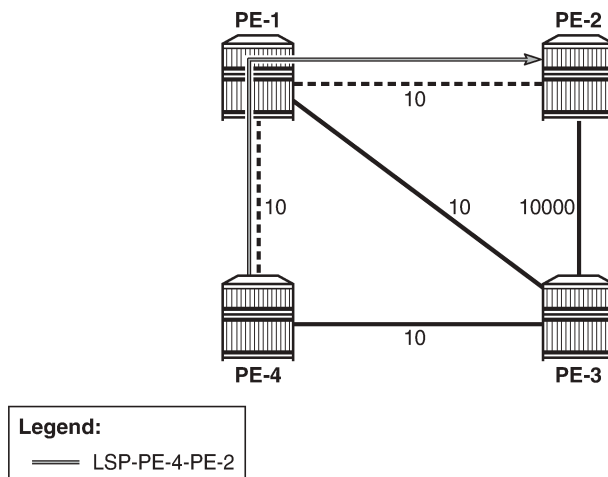
```

path "dyn" {
    admin-state enable
}
lsp "LSP-PE-4-PE-2" {
    admin-state enable
    type p2p-rsvp
    to 192.0.2.2
    path-computation-method local-cspf
    metric-type te
    primary "dyn" {
    }
}
exit all

```

The LSP from PE-4 to PE-2 has TE metric 20 when the next hop is PE-1, and TE metric 30 or 10010 when the next hop is PE-3. [Figure 151: LSP-PE-4-PE-2 on unnumbered interfaces](#) shows LSP-PE-4-PE-2, which uses only unnumbered interfaces.

Figure 151: LSP-PE-4-PE-2 on unnumbered interfaces



25685

The following tunnel table shows a next hop `int-PE-4-PE-1`, which implies that it is an unnumbered interface. The only unnumbered interface at PE-4 is `int-PE-4-PE-1`. The metric in this tunnel table is **16777215** because the IGP metric is not used.

```

[/]
A:admin@PE-4# show router tunnel-table

=====
IPv4 Tunnel Table (Router: Base)
=====
Destination      Owner      Encap TunnelId  Pref  Nexthop      Metric
  Color
-----
192.0.2.2/32     rsvp      MPLS  1          7      int-PE-4-PE-1 16777215
-----
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 actual and computed hops can be verified, as well as the CSPF metric (TE metric), as follows:

```
[/]
A:admin@PE-4# show router mpls lsp "LSP-PE-4-PE-2" path detail

=====
MPLS LSP LSP-PE-4-PE-2 Path (Detail)
=====
Legend :
  @ - Detour Available          # - Detour In Use
  b - Bandwidth Protected      n - Node Protected
  s - Soft Preemption
  S - Strict                    L - Loose
  A - ABR                       + - Inherited
=====

-----
LSP LSP-PE-4-PE-2
Path dyn
-----
LSP Name      : LSP-PE-4-PE-2
From          : 192.0.2.4
To            : 192.0.2.2
Admin State   : Up              Oper State      : Up
Path Name     : dyn
Path LSP ID   : 12288           Path Type       : Primary
Path Admin    : Up              Path Oper       : Up
Out Interface : 1/1/c1/1        Out Label      : 524287
---snip---
Explicit Hops :
  No Hops Specified
Actual Hops   :
  192.0.2.4, If Index : 2          Record Label    : N/A
  -> 192.0.2.1, If Index : 4       Record Label    : 524287
  -> 192.0.2.2, If Index : 2       Record Label    : 524287
Computed Hops :
  192.0.2.4, If Index : 2(S)
  -> 192.0.2.1, If Index : 4(S)
  -> 192.0.2.2, If Index : 2(S)
Resignal Eligible: False
Last Resignal   : n/a           CSPF Metric     : 20
=====
```

The computed hops are strict hops, as indicated by **(S)**. Because the interfaces are unnumbered, the system address and the If Index are displayed. The CSPF metric is 20.

## Configuring the borrowed IP address

The borrowed IP address does not need to be the system address, but the address must exist on the node. When the unnumbered interface is configured with a borrowed IP address that does not exist on the node, the interface goes down. This can be verified by assigning a non-existent address to the unnumbered interface int-PE-1-PE-2, as follows:

```
# on PE-1: configure {
  router "Base" {
    interface "int-PE-1-PE-2" {
      ipv4 {
        unnumbered {
          ip-address 192.1.2.1
        }
      }
    }
  }
}
```

The operational state of this interface goes down, which can be verified as follows:

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

=====
Interface Table (Router: Base)
=====
Interface-Name          Adm    Opr(v4/v6)  Mode    Port/SapId
IP-Address              PfxState
-----
int-PE-1-PE-2         Up     Down/Down  Network 1/1/c1/1
Unnumbered If[192.1.2.1]
192.168.13.1/30        Up      Up/Down     Network  1/1/c1/3
192.168.13.1/30        n/a
int-PE-1-PE-4          Up      Up/Down     Network  1/1/c1/2
Unnumbered If[system]
system                  Up      Up/Down     Network  system
192.0.2.1/32           n/a
-----
Interfaces : 4
=====
```

The borrowed IP address is indicated between square brackets. The interface is down because the IP address is not known on PE-1. The down reason code **noIfAddress** can be retrieved as follows:

```
[/]
A:admin@PE-1# show router interface "int-PE-1-PE-2" detail | match "Down Reason Code"
Down Reason Code : noIfAddress
```

The IP address can be configured as a loopback address on PE-1 and assigned to all unnumbered interfaces, as follows:

```
# on PE-1: configure {
  router "Base" {
    interface "loopback1" {
      loopback
      ipv4 {
        primary {
          address 192.1.2.1
          prefix-length 32
        }
      }
    }
    interface "int-PE-1-PE-2" {
      port 1/1/c1/1
      ipv4 {
        unnumbered {
          ip-address 192.1.2.1
        }
      }
    }
    interface "int-PE-1-PE-4" {
      port 1/1/c1/2
      ipv4 {
        unnumbered {
          ip-address 192.1.2.1
        }
      }
    }
  }
}
exit all
```



When the borrowed IP address is known on node PE-1, the unnumbered interface is operationally up, which can be verified as follows:

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

=====
Interface Table (Router: Base)
=====
Interface-Name      Adm    Opr(v4/v6)  Mode    Port/SapId
IP-Address          PfxState
-----
int-PE-1-PE-2      Up     Up/Down    Network 1/1/c1/1
Unnumbered If[192.1.2.1]
192.168.13.1/30    Up      Up/Down     Network  1/1/c1/3
n/a
int-PE-1-PE-4      Up     Up/Down    Network 1/1/c1/2
Unnumbered If[192.1.2.1]
loopback1         Up     Up/Down    Network loopback
192.1.2.1/32      n/a
system             Up      Up/Down     Network  system
192.0.2.1/32      n/a
-----
Interfaces : 5
=====
```

In a similar way, the borrowed IP address on PE-2 is configured as 192.1.2.2 and on PE-4 as 192.1.2.4.

## TE required for unnumbered interfaces in RSVP

For unnumbered interfaces, the IGP looks up the Router ID in the TE database. Therefore, TE must be enabled even if CSPF is disabled.

TE is disabled in IS-IS and CSPF is disabled in the LSP on PE-4, as follows:

```
# on PE-4: configure {
  router "Base" {
    isis 0 {
      delete traffic-engineering
    }
    mpls {
      lsp "LSP-PE-4-PE-2" {
        admin-state disable
        delete metric-type
        delete path-computation-method
      }
    }
  }
exit all

configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-4-PE-2" {
        admin-state enable
      }
    }
  }
exit all
```

LSP-PE-4-PE-2 is operationally down with failure code **noRouteToDestination**, which can be verified as follows:

```
[/]
A:admin@PE-4# show router mpls lsp "LSP-PE-4-PE-2" path detail

=====
MPLS LSP LSP-PE-4-PE-2 Path (Detail)
=====
Legend :
  @ - Detour Available          # - Detour In Use
  b - Bandwidth Protected      n - Node Protected
  s - Soft Preemption
  S - Strict                    L - Loose
  A - ABR                      + - Inherited
=====
-----
LSP LSP-PE-4-PE-2
Path dyn
-----
LSP Name      : LSP-PE-4-PE-2
From          : 192.0.2.4
To            : 192.0.2.2
Admin State   : Up              Oper State      : Down
Path Name     : dyn
Path LSP ID   : 12294           Path Type       : Primary
Path Admin    : Up              Path Oper       : Down
---snip---
MetricType    : igp             Oper MetricType : N/A
---snip---
Failure Code : noRouteToDestination
Failure Node  : 192.0.2.4
---snip---
=====
```

The configuration is restored by enabling TE in IS-IS and CSPF in the **lsp** context, as follows:

```
# on PE-4: configure {
  router "Base" {
    isis 0 {
      traffic-engineering true
    }
    mpls {
      lsp "LSP-PE-4-PE-2" {
        admin-state disable
        path-computation-method local-cspf
        metric-type te
      }
    }
  }
exit all

configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-4-PE-2" {
        admin-state enable
      }
    }
  }
exit all
```

## FRR facility

FRR facility is enabled on the LSP as follows:

```
# on PE-4: configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-4-PE-2" {
        fast-reroute {
          frr-method facility
        }
      }
    }
  }
}
```

The following LSP path detail output shows where an FRR detour is available (@) and in which node a bypass tunnel originates. The letter **n** indicates that a node is protected, as in hop 192.0.2.4. When there is a detour available, but there is no **n**, link protection is available, as in hop 192.0.2.1:

```
[/]
A:admin@PE-4# show router mpls lsp "LSP-PE-4-PE-2" path detail

=====
MPLS LSP LSP-PE-4-PE-2 Path (Detail)
=====
Legend :
@ - Detour Available           # - Detour In Use
b - Bandwidth Protected         n - Node Protected
s - Soft Preemption
S - Strict                      L - Loose
A - ABR                         + - Inherited

-----
LSP LSP-PE-4-PE-2
Path dyn
-----
LSP Name      : LSP-PE-4-PE-2
From          : 192.0.2.4
To           : 192.0.2.2
Admin State   : Up
Path Name     : dyn
Path LSP ID   : 12296
Path Admin    : Up
Out Interface : 1/1/c1/1
Oper State    : Up
Path Type     : Primary
Path Oper     : Up
Out Label     : 524286
---snip---
Explicit Hops :
  No Hops Specified
Actual Hops    :
  192.0.2.4, If Index : 2 @ n
  -> 192.0.2.1, If Index : 4 @
  -> 192.0.2.2, If Index : 2
Computed Hops  :
  192.0.2.4, If Index : 2(S)
  -> 192.0.2.1, If Index : 4(S)
  -> 192.0.2.2, If Index : 2(S)
Resignal Eligible: False
Last Resignal   : n/a
CSPF Metric     : 20
---snip---
=====
* indicates that the corresponding row element may have been truncated.
```

Information about the bypass tunnel originating in PE-4 can be retrieved as follows:

```
[/]
A:admin@PE-4# show router mpls bypass-tunnel protected-lsp detail

=====
MPLS Bypass Tunnels (Detail)
=====
-----
bypass-node192.0.2.1-61441
-----
To          : 192.168.23.1      State          : Up
Out I/F     : 1/1/c1/2         Out Label      : 524286
Up Time     : 0d 00:02:27     Active Time    : n/a
Reserved BW : 0 Kbps          Protected LSP Count : 1
Type        : Dynamic         Bypass Path Cost : 10010
Setup Priority : 7             Hold Priority   : 0
Class Type  : 0
Exclude Node : None           Inter-Area     : False
Computed Hops :
  192.168.34.2(S)           Egress Admin Groups : None
  -> 192.168.34.1(S)        Egress Admin Groups : None
  -> 192.168.23.1(S)        Egress Admin Groups : None
Actual Hops :
  192.168.34.2(192.0.2.4)   Record Label     : N/A
  -> 192.168.34.1(192.0.2.3) Record Label     : 524286
  -> 192.168.23.1(192.0.2.2) Record Label     : 524284
Last Resignal :
Attempted At : n/a           Resignal Reason  : n/a
Resignal Status: n/a        Reason           : n/a

Protected LSPs -
LSP Name     : LSP-PE-4-PE-2::dyn
From         : 192.0.2.4      To               : 192.0.2.2
Avoid Node/Hop : 192.0.2.1   Downstream Label : 524286
Bandwidth    : 0 Kbps

=====
```

This bypass tunnel, via PE-3 to PE-2, offers node protection for node PE-1. There are no unnumbered interfaces in this path. In a similar way, information about the bypass tunnel to protect the link between PE-1 and PE-2 can be retrieved in PE-1, as follows:

```
[/]
A:admin@PE-1# show router mpls bypass-tunnel protected-lsp detail

=====
MPLS Bypass Tunnels (Detail)
=====
-----
bypass-link192.0.2.2-61441
-----
To          : 192.168.23.1      State          : Up
Out I/F     : 1/1/c1/3         Out Label      : 524287
Up Time     : 0d 00:02:24     Active Time    : n/a
Reserved BW : 0 Kbps          Protected LSP Count : 1
Type        : Dynamic         Bypass Path Cost : 10010
Setup Priority : 7             Hold Priority   : 0
Class Type  : 0
Exclude Node : None           Inter-Area     : False
Computed Hops :
  192.168.13.1(S)          Egress Admin Groups : None

=====
```

```

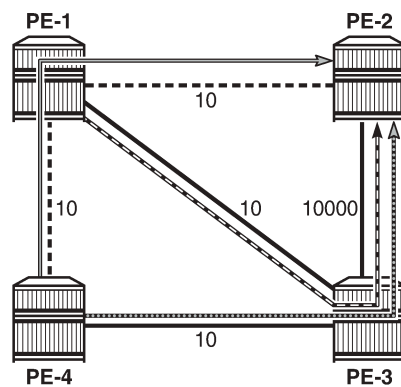
-> 192.168.13.2(S)          Egress Admin Groups : None
-> 192.168.23.1(S)         Egress Admin Groups : None
Actual Hops      :
  192.168.13.1(192.0.2.1)  Record Label       : N/A
-> 192.168.13.2(192.0.2.3) Record Label       : 524287
-> 192.168.23.1(192.0.2.2) Record Label       : 524285
Last Resignal   :
Attempted At    : n/a      Resignal Reason      : n/a
Resignal Status: n/a      Reason              : n/a

Protected LSPs -
LSP Name       : LSP-PE-4-PE-2::dyn
From           : 192.0.2.4      To               : 192.0.2.2
Avoid Node/Hop : 192.0.2.2      Downstream Label : 524286
Bandwidth      : 0 Kbps
=====

```

Figure 152: LSP and FRR facility bypass tunnels shows the LSP and the two bypass tunnels: one in PE-4, offering node protection for node PE-1, and another in PE-1, bypassing the link between PE-1 and PE-2.

Figure 152: LSP and FRR facility bypass tunnels



**Legend:**

- LSP-PE-4-PE-2 Primary path
- Bypass node PE-1
- Bypass link between PE-1 and PE-2

25686

For each bypass tunnel, an additional RSVP session is set up. The following output shows that, in PE-4, two LSPs are signaled: the primary LSP and the bypass tunnel for node PE-1.

```

[/]
A:admin@PE-4# show router rsvp session

=====
RSVP Sessions
=====
RSVP Session Name
From           To           Tunnel ID   LSP ID      State
-----
LSP-PE-4-PE-2::dyn
192.0.2.4     192.0.2.2   1           12296       Up
bypass-node192.0.2.1-61441

```

```

192.0.2.4          192.168.23.1      61441      2          Up
-----
Sessions : 2
=====

```

Similarly, PE-1 has an RSVP session for the primary LSP, but also for the bypass tunnel for the link toward PE-2, as follows:

```

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

=====
RSVP Sessions
=====
RSVP Session Name
  From          To          Tunnel ID  LSP ID     State
-----
LSP-PE-4-PE-2::dyn
192.0.2.4      192.0.2.2      1          12296      Up
bypass-link192.0.2.2-61441
192.0.2.1      192.168.23.1  61441      2          Up
-----
Sessions : 2
=====

```

PE-3 is only used by the bypass tunnels, as follows:

```

[/]
A:admin@PE-3# show router rsvp session

=====
RSVP Sessions
=====
RSVP Session Name
  From          To          Tunnel ID  LSP ID     State
-----
bypass-link192.0.2.2-61441
192.0.2.1      192.168.23.1  61441      2          Up
bypass-node192.0.2.1-61441
192.0.2.4      192.168.23.1  61441      2          Up
-----
Sessions : 2
=====

```

PE-2 terminates the LSP and the bypass tunnels, as follows:

```

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

=====
RSVP Sessions
=====
RSVP Session Name
  From          To          Tunnel ID  LSP ID     State
-----
LSP-PE-4-PE-2::dyn
192.0.2.4      192.0.2.2      1          12296      Up

```

```
bypass-link192.0.2.2-61441
192.0.2.1      192.168.23.1      61441      2      Up

bypass-node192.0.2.1-61441
192.0.2.4      192.168.23.1      61441      2      Up

-----
Sessions : 3
=====
```

## FRR one-to-one only supported on numbered interfaces

When FRR one-to-one is enabled on the LSP, the LSP does not use unnumbered interfaces. FRR is reconfigured on the LSP as follows:

```
# on PE-4: configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-4-PE-2" {
        fast-reroute {
          frr-method one-to-one
        }
      }
    }
  }
}
```

The LSP only comes up if it can use numbered interfaces end-to-end. In this case, the LSP takes the path via PE-3 with CSPF metric 10010, as follows:

```
[/]
A:admin@PE-4# show router mpls lsp "LSP-PE-4-PE-2" path detail

=====
MPLS LSP LSP-PE-4-PE-2 Path (Detail)
=====
Legend :
  @ - Detour Available          # - Detour In Use
  b - Bandwidth Protected      n - Node Protected
  s - Soft Preemption
  S - Strict                    L - Loose
  A - ABR                       + - Inherited
=====

LSP LSP-PE-4-PE-2
Path dyn
-----
LSP Name      : LSP-PE-4-PE-2
From          : 192.0.2.4
To            : 192.0.2.2
Admin State   : Up                Oper State    : Up
Path Name     : dyn
Path LSP ID   : 12300              Path Type     : Primary
Path Admin    : Up                Path Oper     : Up
Out Interface : 1/1/c1/2           Out Label     : 524287
---snip---
Explicit Hops :
  No Hops Specified
Actual Hops   :
  192.168.34.2(192.0.2.4)         Record Label  : N/A
  -> 192.168.34.1(192.0.2.3)      Record Label  : 524287
  -> 192.168.23.1(192.0.2.2)      Record Label  : 524287
Computed Hops :
```

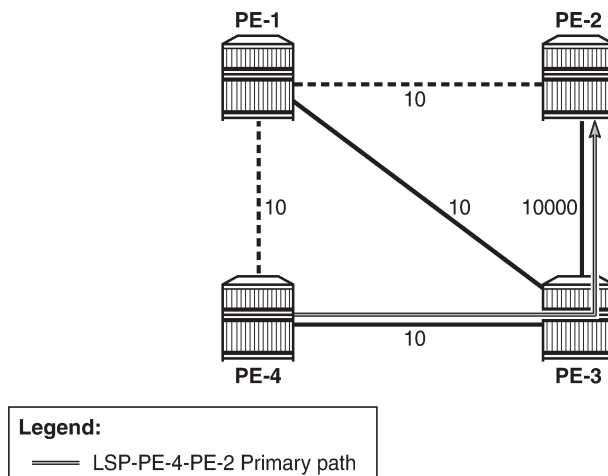
```

192.168.34.2(S)
-> 192.168.34.1(S)
-> 192.168.23.1(S)
Resignal Eligible: False
Last Resignal      : n/a
CSPF Metric        : 10010
=====
* indicates that the corresponding row element may have been truncated.

```

Figure 153: FRR one-to-one only supported on numbered interfaces shows the LSP in case of FRR one-to-one. Only numbered interfaces are used. Unfortunately, there is no bypass tunnel possible with only numbered interfaces; therefore, there is no protection.

Figure 153: FRR one-to-one only supported on numbered interfaces



25687

If there is no path available with only numbered interfaces, the LSP remains operationally down with failure code **noCspfRouteToDestination**". This can be verified by disabling port 1/1/c1/2 toward PE-3, as follows:

```

# on PE-4: configure {
  port 1/1/c1/2 {
    admin-state disable
  }
exit all

```

```

[/]
A:admin@PE-4# show router mpls lsp "LSP-PE-4-PE-2" path detail

```

```

=====
MPLS LSP LSP-PE-4-PE-2 Path (Detail)
=====

```

```

---snip---

```

```

LSP LSP-PE-4-PE-2
Path dyn

```

```

-----
LSP Name      : LSP-PE-4-PE-2
From          : 192.0.2.4
To            : 192.0.2.2
Admin State   : Up
Path Name     : dyn
Path LSP ID   : 12302
Path Admin    : Up
Oper State    : Down
Path Type     : Primary
Path Oper     : Down

```



```

Out Interface      : n/a                Out Label        : n/a
---snip---
Failure Code     : noCspfRouteToDestination
Failure Node      : 192.0.2.4
Explicit Hops     :
  No Hops Specified
Actual Hops       :
  No Hops Specified
Computed Hops     :
  No Hops Specified
Resignal Eligible: False
Last Resignal     : n/a                CSPF Metric      : N/A
=====
* indicates that the corresponding row element may have been truncated.

```

The port is enabled again and the LSP configuration is restored to FRR facility, as follows:

```

# on PE-4: configure {
  port 1/1/c1/2 {
    admin-state enable
  }
  exit all

  configure {
    router "Base" {
      mpls {
        lsp "LSP-PE-4-PE-2" {
          admin-state disable
          fast-reroute {
            frr-method facility
          }
        }
      }
    }
  }

  configure {
    router "Base" {
      mpls {
        lsp "LSP-PE-4-PE-2" {
          admin-state enable
        }
      }
    }
  }
}

```

### FRR bypass not possible on iLER on unnumbered interfaces

FRR facility is not supported on the iLER PE-4 if the bypass is over an unnumbered interface. This restriction only applies to the iLER, not to the LSRs. The interface toward PE-3 is reconfigured as unnumbered, as follows:

```

# on PE-3: configure {
  router "Base" {
    interface "int-PE-3-PE-4" {
      ipv4 {
        delete primary
        unnumbered {
          system
        }
      }
    }
  }
  mpls {
    interface "int-PE-3-PE-4" {
      te-metric 10
    }
  }
}

```

```

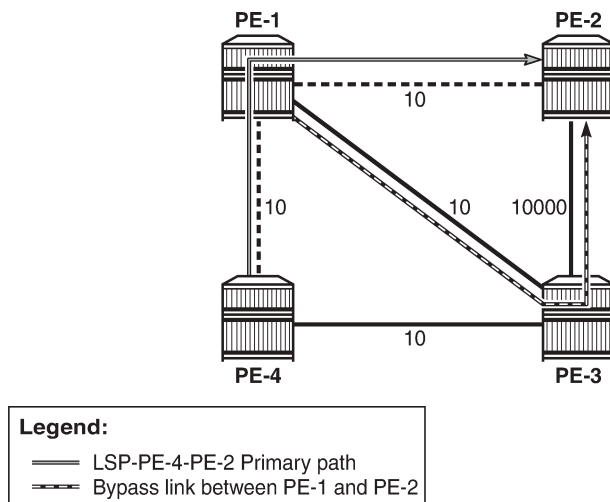
exit all

# on PE-4: configure {
  router "Base" {
    interface "int-PE-4-PE-3" {
      ipv4 {
        delete primary
        unnumbered {
          system
        }
      }
    }
  }
  mpls {
    interface "int-PE-4-PE-3" {
      te-metric 10
    }
  }
}
exit all

```

When an interface changes from numbered to unnumbered or the other way around, it is no longer known in the **mpls** context. Therefore, the interface must be added in the **mpls** context again. When the interface toward PE-3 is numbered, there is a bypass tunnel in PE-4 to protect node PE-1, but this bypass tunnel cannot be established on an unnumbered interface. The only remaining protection for the LSP is the bypass tunnel originating in PE-1 to protect the link between PE-1 and PE-2, as shown in [Figure 154: FRR on iLER: no bypass on unnumbered interfaces](#).

Figure 154: FRR on iLER: no bypass on unnumbered interfaces



25688

The following output shows that there is only a detour available in PE-1:

```

[/]
A:admin@PE-4# show router mpls lsp "LSP-PE-4-PE-2" path detail

=====
---snip--
-----
LSP LSP-PE-4-PE-2
Path dyn

```

```

-----
LSP Name      : LSP-PE-4-PE-2
From          : 192.0.2.4
To           : 192.0.2.2
Admin State   : Up                               Oper State    : Up
Path Name     : dyn
Path LSP ID   : 12308                             Path Type     : Primary
Path Admin    : Up                               Path Oper     : Up
Out Interface : 1/1/c1/1                           Out Label    : 524287
---snip---
Explicit Hops :
  No Hops Specified
Actual Hops   :
  192.0.2.4, If Index : 2                       Record Label  : N/A
  -> 192.0.2.1, If Index : 4 @                 Record Label  : 524287
  -> 192.0.2.2, If Index : 2                       Record Label  : 524286
Computed Hops :
  192.0.2.4, If Index : 2(S)
  -> 192.0.2.1, If Index : 4(S)
  -> 192.0.2.2, If Index : 2(S)
Resignal Eligible: False
Last Resignal  : n/a                             CSPF Metric   : 20
=====
* indicates that the corresponding row element may have been truncated.

```

In iLER PE-4, there is only the LSP tunnel, no bypass tunnel, as follows:

```

[/]
A:admin@PE-4# show router rsvp session

=====
RSVP Sessions
=====
RSVP Session Name
  From          To          Tunnel ID  LSP ID    State
-----
LSP-PE-4-PE-2::dyn
192.0.2.4      192.0.2.2      1          12308     Up
-----
Sessions : 1
=====

```

The original configuration is restored with numbered interfaces between PE-3 and PE-4, as follows:

```

# on PE-3: configure {
  router "Base" {
    interface "int-PE-3-PE-4" {
      ipv4
        delete unnumbered
        primary {
          address 192.168.34.1
          prefix-length 30
        }
      }
    }
  mpls {
    interface "int-PE-3-PE-4" {
      te-metric 10
    }
  }
}

```

```
exit all

# on PE-4: configure {
  router "Base" {
    interface "int-PE-4-PE-3" {
      ipv4
        delete unnumbered
        primary {
          address 192.168.34.2
          prefix-length 30
        }
      }
    }
  mpls {
    interface "int-PE-4-PE-3" {
      te-metric 10
    }
  }
}
exit all
```

## Admin groups for unnumbered interfaces in RSVP

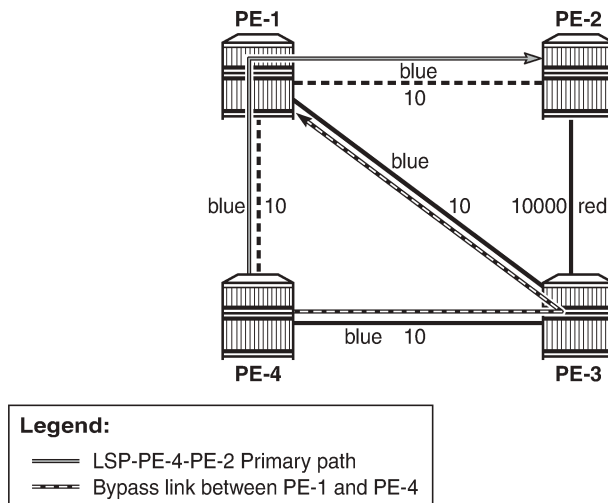
Administrative groups (link coloring) can be used to calculate a path with the restriction to only include, or exclude, links of a particular admin group (color). Paths can be disjointed from each other, without the need for an explicit hops list. For unnumbered interfaces, an explicit hops list is not an option, but admin groups are.

Two admin groups are configured on all nodes, as follows:

```
# on all nodes: configure {
  routing-options {
    if-attribute {
      admin-group "red" {
        value 0
      }
      admin-group "blue" {
        value 1
      }
    }
  }
}
exit all
```

Admin group "blue" is assigned to all MPLS interfaces, except for the link between PE-2 and PE-3; see [Figure 155: FRR facility and admin groups](#).

Figure 155: FRR facility and admin groups



25689

The admin groups are assigned to the interfaces in the **mpls** context, as follows:

```
# on PE-2: configure {
  router "Base" {
    mpls {
      resignal-timer 30
      interface "int-PE-2-PE-1" {
        admin-group ["blue"]
      }
      interface "int-PE-2-PE-3" {
        admin-group ["red"]
      }
    }
  }
  exit all
```

To ensure that FRR bypass tunnels only use links belonging to the same admin group, the following is configured on all nodes. It is required on all PLRs.

```
# on all nodes: configure {
  router "Base" {
    mpls {
      admin-group-frr true
    }
  }
  exit all
```

In the **lsp** context, the admin group "blue" is included. The option **propagate-admin-group** implies that the tunnels must use links belonging to the admin group "blue". This is configured for the LSP tunnel, and for the FRR bypass tunnels, as follows:

```
# on PE-4: configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-4-PE-2" {
        admin-state enable
        type p2p-rsvp
        to 192.0.2.2
        propagate-admin-group true
        path-computation-method local-cspf
      }
    }
  }
  exit all
```

```
metric-type te
  include-admin-group ["blue"]
  fast-reroute {
    propagate-admin-group true
    frr-method facility
  }
  primary "dyn" {
  }
}
exit all
```

This configuration implies that the link that does not belong to admin group "blue" is excluded, and cannot be used by the LSP nor by a bypass tunnel. Therefore, there is no bypass tunnel to protect node PE-1 and no bypass tunnel originating in PE-1 protecting the link to PE-2. There is a bypass tunnel originating in PE-4 to protect the link between PE-4 and PE-1, as shown in [Figure 155: FRR facility and admin groups](#). The following output shows that a detour is available for link protection in PE-4:

```
[/]
A:admin@PE-4# show router mpls lsp "LSP-PE-4-PE-2" path detail

=====
MPLS LSP LSP-PE-4-PE-2 Path (Detail)
=====
Legend :
  @ - Detour Available          # - Detour In Use
  b - Bandwidth Protected      n - Node Protected
  s - Soft Preemption
  S - Strict                    L - Loose
  A - ABR                       + - Inherited

-----
LSP LSP-PE-4-PE-2
Path dyn
-----
LSP Name      : LSP-PE-4-PE-2
From          : 192.0.2.4
To            : 192.0.2.2
Admin State   : Up                Oper State    : Up
Path Name     : dyn
Path LSP ID   : 12310             Path Type     : Primary
Path Admin    : Up                Path Oper     : Up
Out Interface : 1/1/c1/1          Out Label     : 524286
---snip---
Include Groups      :           Oper IncludeGroups:
blue                :           blue
Exclude Groups      :           Oper ExcludeGroups:
None                 :           None
---snip---
Explicit Hops      :
  No Hops Specified
Actual Hops        :
  192.0.2.4, If Index : 2 @           Record Label   : N/A
  -> 192.0.2.1, If Index : 4           Record Label   : 524286
  -> 192.0.2.2, If Index : 2           Record Label   : 524285
Computed Hops      :
  192.0.2.4, If Index : 2(S)
  -> 192.0.2.1, If Index : 4(S)
  -> 192.0.2.2, If Index : 2(S)
---snip---
=====
* indicates that the corresponding row element may have been truncated.
```

The following output shows two RSVP sessions in PE-4: one for the LSP and one for the bypass tunnel to protect the link between PE-4 and PE-1.

```
[/]
A:admin@PE-4# show router rsvp session

=====
RSVP Sessions
=====
RSVP Session Name
  From              To              Tunnel ID   LSP ID   State
-----
LSP-PE-4-PE-2::dyn
192.0.2.4          192.0.2.2      1           12310    Up
bypass-link192.0.2.1-61457
192.0.2.4          192.168.13.1  61457       22       Up
-----
Sessions : 2
=====
```

The configuration is restored as follows:

```
# on PE-4: configure {
  router "Base" {
    mpls {
      lsp "LSP-PE-4-PE-2" {
        fast-reroute {
          delete propagate-admin-group
        }
        delete propagate-admin-group
        delete include-admin-group ["blue"]
      }
      delete admin-group-frr
    }
  }
  exit all
```

## SRLGs for unnumbered interfaces in RSVP

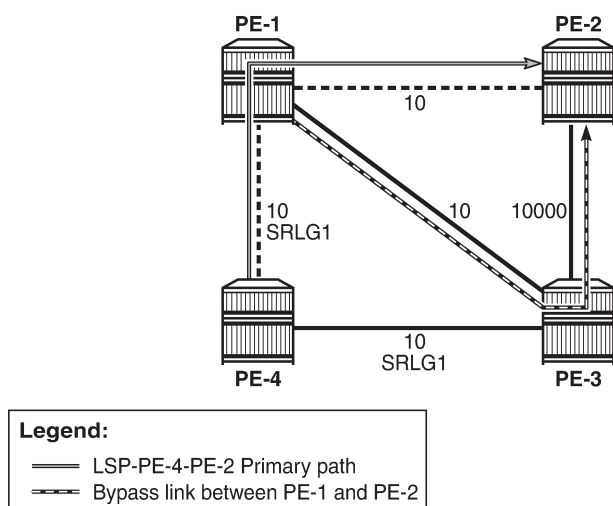
SRLGs allow operators to create automatic secondary LSPs or FRR tunnels that are disjointed from the protected primary tunnel. See chapter [Shared Risk Link Groups for RSVP-Based LSPs](#) for more information.

One SRLG group is configured on all nodes, as follows:

```
# on all nodes: configure {
  routing-options {
    if-attribute {
      srlg-group "SRLG1" {
        value 1
      }
    }
  }
  exit all
```

SRLG "SRLG1" is assigned to the interface between PE-4 and PE-1, and to the interface between PE-4 and PE-3, as shown in [Figure 156: SRLG-FRR strict: no bypass on PE-4](#).

Figure 156: SRLG-FRR strict: no bypass on PE-4



25690

The SRLG is assigned to the interfaces in the **mpls** context, as follows:

```
# on PE-4: configure {
  router "Base" {
    mpls {
      interface "int-PE-4-PE-1" {
        srlg-group "SRLG1" { }
      }
      interface "int-PE-4-PE-3" {
        srlg-group "SRLG1" { }
      }
    }
  }
  exit all
```

The configuration on PE-1 and PE-3 is similar.

When SRLG for FRR is enabled in strict mode, CSPF does not establish any detour LSP if there is no path that meets the SRLG constraint. This configuration implies that there is no bypass tunnel in PE-4. The following enables SRLG for FRR in strict mode on all nodes:

```
# on all nodes: configure {
  router "Base" {
    mpls {
      srlg-frr strict
    }
  }
  exit all
```



**Note:**

Enabling or disabling SRLG for FRR is a system-wide configuration that requires the MPLS routing instance to be manually disabled and then re-enabled, to activate the change. This can be service affecting. Nokia recommends that the operator include the SRLG in the initial network design and implementation to minimize the traffic loss.

The following output shows that there is only a detour available in PE-1:

```
[/]
A:admin@PE-4# show router mpls lsp "LSP-PE-4-PE-2" path detail
```



```

=====
MPLS LSP LSP-PE-4-PE-2 Path (Detail)
=====
Legend :
  @ - Detour Available          # - Detour In Use
  b - Bandwidth Protected       n - Node Protected
  s - Soft Preemption
  S - Strict                    L - Loose
  A - ABR                      + - Inherited
=====
-----
LSP LSP-PE-4-PE-2
Path dyn
-----
LSP Name      : LSP-PE-4-PE-2
From          : 192.0.2.4
To            : 192.0.2.2
Admin State   : Up                Oper State      : Up
Path Name     : dyn
Path LSP ID   : 12314             Path Type       : Primary
Path Admin    : Up                Path Oper       : Up
Out Interface : 1/1/c1/1          Out Label       : 524287
---snip---
Explicit Hops :
  No Hops Specified
Actual Hops    :
  192.0.2.4, If Index : 2          Record Label    : N/A
-> 192.0.2.1, If Index : 4 @      Record Label    : 524287
-> 192.0.2.2, If Index : 2          Record Label    : 524287
Computed Hops  :
  192.0.2.4, If Index : 2(S)
-> 192.0.2.1, If Index : 4(S)
-> 192.0.2.2, If Index : 2(S)
---snip---
=====
* indicates that the corresponding row element may have been truncated.

```

The following output shows that PE-1 has two RSVP sessions: one for the LSP and one for the bypass tunnel to protect the link between PE-1 and PE-2.

```

[/]
A:admin@PE-1# show router rsvp session
=====
RSVP Sessions
=====
RSVP Session Name
  From          To          Tunnel ID  LSP ID    State
-----
LSP-PE-4-PE-2::dyn
192.0.2.4      192.0.2.2      1          12314     Up
bypass-link192.0.2.2-61496
192.0.2.1      192.168.23.1   61496     10        Up
-----
Sessions : 2
=====

```

This was the last example for unnumbered interfaces in RSVP. MPLS and RSVP are disabled in all nodes as follows:

```
# on all nodes: configure {
  router "Base" {
    rsvp {
      admin-state disable
    }
    mpls {
      admin-state disable
      delete srlg-frr
    }
  }
  exit all
```

## Unnumbered interfaces in LDP

Link LDP is configured on PE-4, as follows:

```
# on PE-4: configure {
  router "Base" {
    ldp {
      interface-parameters {
        interface "int-PE-4-PE-1" {
          ipv4 {
          }
        }
        interface "int-PE-4-PE-3" {
          ipv4 {
          }
        }
      }
    }
  }
  exit all
```

The configuration of link LDP on the other nodes is similar. Link LDP sessions are established, which can be verified as follows:

```
[/]
A:admin@PE-4# show router ldp session family ipv4

=====
LDP IPv4 Sessions
=====
Peer LDP Id      Adj Type  State      Msg Sent  Msg Recv  Up Time
-----
192.0.2.1:0     Link     Established  25        26        0d 00:00:41
192.0.2.3:0     Link     Established  79        80        0d 00:02:49
-----
No. of IPv4 Sessions: 2
=====
```

The **Peer LDP Id** is the LSR ID, which is the system address, by default. The IP address configured on the unnumbered interface (such as 192.1.2.1) is not used. The following tunnel table shows a distinction between numbered and unnumbered interfaces in the next hop:

```
[/]
A:admin@PE-4# show router tunnel-table

=====
IPv4 Tunnel Table (Router: Base)
```

```

=====
Destination      Owner      Encap TunnelId Pref  Nexthop      Metric
  Color
-----
192.0.2.1/32     ldp        MPLS  65540   9    int-PE-4-PE-1 10
192.0.2.2/32     ldp        MPLS  65541   9    int-PE-4-PE-1 20
192.0.2.3/32     ldp        MPLS  65537   9    192.168.34.1  10
-----
Flags: B = BGP or MPLS backup hop available
      L = Loop-Free Alternate (LFA) hop available
      E = Inactive best-external BGP route
      k = RIB-API or Forwarding Policy backup hop
=====

```

For destination 192.0.2.1 or 192.0.2.2, the unnumbered interface toward PE-1 is taken. The next hop is represented by int-PE-4-PE-1. When a node has several unnumbered interfaces, the corresponding next hop values are different, as follows, for PE-1:

```

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

=====
IPv4 Tunnel Table (Router: Base)
=====
Destination      Owner      Encap TunnelId Pref  Nexthop      Metric
  Color
-----
192.0.2.2/32     ldp        MPLS  65541   9    int-PE-1-PE-2 10
192.0.2.3/32     ldp        MPLS  65542   9    192.168.13.2  10
192.0.2.4/32     ldp        MPLS  65540   9    int-PE-1-PE-4 10
-----
Flags: B = BGP or MPLS backup hop available
      L = Loop-Free Alternate (LFA) hop available
      E = Inactive best-external BGP route
      k = RIB-API or Forwarding Policy backup hop
=====

```

The LDP active prefix bindings only contain system addresses, no other loopback prefixes, as follows:

```

[/]
A:admin@PE-4# show router ldp bindings active prefixes ipv4

=====
LDP Bindings (IPv4 LSR ID 192.0.2.4)
(IPv6 LSR ID ::)
=====
---snip---
=====
LDP IPv4 Prefix Bindings (Active)
=====
Prefix           Op
IngLbl           EgrLbl
EgrNextHop       EgrIf/LspId
-----
192.0.2.1/32     Push
--              524283
Unnumbered       1/1/c1/1

192.0.2.1/32     Swap
524283           524283
Unnumbered       1/1/c1/1

```

```

192.0.2.2/32          Push
--                  524280
Unnumbered           1/1/c1/1

192.0.2.2/32          Swap
524282              524280
Unnumbered           1/1/c1/1

192.0.2.3/32          Push
--                  524287
192.168.34.1         1/1/c1/2

192.0.2.3/32          Swap
524286              524287
192.168.34.1         1/1/c1/2

192.0.2.4/32          Pop
524287              --
--                  --

-----
No. of IPv4 Prefix Active Bindings: 7
=====

```

For prefixes 192.0.2.1 and 192.0.2.2, the egress next hop is unnumbered. The egress interface for both is 1/1/c1/1. There is no If Index. Local addresses are advertised in LDP address messages, such as:

```

# on PE-4:
8 2023/03/21 13:24:43.317 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Send Address packet (msgId 5) to 192.0.2.1:0
Protocol version = 1
Address Family = 1 Number of addresses = 3
Address 1 = 192.0.2.4
Address 2 = 192.1.2.4
Address 3 = 192.168.34.2
"

9 2023/03/21 13:24:43.378 UTC MINOR: DEBUG #2001 Base LDP
"LDP: LDP
Recv Address packet (msgId 5) from 192.0.2.1:0
Protocol version = 1
Address Family = 1 Number of addresses = 3
Address 1 = 192.0.2.1
Address 2 = 192.1.2.1
Address 3 = 192.168.13.1
"

```

The list of advertised local addresses includes the loopback addresses "loopback1": 192.1.2.1 and 192.1.2.4. These loopback addresses did not occur in the preceding lists of LDP sessions or LDP bindings, but they occur in the LDP session local addresses, as follows:

```

[/]
A:admin@PE-4# show router ldp session local-addresses ipv4

=====
LDP Session Local-Addresses
=====
-----
Session with Peer 192.0.2.1:0,
      Local 192.0.2.4:0
-----

```

```
IPv4 Sent Addresses:
                192.0.2.4      192.1.2.4      192.168.34.2

IPv4 Recv Addresses:
                192.0.2.1      192.1.2.1      192.168.13.1
-----
---snip---
```

If there were only unnumbered addresses and no additional loopback addresses, only the system address and other loopback addresses would be sent or received. The interface addresses in the list of local addresses are from numbered interfaces.

## Configuring the local LSR ID

To use the loopback address "loopback1" in the LDP sessions, the local LSR ID is configured as follows:

```
# on PE-4: configure {
  router "Base" {
    ldp {
      interface-parameters {
        interface "int-PE-4-PE-1" {
          ipv4 {
            local-lsr-id {
              interface-name "loopback1"
            }
          }
          transport-address interface
        }
      }
    }
  }
  exit all
```

The transport address is the system address, by default, but here it is changed to the address of "loopback1", which is 192.1.2.4. The configuration is similar on PE-1. On PE-2, the system addresses are kept and no additional configuration is required.

LDP hello messages are sent from the transport address to 224.0.0.2 to establish hello adjacencies. The transport address is 192.1.2.4 for the unnumbered interface toward PE-1, and 192.0.2.4 (system address of PE-4) for the numbered interface toward PE-3. LDP hello adjacencies are verified as follows:

```
[/]
A:admin@PE-4# show router ldp discovery ipv4

=====
LDP IPv4 Hello Adjacencies
=====
Interface Name          Local Addr          State
AdjType                 Peer Addr
-----
int-PE-4-PE-1          192.1.2.4:0        Estab
link                    192.1.2.1:0
int-PE-4-PE-3          192.0.2.4:0        Estab
link                    192.0.2.3:0

-----
No. of IPv4 Hello Adjacencies: 2
=====
```

The LDP hello adjacencies are established, but the LDP session on the unnumbered interface is non-existent, as follows:

```
[/]
A:admin@PE-4# show router ldp session family ipv4

=====
LDP IPv4 Sessions
=====
Peer LDP Id      Adj Type  State      Msg Sent  Msg Recv  Up Time
-----
192.0.2.3:0      Link      Established 119       120       0d 00:04:32
192.1.2.1:0     Link     Nonexistent 10        11        0d 00:00:36
-----
No. of IPv4 Sessions: 2
=====
```

The LDP session is non-existent because the prefix 192.1.2.1/32 is not in the routing table and the LDP session is to be established between 192.1.2.4 and 192.1.2.1. The following export policy is configured and added in the IS-IS context on the nodes:

```
# on PE-4: configure {
  policy-options {
    policy-statement "export_ISIS" {
      entry 10 {
        from {
          protocol {
            name [direct]
          }
        }
        action {
          action-type accept
        }
      }
      default-action {
        action-type reject
      }
    }
  }
  router "Base" {
    isis 0 {
      export-policy ["export_ISIS"]
    }
  }
}
exit all
```

The loopback addresses are now exported in IS-IS. When the loopback addresses are in the routing table, the LDP session is established, as follows:

```
[/]
A:admin@PE-4# show router ldp session family ipv4

=====
LDP IPv4 Sessions
=====
Peer LDP Id      Adj Type  State      Msg Sent  Msg Recv  Up Time
-----
192.0.2.3:0      Link      Established 130       128       0d 00:04:57
192.1.2.1:0     Link     Established 29        30        0d 00:01:01
-----
No. of IPv4 Sessions: 2
=====
```

The local LSR ID can also be configured for targeted LDP sessions, as follows:

```
# on PE-4: configure {
  router "Base" {
    ldp {
      targeted-session {
        peer 192.1.2.1 {
          local-lsr-id {
            interface-name "loopback1"
          }
        }
      }
    }
  }
}
```

The configuration on PE-1 is similar for peer 192.0.2.4. On PE-4, the LDP adjacency type is now both link and targeted for peer 192.1.2.1, as follows:

```
[/]
A:admin@PE-4# show router ldp session family ipv4

=====
LDP IPv4 Sessions
=====
Peer LDP Id      Adj Type  State      Msg Sent  Msg Recv  Up Time
-----
192.0.2.3:0     Link     Established 144       144       0d 00:05:36
192.1.2.1:0     Both     Established 45        52        0d 00:01:40
-----
No. of IPv4 Sessions: 2
=====
```

Even though the LDP sessions are established, there is no LDP prefix binding for the loopback address, as follows:

```
[/]
A:admin@PE-4# show router ldp bindings active prefixes ipv4

=====
LDP Bindings (IPv4 LSR ID 192.0.2.4)
(IPv6 LSR ID ::)
=====
---snip---
=====
LDP IPv4 Prefix Bindings (Active)
=====
Prefix          Op
IngLbl          EgrLbl
EgrNextHop      EgrIf/LspId
-----
192.0.2.1/32    Push
--              524283
Unnumbered      1/1/c1/1

192.0.2.1/32    Swap
524283          524283
Unnumbered      1/1/c1/1

192.0.2.2/32    Push
--              524280
Unnumbered      1/1/c1/1

192.0.2.2/32    Swap
```

```

524282                               524280
Unnumbered                           1/1/c1/1

192.0.2.3/32                          Push
--                                   524287
192.168.34.1                          1/1/c1/2

192.0.2.3/32                          Swap
524286                                524287
192.168.34.1                          1/1/c1/2

192.0.2.4/32                          Pop
524287                                --
--                                     --

-----
No. of IPv4 Prefix Active Bindings: 7
=====

```

There is no label mapping for prefix 192.1.2.1/32.

SDPs are created toward all other nodes, as follows:

```

# on PE-4: configure {
  service {
    sdp 411 {
      admin-state enable
      delivery-type mpls
      ldp true
      far-end {
        ip-address 192.0.2.1
      }
    }
    sdp 412 {
      admin-state enable
      delivery-type mpls
      ldp true
      far-end {
        ip-address 192.1.2.1
      }
    }
    sdp 421 {
      admin-state enable
      delivery-type mpls
      ldp true
      far-end {
        ip-address 192.0.2.2
      }
    }
    sdp 431 {
      admin-state enable
      delivery-type mpls
      ldp true
      far-end {
        ip-address 192.0.2.3
      }
    }
  }
exit all

```

The following output shows that, between PE-4 and PE-1, there are two LDP SDPs: one using the system address and another using the loopback address 192.1.2.1. The configuration on the other nodes is



similar. All SDPs that have a system address as the far end are operationally up, whereas the SDP toward 192.1.2.1 is down:

```
[/]
A:admin@PE-4# show service sdp

=====
Services: Service Destination Points
=====
SdpId  AdmMTU  OprMTU  Far End      Adm  Opr      Del  LSP  Sig
-----
411    0        1556    192.0.2.1    Up  Up       MPLS L    TLDP
412    0        0        192.1.2.1    Up  Down     MPLS L    TLDP
421    0        1556    192.0.2.2    Up  Up       MPLS L    TLDP
431    0        1556    192.0.2.3    Up  Up       MPLS L    TLDP
-----
Number of SDPs : 4
-----
Legend: R = RSVP, L = LDP, B = BGP, M = MPLS-TP, n/a = Not Applicable
        I = SR-ISIS, 0 = SR-OSPF, T = SR-TE, F = FPE
=====
```

The SDP toward 192.1.2.1 is down because the transport tunnel is down, as follows:

```
[/]
A:admin@PE-4# show service sdp 412 detail

=====
Service Destination Point (Sdp Id : 412) Details
=====
Sdp Id 412 -192.1.2.1
-----
Description      : (Not Specified)
SDP Id           : 412                SDP Source       : manual
Admin Path MTU   : 0                  Oper Path MTU    : 0
Delivery         : MPLS
Far End          : 192.1.2.1          Tunnel Far End   :
Oper Tunnel Far End : 192.1.2.1
LSP Types        : LDP

Admin State      : Up                  Oper State       : Down
Signaling        : TLDP                Metric           : 0
---snip---
Flags           : TranspTunnDown
---snip---
=====
```

The solution is to manually add an LDP prefix binding, as described in the following section.

## Configuring the FEC originate

The labels to be used for a manually created LDP prefix binding must be chosen from the range for static labels: from 32 to 18431. This range can be retrieved as follows:

```
[/]
A:admin@PE-1# show router mpls-labels label-range

=====
```

Label Ranges					
Label Type	Start Label	End Label	Aging	Available	Total
<b>Static</b>	<b>32</b>	<b>18431</b>	-	<b>18400</b>	<b>18400</b>
Dynamic	18432	524287	0	505852	505856
Seg-Route	0	0	-	0	0

To manually add an LDP prefix binding for the loopback prefixes, configure the following:

```
# on PE-4: configure {
  router "Base" {
    ldp {
      fec-originate 192.1.2.1/32 {
        next-hop 192.0.2.4
        interface "int-PE-4-PE-1"
        swap-label 101
      }
      fec-originate 192.1.2.4/32 {
        advertised-label 104
        pop true
      }
    }
  }
  exit all
```

```
# on PE-1: configure {
  router "Base" {
    ldp {
      fec-originate 192.1.2.1/32 {
        advertised-label 101
        pop true
      }
      fec-originate 192.1.2.4/32 {
        next-hop 192.0.2.1
        interface "int-PE-1-PE-4"
        swap-label 104
      }
    }
  }
  exit all
```

This configuration for unnumbered interfaces includes the interface name, such as int-PE-4-PE-1. This parameter is optional for numbered interfaces.

If the label is chosen from the dynamic range instead of the static range, an error is raised for the pop operation, as follows:

```
# on PE-1: configure {
  router "Base" {
    ldp {
      fec-originate 192.1.2.1/32 {
        pop true
        advertised-label 100001
      }
    }
  }
}
MINOR: MGMT_CORE #4001: configure router "Base" ldp fec-originate 192.1.2.1/32 advertised-label
- advertised-label 100001 is not within the valid range for static labels
- configure router "Base" mpls-labels static-label-range
```

For interoperability, no error is raised for the swap operation.

As a result, three active LDP bindings are added: one pop operation for the local loopback prefix, and a swap and a push operation for the remote loopback prefix, as follows:

```
[/]
```

```
A:admin@PE-4# show router ldp bindings active prefixes ipv4
```

```
=====
LDP Bindings (IPv4 LSR ID 192.0.2.4)
(IPv6 LSR ID ::)
=====
```

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

```
=====
LDP IPv4 Prefix Bindings (Active)
=====
```

Prefix	Op
IngLbl	EgrLbl
EgrNextHop	EgrIf/LspId
192.0.2.1/32	Push
--	524283
Unnumbered	1/1/c1/1
192.0.2.1/32	Swap
524283	524283
Unnumbered	1/1/c1/1
192.0.2.2/32	Push
--	524280
Unnumbered	1/1/c1/1
192.0.2.2/32	Swap
524282	524280
Unnumbered	1/1/c1/1
192.0.2.3/32	Push
--	524287
192.168.34.1	1/1/c1/2
192.0.2.3/32	Swap
524286	524287
192.168.34.1	1/1/c1/2
192.0.2.4/32	Pop
524287	--
--	--
<b>192.1.2.1/32</b>	<b>Push</b>
--	<b>101</b>
<b>Unnumbered</b>	<b>1/1/c1/1</b>
<b>192.1.2.1/32</b>	<b>Swap</b>
<b>524285</b>	<b>101</b>
<b>Unnumbered</b>	<b>1/1/c1/1</b>
<b>192.1.2.4/32(S)</b>	<b>Pop</b>
<b>104</b>	--
--	--

```
-----
No. of IPv4 Prefix Active Bindings: 10
=====
```

The following output shows that the SDPs are all operationally up, including the one toward the loopback address:

```
[/]
A:admin@PE-4# show service sdp
```

```

=====
Services: Service Destination Points
=====
SdpId  AdmMTU  OprMTU  Far End          Adm  Opr          Del  LSP  Sig
-----
411    0        1556    192.0.2.1        Up  Up           MPLS L    TLDP
412    0        1556    192.1.2.1        Up  Up           MPLS L    TLDP
421    0        1556    192.0.2.2        Up  Up           MPLS L    TLDP
431    0        1556    192.0.2.3        Up  Up           MPLS L    TLDP
-----
Number of SDPs : 4
-----
Legend: R = RSVP, L = LDP, B = BGP, M = MPLS-TP, n/a = Not Applicable
        I = SR-ISIS, 0 = SR-OSPF, T = SR-TE, F = FPE
=====

```

### LDP FRR Loop-Free Alternate on unnumbered interfaces

LDP FRR Loop-Free Alternate (LFA) is supported on unnumbered interfaces and on numbered interfaces. For information about LDP FRR LFA, see chapter [MPLS LDP FRR using ISIS as IGP](#). LDP FRR LFA can be configured as follows:

```

# on all nodes: configure {
  router "Base" {
    isis 0 {
      loopfree-alternate {
      }
    }
    ldp {
      fast-reroute {
      }
    }
  }
  routing-options {
    ip-fast-reroute true
  }
}
exit all

```

Enabling FRR LFA is a local decision. In this configuration, it is configured on all nodes. The LFA coverage can be retrieved as follows:

```

[/]
A:admin@PE-4# show router isis lfa-coverage

=====
Rtr Base ISIS Instance 0 LFA Coverage
=====
Topology          Level  Node          IPv4          IPv6
-----
IPv4 Unicast    L1     3/3(100%)    7/7(100%)    0/0(0%)
IPv6 Unicast      L1     0/0(0%)      0/0(0%)      0/0(0%)
IPv4 Multicast    L1     0/0(0%)      0/0(0%)      0/0(0%)
IPv6 Multicast    L1     0/0(0%)      0/0(0%)      0/0(0%)
IPv4 Unicast      L2     0/0(0%)      7/7(100%)    0/0(0%)
IPv6 Unicast      L2     0/0(0%)      0/0(0%)      0/0(0%)
IPv4 Multicast    L2     0/0(0%)      0/0(0%)      0/0(0%)
IPv6 Multicast    L2     0/0(0%)      0/0(0%)      0/0(0%)
=====

```

There is protection for the three other nodes and for all remote prefixes in the routing table, which can be verified as follows:

```
[/]
A:admin@PE-4# show router route-table alternative

=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]
Next Hop[Interface Name]
Alt-NextHop
Type Proto Age Metric Alt-Metric Pref
-----
192.0.2.1/32 Remote ISIS 00h43m13s 15
int-PE-4-PE-1 10
192.168.34.1 (LFA) 20
192.0.2.2/32 Remote ISIS 00h43m13s 15
int-PE-4-PE-1 20
192.168.34.1 (LFA) 30
192.0.2.3/32 Remote ISIS 00h22m23s 15
192.168.34.1 10
int-PE-4-PE-1 (LFA) 20
192.0.2.4/32 Local Local 01h08m36s 0
system 0
192.1.2.1/32 Remote ISIS 00h06m40s 15
int-PE-4-PE-1 10
192.168.34.1 (LFA) 20
192.1.2.2/32 Remote ISIS 00h06m39s 15
int-PE-4-PE-1 20
192.168.34.1 (LFA) 30
192.1.2.4/32 Local Local 00h43m14s 0
loopback1 0
192.168.13.0/30 Remote ISIS 00h06m40s 160
int-PE-4-PE-1 10
192.168.34.1 (LFA) 20
192.168.23.0/30 Remote ISIS 00h22m23s 15
192.168.34.1 10010
int-PE-4-PE-1 (LFA) 10020
192.168.34.0/30 Local Local 00h22m24s 0
int-PE-4-PE-3 0
-----
No. of Routes: 10
Flags: n = Number of times nexthop is repeated
Backup = BGP backup route
LFA = Loop-Free Alternate nexthop
S = Sticky ECMP requested
=====
```

For unnumbered interfaces, the interface name is shown (int-PE-4-PE-1); for numbered interfaces, the next hop IP address is shown (192.168.34.1). The LFA type is link protection for the three nodes, as follows:

```
[/]
A:admin@PE-4# show router isis topology lfa detail

=====
Rtr Base ISIS Instance 0 Topology Table
=====
IS-IS IP paths (MT-ID 0), Level 1
-----
Node : PE-1.00
```

```

Nexthop   : PE-1
Interface : int-PE-4-PE-1
SNPA      : none                               Metric    : 10

LFA nh    : PE-3
LFA intf  : int-PE-4-PE-3                     LFA Metric : 20
LFA type  : linkProtection

Node      : PE-2.00
Nexthop   : PE-1
Interface : int-PE-4-PE-1
SNPA      : none                               Metric    : 20

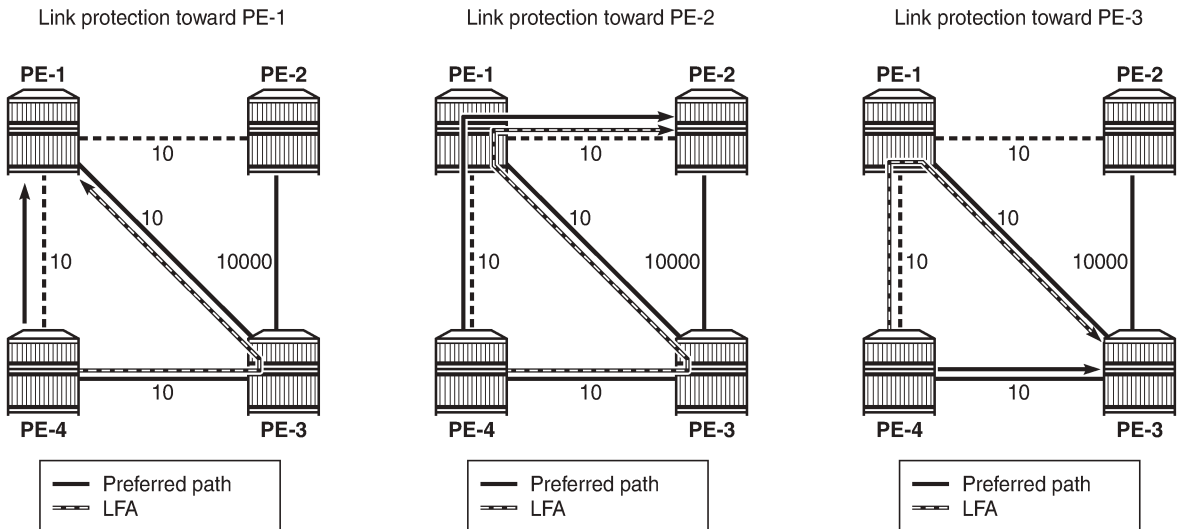
LFA nh    : PE-3
LFA intf  : int-PE-4-PE-3                     LFA Metric : 30
LFA type  : linkProtection

Node      : PE-3.00
Nexthop   : PE-3
Interface : int-PE-4-PE-3
SNPA      : none                               Metric    : 10

LFA nh    : PE-1
LFA intf  : int-PE-4-PE-1                     LFA Metric : 20
LFA type  : linkProtection
=====
    
```

The LFA protection is shown in [Figure 157: LDP FRR LFA link protection on PE-4](#).

Figure 157: LDP FRR LFA link protection on PE-4



25691

The LDP bindings for FRR LFA indicate alternate (**BU**) in the list, as follows:

```

[/]
A:admin@PE-4# show router ldp bindings prefixes ipv4
=====
    
```

```

LDP Bindings (IPv4 LSR ID 192.0.2.4)
              (IPv6 LSR ID ::)
=====
Label Status:
  U - Label In Use, N - Label Not In Use, W - Label Withdrawn
  WP - Label Withdraw Pending, BU - Alternate For Fast Re-Route
  e - Label ELC
FEC Flags:
  LF - Lower FEC, UF - Upper FEC, M - Community Mismatch,
  BA - ASBR Backup FEC
=====
LDP IPv4 Prefix Bindings
=====
Prefix
Peer          FEC-Flags
IgrLbl        EgrLbl
EgrNextHop    EgrIntf/LspId
-----
192.0.2.1/32
192.0.2.3:0
524283U          524283BU
192.168.34.1    1/1/c1/2

192.0.2.1/32
192.1.2.1:0
--              524283
Unnumbered      1/1/c1/1

192.0.2.2/32
192.0.2.3:0
524282U          524285BU
192.168.34.1    1/1/c1/2

192.0.2.2/32
192.1.2.1:0
524282N          524280
Unnumbered      1/1/c1/1

192.0.2.3/32
192.0.2.3:0
--              524287
192.168.34.1    1/1/c1/2

192.0.2.3/32
192.1.2.1:0
524286U          524281BU
Unnumbered      1/1/c1/1

192.0.2.4/32
192.0.2.3:0
524287U          --
--              --

192.0.2.4/32
192.1.2.1:0
524287U          --
--              --

192.1.2.1/32
192.0.2.3:0
524285U          524286BU
192.168.34.1    1/1/c1/2

192.1.2.1/32
    
```

```
192.1.2.1:0
--
Unnumbered
192.1.2.4/32
192.0.2.3:0
104U
--
192.1.2.4/32
192.1.2.1:0
104U
--
-----
No. of IPv4 Prefix Bindings: 12
=====
```

## Conclusion

Unnumbered interfaces were initially supported for SONET/SDH/ATM/FR, and later also on Ethernet access ports. IS-IS adjacencies and OSPF neighbors can be established on unnumbered interfaces. This chapter shows that unnumbered interfaces can be added to RSVP or LDP. Most features that are supported on numbered interfaces are also supported on unnumbered interfaces.





# Customer document and product support



## Customer documentation

[Customer documentation welcome page](#)



## Technical support

[Product support portal](#)



## Documentation feedback

[Customer documentation feedback](#)