7450 ETHERNET SERVICE SWITCH
7750 SERVICE ROUTER
7950 EXTENSIBLE ROUTING SYSTEM
VIRTUALIZED SERVICE ROUTER

MPLS GUIDE
RELEASE 21.2.R1

3HE 17154 AAAA TQZZA 01

Issue: 01

February 2021

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

Table of Contents

1 Getting Started

1.1 About This Guide

This guide describes the services and protocol support provided by the router and presents examples to configure and implement MPLS, RSVP, and LDP protocols.

This guide is organized into functional chapters and provides concepts and descriptions of the implementation flow, as well as Command Line Interface (CLI) syntax and command usage.

The topics and commands described in this document apply to the:

- 7450 ESS
- 7750 SR
- 7950 XRS
- VSR

Table 1 lists the available chassis types for each SR OS router.

Table 1 **Supported SR OS Router Chassis Types**

7450 ESS	7750 SR	7950 XRS
• 7450 ESS-7/12	• 7750 SR-a4/a8 • 7750 SR-1e/2e/3e • 7750 SR-12e • 7750 SR-1s/2s • 7750 SR-1 • 7750 SR-7/12 • 7750 SR-7s/14-s	• 7950 XRS-16c • 7950 XRS-20/40 • 7950 XRS-20e

For a list of unsupported features by platform and chassis, refer to the SR OS 21.*x*.R*x* Software Release Notes, part number 3HE 17177 000*x* TQZZA.

Command outputs shown in this guide are examples only; actual displays may differ depending on supported functionality and user configuration.

Note: The SR OS CLI trees and command descriptions have been removed from this guide and can now be found in the following guides:

- *7450 ESS, 7750 SR, 7950 XRS, and VSR Classic CLI Command Reference Guide*
- *7450 ESS, 7750 SR, 7950 XRS, and VSR Clear, Show, and Tools Command Reference Guide* (for both MD-CLI and Classic CLI)
- *7450 ESS, 7750 SR, 7950 XRS, and VSR MD-CLI Command Reference Guide*

Note: This guide generically covers Release 21.*x*.R*x* content and may contain some content that will be released in later maintenance loads. Please refer to the SR OS 21.*x*.R*x* Software Release Notes, part number 3HE 17177 000*x* TQZZA, for information about features supported in each load of the Release 21.*x*.R*x* software.

1.2 Nokia Router Configuration Process

Table 2 lists the tasks necessary to configure MPLS applications functions.

This guide is presented in an overall logical configuration flow. Each section describes a software area and provides CLI syntax and command usage to configure parameters for a functional area.

Table 2 **Configuration Process**

Area	Task	Section
MPLS and RSVP protocol configuration	MPLS Configuration	Common Configuration Tasks
	Configure RSVP parameters	Configuring RSVP Parameters
	MPLS configuration management	MPLS Configuration Management Tasks
	RSVP configuration management	RSVP Configuration Management Tasks
GMPLS protocol configuration	Configure LMP and IPCC	LMP and IPCC Configuration
	Configure MPLS paths for GMPLS	Configuring MPLS Paths for GMPLS
	Configure RSVP in GMPLS	Configuring RSVP in GMPLS
	Configure a GMPLS LSP on the UNI	Configuring a GMPLS LSP on the UNI
	Configure Bandwidth	Bandwidth
	Configure end-to-end GMPLS recovery	Configuration of End-to-End GMPLS Recovery
	Configure IP and MPLS in an overlay network to use a GMPLS LSP	Configuring IP and MPLS in an Overlay Network to Use a GMPLS LSP
PCEP configuration	Configure PCC and PCE	PCC and PCE Configuration
	Configure and Operate RSVP-TE LSP with PCEP	Configuring and Operating RSVP-TE LSP with PCEP
Label Distribution Protocol (LDP) configuration	Configure LDP	Configuring LDP with CLI
	LDP configuration management	LDP Configuration Management Tasks

2 MPLS and RSVP

2.1 MPLS

Multiprotocol Label Switching (MPLS) is a label switching technology that provides the ability to set up connection-oriented paths over a connectionless IP network. MPLS facilitates network traffic flow and provides a mechanism to engineer network traffic patterns independently from routing tables. MPLS sets up a specific path for a sequence of packets. The packets are identified by a label inserted into each packet. MPLS is not enabled by default and must be explicitly enabled.

MPLS is independent of any routing protocol but is considered multiprotocol because it works with the Internet Protocol (IP), Asynchronous Transport Mode (ATM), and frame relay network protocols.

2.1.1 MPLS Label Stack

MPLS requires a set of procedures to enhance network layer packets with label stacks which thereby turns them into labeled packets. Routers that support MPLS are known as Label Switching Routers (LSRs). In order to transmit a labeled packet on a particular data link, an LSR must support the encoding technique which, when given a label stack and a network layer packet, produces a labeled packet.

In MPLS, packets can carry not just one label, but a set of labels in a stack. An LSR can swap the label at the top of the stack, pop the stack, or swap the label and push one or more labels into the stack. The processing of a labeled packet is completely independent of the level of hierarchy. The processing is always based on the top label, without regard for the possibility that some number of other labels may have been above it in the past, or that some number of other labels may be below it at present.

As described in RFC 3032, *MPLS Label Stack Encoding*, the label stack is represented as a sequence of label stack entries. Each label stack entry is represented by 4 octets. Figure 1 displays the label placement in a packet.

Figure 1 **Label Placement**

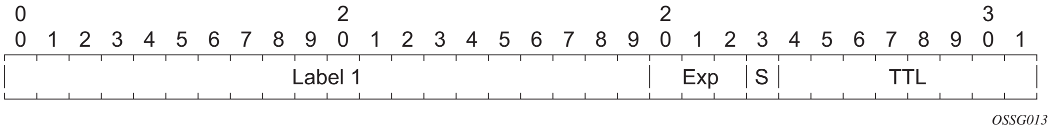
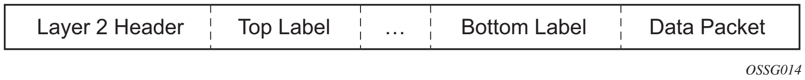


Table 3 **Packet/Label Field Description**

Field	Description
Label	This 20-bit field carries the actual value (unstructured) of the label.
Exp	This 3-bit field is reserved for experimental use. It is currently used for Class of Service (CoS).
S	This bit is set to 1 for the last entry (bottom) in the label stack, and 0 for all other label stack entries.
TTL	This 8-bit field is used to encode a TTL value.

A stack can carry several labels, organized in a last in/first out order. The top of the label stack appears first in the packet and the bottom of the stack appears last, as shown in Figure 2.

Figure 2 **Label Packet Placement**



The label value at the top of the stack is looked up when a labeled packet is received. A successful lookup reveals:

- The next hop where the packet is to be forwarded.
- The operation to be performed on the label stack before forwarding.

In addition, the lookup may reveal outgoing data link encapsulation and other information needed to properly forward the packet.

An empty label stack can be thought of as an unlabeled packet. An empty label stack has zero (0) depth. The label at the bottom of the stack is referred to as the Level 1 label. The label above it (if it exists) is the Level 2 label, and so on. The label at the top of the stack is referred to as the Level *m* label.

Labeled packet processing is independent of the level of hierarchy. Processing is always based on the top label in the stack which includes information about the operations to perform on the packet's label stack.

2.1.1.1 Label Values

Packets traveling along an LSP (see Label Switching Routers) are identified by its label, the 20-bit, unsigned integer. The range is 0 through 1,048,575. Label values 0 to 15 are reserved and are defined below as follows:

- A value of 0 represents the IPv4 Explicit NULL label. It indicates that the label stack must be popped, and the packet forwarding must be based on the IPv4 header. SR OS implementation does not support advertising an explicit-null label value, but can properly process in a received packet.
- A value of 1 represents the router alert label. This label value is legal anywhere in the label stack except at the bottom. When a received packet contains this label value at the top of the label stack, it is delivered to a local software module for processing. The actual packet forwarding is determined by the label beneath it in the stack. However, if the packet is further forwarded, the router alert label should be pushed back onto the label stack before forwarding. The use of this label is analogous to the use of the router alert option in IP packets. Since this label cannot occur at the bottom of the stack, it is not associated with a particular network layer protocol.
- A value of 2 represents the IPv6 explicit NULL label. It indicates that the label stack must be popped, and the packet forwarding must be based on the IPv6 header. SR OS implementation does not support advertising an explicit-null label value, but can properly process in a received packet.
- A value of 3 represents the Implicit NULL label. This is a label that a Label Switching Router (LSR) can assign and distribute, but which never actually appears in the encapsulation. When an LSR would otherwise replace the label at the top of the stack with a new label, but the new label is Implicit NULL, the LSR pops the stack instead of doing the replacement. Although this value may never appear in the encapsulation, it needs to be specified in the Label Distribution Protocol (LDP) or RSVP-TE protocol, so a value is reserved.
- A value of 7 represents the Entropy Label Indicator (ELI) which precedes in the label stack the actual Entropy Label (EL) which carries the entropy value of the packet.
- A value of 13 represents the Generic-ACH Label (GAL), an alert mechanism used to carry OAM payload in MPLS-TP LSP.
- Values 5-6, 8-12, and 14-15 are reserved for future use.

The router uses labels for MPLS, RSVP-TE, LDP, BGP Label Unicast, Segment Routing, as well as packet-based services such as VLL and VPLS.

Label values 16 through 1,048,575 are defined as follows:

- label values 16 through 31 are reserved for future use
- label values 32 through 18,431 are available for static LSP, MPLS-TP LSP, and static service label assignments. The upper bound of this range, which is also the lower bound of the dynamic label range, is configurable such that the user can expand or shrink the static or dynamic label range.
- label values 18,432 through 524,287 (1,048,575 in FP4 system profile B) are assigned dynamically by RSVP, LDP, and BGP control planes for both MPLS LSP and service labels.
- label values 524,288 through 1,048,575 are not assigned by SR OS in system profiles other than FP4 profile B, and thus no POP or SWAP label operation is possible in that range and for those system profiles. However, a PUSH operation, with a label from the full range 32 through 1,048,575 if signaled by some downstream LSR for LSP or service, is supported.
- The user can carve out a range of the dynamic label space dedicated for labels of the following features:
  - Segment Routing Global Block (SRGB) and usable by Segment Routing in OSPF and ISIS.
  - Reserved Label Block for applications such as SR policy, MPLS forwarding policy, and the assignment of a static label to the SID of a ISIS or OSPF adjacency and adjacency set.

2.1.1.2 Reserved Label Blocks

Reserved label blocks are used to reserve a set of labels for allocation for various applications. These reserved label blocks are separate from the existing ranges such as the static-labels-range, and are not tied to the bottom of the labels range. For example, a reserved range may be used as a Segment Routing Local Block (SRLB) for local segment identifiers (SIDs). Ranges are reserved from the dynamic label range and up to four reserved label block ranges may be configured on a system.

A reserved label block is configured using the following:

```
config
   router
      mpls-labels
         reserved-label-block <name>
            start <start-value> end <end-value>
            exit
         no reserved-label-block <name
```

A range can be configured up to the maximum supported MPLS label value on the system.

2.1.2 MPLS Entropy Label and Hash Label

The router supports both the MPLS entropy label, as specified in RFC 6790, and the flow-aware transport (FAT) label (the FAT label is also known as the hash label), as specified in RFC 6391. LSR nodes in a network can load-balance labeled packets in a more granular way than by hashing on the standard label stack by demarking the presence of individual flows on the LSP. The labels also remove the need to have an LSR inspect the payload below the label stack and check for an IPv4 or IPv6 header to determine how to apply load balancing.

The hash label is primarily applicable to Layer 2 services such as VLL and VPLS, while the entropy label (EL) is applicable to more general scenarios where a common way to indicate flows on a wide range of services suitable for load balancing is required.

The application of a hash label or an entropy label is mutually exclusive for a service.

2.1.2.1 Hash Label

The hash label is supported on VLL, VPRN, or VPLS services bound to any MPLS type encapsulated SDPs, as well as to a VPRN service using **auto-bind-tunnel** with the **resolution-filter** set to any MPLS tunnel type. When enabled, the ingress data path is modified such that the result of the hash on the payload packet header is communicated to the egress data path for use as the value of the label field of the hash label. The egress data path appends the hash label to the bottom of the stack (BoS) and sets the S-bit to 1. The user enables the signaling of the hash-label capability under a VLL spoke SDP, a VPLS spoke SDP or mesh SDP, or an IES or VPRN spoke SDP interface by adding the **signal-capability** option. When this capability is enabled, the decision to insert the hash label on the user and control plane packets by the local PE is determined by the outcome of the signaling process and may override the local PE configuration.

2.1.2.2 Entropy Label

The MPLS entropy label provides a similar function to the hash label but is applicable to a wider range of services. The entropy label is appended directly below the tunnel label. As with the hash label, the value of the entropy label is calculated based on a hash of the packet payload header.

The router supports the entropy label for the following services and protocols:

- VPRN
- EVPN VPLS and Epipe
- RFC 3107 MP-BGP tunnels
- RSVP and LDP LSPs used as shortcuts for static, IGP and BGP route resolution
- VLLs, including BGP VPWS, IES/VPRN, and VPLS spoke SDP termination, but not including Apipe and Cpipe
- LDP VPLS and BGP-AD VPLS

It is supported when used with the following tunnel types:

- RSVP-TE: Configured and auto-LSPs
- LDP
- Segment Routing (shortest path, configured SR-TE and SR-TE auto-LSPs)
- BGP

The entropy label is not supported on P2MP LSPs.

The entropy label indicated (ELI) label (value=7) is a special-purpose label that indicates that the entropy label follows in the stack. It is always placed immediately below the tunnel label to which hashing applies. Therefore, the EL results in two labels being inserted in the MPLS label stack; the EL and its accompanying ELI.

Three criteria are used to determine if an EL and an ELI are inserted on a labeled packet belonging to a service by an ingress LER:

- The Entropy Label Capability (ELC), which is the ability of the egress LER to receive and process the EL

  The ingress LER associates the ELC with the LSP tunnel to be used to transport the service. ELC signaling is supported for RSVP and LDP and causes the router to signal ELC to upstream peers. ELC is configured on these services by using the **config**>**router**>**rsvp**>**entropy-label-capability** and **config**>**router**>**ldp**>**entropy-label-capability** commands.

ELC signaling is not supported for BGP or SR tunnels. For these services, configure the ingress LER (or LSR at a stitching point to a BGP or SR segment) with ELC for this tunnel type using the **override-tunnel-elc** command for BGP or for the IGP if using SR.

- Whether a specific tunnel at the ingress LER supports EL

  Support for EL on a specific tunnel is configurable to prevent exceeding the maximum supported label stack depth due to the additional EL and ELI label (see Impact of EL and ELI on MTU and Label Stack Depth for more information). For RSVP and SR-TE LSPs, it is configured using the **entropy-label** command under the LSP, LSP template, or MPLS contexts.

- Whether the use of EL has been configured for the service

  Refer to the *L2 Services and EVPN Guide*, *L3 Services Guide*, and the *Unicast Routing Protocols Guide* for more information about entropy label configuration on services.

Each of these conditions must be true before the ingress LER inserts the EL and ELI into the label stack.

An LSR for RSVP and LDP tunnels passes the ELC from the downstream LSP segment to upstream peers. However, releases of SR OS that do not support EL functionality do not pass the ELC to their peers.

2.1.2.3 Inserting and Processing the Entropy Label at LERs and LSRs

This section describes entropy label processing. Details specific to particular services or other tunnel types are described in the *L2 Services and EVPN Guide*, *L3 Services Guide*, and the *Unicast Routing Protocols Guide*.

2.1.2.3.1 Ingress LER

The SR OS router follows the procedures at the ingress LER as specified in Section 4.2 of RFC 6790. In general, the router inserts an EL in a packet if the egress LER for the LSP tunnel has signaled support for ELs, the EL is configured for the service that the packet belongs to, and the EL is not disabled for an RSVP LSP. If there are multiple LSPs in a hierarchy (for example, LDP over RSVP), the router only inserts a single EL and ELI pair under the innermost LSP label closest to the service payload that has advertised EL capability. The router does not insert an EL in a packet belonging to a service for which the hash label has been configured, even if the far end for the LSP tunnel has advertised ELC. The system instead inserts a hash label, as specified by the hash label feature.

If the downstream LSR or LER has signaled implicit or explicit NULL label for a tunnel that is ELC, the router still inserts the EL when required by the service. This ensures consistent behavior as well as ensuring that entropy as determined by the ingress LER is maintained where a tunnel with an implicit NULL label is stitched at a downstream LSR.

2.1.2.3.2 LSR

If an LSR is configured for load balancing and an EL is found in the label stack, the LSR takes the EL into account in the hashing algorithm as follows:

- **label-only**: Only use the EL as input to the hash routine. The rest of the label stack is ignored.
- **label-ip**: Only use the EL and the IP packet as input to the hash routine. The rest of the label stack is ignored.

An EL and its associated ELI are not exposed when a tunnel label is swapped at an LSR acting as an LSP stitching point. Therefore, the EL and ELI are forwarded as any other packet on the LSP.

2.1.2.3.3 Egress LER

If an EL is detected in the label stack at an egress LER for a tunnel where the tunnel label that the EL is associated with is popped, then the EL is also popped and the packet is processed as normal. This occurs whether or not the system has signaled ELC.

If an ELI is popped that has the BoS bit set, then the system discards the packet and raises a trap.

2.1.2.4 Mapping Entropy Label Capability at LSP Stitching Points

A router acting as a stitching point between two LSPs maps the ELC received in signaling for a downstream segment to the upstream segment for the level in the LSP hierarchy being stitched.

If an LSR is stitching an RSVP or LDP segment to a downstream segment of a tunnel type that does not support ELC signaling (for example, BGP) and **override-tunnel-elc** is configured at the LSR for to downstream segment, then the system signals ELC on the upstream LSP segment. The **override-tunnel-elc** command must be configured to reflect whether all possible downstream LERs are entropy-label-capable; otherwise, packets with an EL are discarded by a downstream LER that is not entropy-label-capable.

The mapping of ELC across LDP-BGP stitching points is not supported. If a downstream tunnel endpoint signals ELC, this signal is not automatically propagated upstream. The EL and ELI are not inserted on these LSPs by the ingress LER.

2.1.2.5 Entropy Label on OAM Packets

Service OAM packets or OAM packets within the context of a shortcut (for example, ICMP Ping or traceroute packets), also include an EL and ELI if ELC is signaled for the corresponding tunnel and the **entropy-label** command is enabled for the service. The EL and ELI is inserted at the same level in the label stack as it is in user data packets, which is under the innermost LSP label closest to the service payload that has advertised ELC. The EL and ELI therefore always reside at a different level in the label stack than the special-purpose labels related to the service payload (such as the Router Alert label). OAM packets at the LSP level, such as LSP ping and LSP trace, do not have the EL and ELI inserted.

2.1.2.6 Impact of EL and ELI on MTU and Label Stack Depth

If EL insertion is configured for a VPLS or VLL service, the MTU of the SDP binding is automatically reduced to account for the overhead of the EL and ELI labels. The MTU is reduced whether or not the LSP tunnel used by the service is entropy-label-capable.

The EL requires the insertion of two additional labels in the label stack. In some cases, the insertion of EL and ELI may result in an unsupported label stack depth or large changes in the label stack depth during the lifetime of an LSP. For RSVP LSPs, the **entropy-label** command under the **config**>**router**>**mpls** and **config**>**router**>**mpls**>**lsp** contexts provides local control at the head-end of an LSP over whether the entropy label is inserted on an LSP irrespective of the entropy label capability signaled from the egress LER, and control over how the additional label stack depth is accounted for. This control allows a user to avoid entropy label insertion where there is a risk of the label stack becoming too deep.

2.1.3 Label Switching Routers

LSRs perform the label switching function. LSRs perform different functions based on its position in an LSP. Routers in an LSP do one of the following:

- The router at the beginning of an LSP is the ingress label edge router (ILER). The ingress router can encapsulate packets with an MPLS header and forward it to the next router along the path. An LSP can only have one ingress router.
- A Label Switching Router (LSR) can be any intermediate router in the LSP between the ingress and egress routers. An LSR swaps the incoming label with the outgoing MPLS label and forwards the MPLS packets it receives to the next router in the MPLS path (LSP). An LSP can have 0 to 253 transit routers.
- The router at the end of an LSP is the egress label edge router (eLER). The egress router strips the MPLS encapsulation which changes it from an MPLS packet to a data packet, and then forwards the packet to its final destination using information in the forwarding table. Each LSP can have only one egress router. The ingress and egress routers in an LSP cannot be the same router.

A router in your network can act as an ingress, egress, or transit router for one or more LSPs, depending on your network design.

An LSP is confined to one IGP area for LSPs using constrained-path. They cannot cross an autonomous system (AS) boundary.

Static LSPs can cross AS boundaries. The intermediate hops are manually configured so the LSP has no dependence on the IGP topology or a local forwarding table.

2.1.3.1 LSP Types

The following are LSP types:

- Static LSPs — A static LSP specifies a static path. All routers that the LSP traverses must be configured manually with labels. No signaling such as RSVP or LDP is required.
- Signaled LSP — LSPs are set up using a signaling protocol such as RSVP-TE or LDP. The signaling protocol allows labels to be assigned from an ingress router to the egress router. Signaling is triggered by the ingress routers. Configuration is required only on the ingress router and is not required on intermediate routers. Signaling also facilitates path selection.

  There are two signaled LSP types:

- Explicit-path LSPs — MPLS uses RSVP-TE to set up explicit path LSPs. The hops within the LSP are configured manually. The intermediate hops must be configured as either 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). You can control how the path is set up. They are similar to static LSPs but require less configuration. See RSVP.
- Constrained-path LSPs — The intermediate hops of the LSP are dynamically assigned. A constrained path LSP relies on the Constrained Shortest Path First (CSPF) routing algorithm to find a path which satisfies the constraints for the LSP. In turn, CSPF relies on the topology database provided by the extended IGP such as OSPF or IS-IS.

  Once the path is found by CSPF, RSVP uses the path to request the LSP set up. CSPF calculates the shortest path based on the constraints provided such as bandwidth, class of service, and specified hops.

If fast reroute is configured, the ingress router signals the routers downstream. Each downstream router sets up a detour for the LSP. If a downstream router does not support fast reroute, the request is ignored and the router continues to support the LSP. This can cause some of the detours to fail, but otherwise the LSP is not impacted.

No bandwidth is reserved for the rerouted path. If the user enters a value in the bandwidth parameter in the **config>router>mpls>lsp>fast-reroute** context, it has no effect on the LSP backup LSP establishment.

Hop-limit parameters specifies the maximum number of hops that an LSP can traverse, including the ingress and egress routers. An LSP is not set up if the hop limit is exceeded. The hop count is set to 255 by default for the primary and secondary paths. It is set to 16 by default for a bypass or detour LSP path.

2.1.4 Bidirectional Forwarding Detection for MPLS LSPs

BFD for MPLS LSPs monitors the LSP between its LERs, regardless of how many LSRs the LSP may traverse. Therefore, it enables local faults on individual LSPs to be detected, whether or not they also affect forwarding for other LSPs or IP packet flows. This makes BFD for MPLS LSPs ideal for monitoring LSPs carrying specific high-value services, where detecting forwarding failures in the minimal amount of time is critical. The system raises an SNMP trap, and indicates the BFD session state in show and tools dump commands if an LSP BFD session goes down. It can also optionally determine the availability of the tunnel in TTM for use by applications, or trigger a switchover of the LSP from the currently active path to a backup path.

The system supports LSP BFD on RSVP LSPs. See Label Distribution Protocol for information about using LSP BFD on LDP LSPs see Seamless BFD for SR-TE LSPs for information about Seamless BFD on SR-TE LSPs. BFD packets are encapsulated in an MPLS label stack corresponding to the FEC that the BFD session is associated with, as described in Section 7 of RFC 5884, *Bidirectional Forwarding Detection (BFD) for MPLS Label Switched Paths (LSPs).*

Since RSVP LSPs are unidirectional, a routed return path is used for the BFD control packets from the egress LER towards the ingress LER.

2.1.4.1 Bootstrapping and Maintaining the BFD Session

A BFD session on an LSP is bootstrapped using LSP ping. LSP ping is used to exchange the local and remote discriminator values to use for the BFD session for a particular MPLS LSP or FEC.

SR OS supports the sending of periodic LSP ping messages on an LSP for which LSP BFD has been configured, as specified in RFC 5884. The ping messages are sent, along with the bootstrap TLV, at a configurable interval for LSPs on which **bfd-enable** has been configured. The default interval is 60 s, with a maximum interval of 300 s. The LSP ping echo request message uses the system IP address as the default source address. An alternative source address consisting of any routable address that is local to the node may be configured, and is used if the local system IP address is not routable from the far-end node.

➡ **Note:** SR OS does not take any action if a remote system fails to respond to a periodic LSP ping message. However, when the **show**>**test-oam**>**lsp-bfd** command is executed, it displays a return code of zero and a replying node address of 0.0.0.0 if the periodic LSP ping times out.

The periodic LSP ping interval is configured using the **config**>**router**>**mpls**>**lsp**>**bfd**>**lsp-ping-interval** *seconds* command.

The **no lsp-ping-interval** command reverts to the default of 60 s.

LSP BFD sessions are recreated after a high-availability switchover between active and standby CPMs. However, some disruption may occur to LSP ping due to LSP BFD.

At the tail end of an LSP, sessions are recreated on the standby CPM following an HA switchover. The following current information is lost from an active **tools dump test-oam lsp-bfd tail** display:

- handle
- seqNum
- rc
- rsc

Any new, incoming bootstrap requests are dropped until LSP BFD has become active. When LSP BFD has finished becoming active, new bootstrap requests are considered.

2.1.4.2 LSP BFD Configuration

There are four steps to configuring LSP BFD.

1. Configure a BFD template.
2. Enable LSP BFD on the tail node and configure the maximum number of LSP BFD sessions at the tail node.

Note: The default number of LSP BFD sessions is zero.

3. Apply a BFD template to the LSP or LSP path.
4. Enable BFD on the LSP or LSP path.

LSP BFD uses BFD templates to set generic BFD session parameters.

The BFD template is configured as follows:

```
config
   router
      bfd
         bfd-template name
            transmit-interval transmit-interval
            receive-interval receive-interval
            echo-receive echo-interval
            multiplier multiplier
            exit
```

Network processor BFD is not supported for LSPs. The minimum supported BFD receive or transmit timer interval for RSVP LSPs is 100 milliseconds. Therefore, an error is generated if a user tries to bind a BFD template with the **type cpm-np** command or any unsupported transmit or receive interval value to an LSP. An error is generated when the user attempts to commit changes to a BFD template that is already bound to an LSP if the new values are invalid for LSP BFD.

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 is bound tries to renegotiate their timers to the new values.

The bfd-template uses a begin-commit model. To edit any value within the BFD template, a <begin> needs to be executed before the template context has been entered. However, a value is stored temporarily in the template-module until the commit is issued. Values are actually used once the commit is issued.

2.1.4.3 Enabling and Implementing Limits for LSP BFD on a Node

The **config>router>lsp-bfd** command enables support for LSP BFD and allows an upper limit to the number of supported sessions at the tail end node for LSPs, where it is disabled by default. This is useful because BFD resources are shared among applications using BFD, so a user may wish to set an upper limit to ensure that a certain number of BFD sessions are reserved for other applications. This is important at the tail end of LSPs where no per-LSP configuration context exists.

LSP BFD is enabled or disabled on a node-wide basis using the **bfd-sessions** *max-limit* command under the **config>router>lsp-bfd** context. This command also enables the maximum number of LSP BFD sessions that can be established at the tail end of LSPs to be limited.

The default is disabled. The *max-limit* parameter specifies the maximum number of LSP BFD sessions that the system allows to be established at the tail end of LSPs.

2.1.4.4 BFD Configuration on RSVP-TE LSPs

LSP BFD is applicable to configured RSVP LSPs as well as mesh-p2p and one-hop-p2p auto-LSPs.

LSP BFD is configured on an RSVP-TE LSP, or on the primary or secondary 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-enable** command.

```
config
   router
      mpls
         lsp
            bfd
               [no] bfd-template <name>
```

```
              [no] bfd-enable
              [no] wait-for-up-timer <value>
              exit
```

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

The **bfd-template** provides the control packet timer values for the BFD session to use at the LSP head end. Since there is no LSP configuration at the tail end of an RSVP LSP, the BFD state machine at the tail end initially uses system-wide default parameters (the timer values are: min-tx: 1sec, min-rx: 1sec). The head end then attempts to adjust the control packet timer values when it transitions to the INIT state.

The BFD command **wait-for-up-timer** allows RSVP LSPs BFD sessions to come up during certain MBB and switchover events when the current active path is not BFD degraded (that is, BFD is not down). It is only applicable in cases where **failure-action failover-or-down** is also configured (see Using LSP BFD for LSP Path Protection) and applies to the following:

- a path undergoing MBB when BFD is up on the original path
- the initial administrative enable of an LSP
- signaling retry of non-standby secondary paths

The **wait-for-up-timer** command is configurable at either the **lsp**>**bfd**, **primary**>**bfd**, or **secondary**>**bfd** context. The value that the system uses is the one configured under the same context in which **bfd-enable** is configured. The **wait-for-up-timer** has a range of 1-60 seconds and a default of 4 seconds.

BFD is configured at the primary path level, as follows:

```
config
   router
      mpls
         lsp
            primary <path-name>
               bfd
                  [no] bfd-template name <name>
                  [no] bfd-enable
                  [no] wait-for-up-timer <value>
                   exit
```

BFD is configured on both standby and non-standby secondary paths as follows:

```
config
   router
      mpls
         lsp
            secondary <path-name>
               bfd
                  [no] bfd-template <name>
                  [no] bfd-enable
                  [no] wait-for-up-timer <value>
                   exit
```

BFD sessions are not established on these paths unless they are made active, unless **failure-action failover-or-down** is configured. See Using LSP BFD for LSP Path Protection. If **failure-action failover-or-down** is configured then the top three best-preference primary and standby paths (primary and up to two standby paths, or three standby paths if no primary is present) are programmed in the IOM, and BFD sessions are established on all of them.

It is not possible to configure LSP BFD on a secondary path or on P2MP LSPs.

LSP BFD at the LSP level and the path level is mutually exclusive. That is, if LSP BFD is already configured for the LSP then 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-LSPs. In this case, LSP BFD is configured on mesh-p2p and one-hop-p2p auto-LSPs using the LSP template, as follows:

```
Config
   router
      mpls
         lsp-template template-name {mesh-p2p | one-hop-p2p}
            bfd
               [no] bfd-template name
               [no] bfd-enable
               exit
```

2.1.4.5 Using LSP BFD for LSP Path Protection

SR OS can determine the forwarding state of an LSP from the LSP BFD session, allowing users of the LSP to determine whether their transport is operational. If BFD is down on an LSP path, then the path is considered to be BFD degraded by the system.

Using the **failure-action** command, a user can configure the action taken by the system if BFD fails for an RSVP LSP or LDP prefix list. There are three possible failure actions:

- **failure-action down** — the LSP is marked as unusable in TTM when BFD on the LSP goes down. This is applicable to RSVP and LDP LSPs.

- **failure-action failover** — when LSP BFD goes down on the currently active path, then the LSP switches from the primary path to the secondary path, or from the currently active secondary path to the next-best preference secondary path. This is applicable to RSVP LSPs.
- **failover-action failover-or-down** — similar to **failover-action failover**, when LSP BFD goes down on the currently active path, then the LSP switches from the primary path to the secondary path, or from the currently active secondary path to the next best preference secondary path. However, **failover-or-down** also supports the ability to run BFD sessions simultaneously on the primary and up to two other secondary or standby paths. The system does not switch to a standby path for which the BFD session is down. If all BFD sessions for the LSP are down, then the LSP is marked as unusable in TTM. This is applicable to RSVP LSPs and SR-TE LSPs. See Support for BFD Failure Action with SR-TE LSPs for further details of its use for SR-TE LSPs.

In all cases, an SNMP trap is raised indicating that BFD has gone down on the LSP.

Note: It is recommended that BFD control packet timers are configured to a value that is large enough to allow for transient data path disruptions that may occur when the underlying transport network recovers following a failure.

2.1.4.5.1 Failure-action down

The **failure-action down** command is supported for point-to-point RSVP (including mesh point-to-point and one-hop point-to-point auto-LSPs) and LDP LSPs. This command is configured within the **config**>**router**>**mpls**>**lsp**>**bfd**, **config**>**router**>**mpls**>**lsp-template**>**bfd**, or **config**>**router**>**ldp**>**lsp-bfd** contexts. For RSVP LSPs, it is only supported at the LSP level and not at the primary or secondary path levels. When configured, an LSP is made unavailable as a transport if BFD on the LSP goes down.

If BFD is disabled, MPLS installs the LSP as “usable” in the TTM. The **failure-action** configuration is ignored.

If BFD is enabled and **no failure-action** is configured, MPLS installs the LSP as “usable” in the TTM regardless of the BFD session state. BFD generates BFD Up and BFD Down traps.

If BFD is enabled and **failure-action down** is configured:

- BFD traps are still generated when the BFD state machine transitions.

- If the BFD session is up for the active path of the LSP, the LSP is installed as “usable” in the TTM. If the BFD session is down for the active path, the LSP is installed as “not-usable” in the TTM.
- When an LSP is first activated using the **no shutdown** command, and its LSP BFD session first starts to come up, the LSP is installed as “not-usable” in the TTM to any user until the BFD session transitions to the up state, despite the FEC for the corresponding LSP being installed by the TTM. Users include all protocols, including those in RTM. A tunnel that is marked as down in the TTM is not available to RTM, and all routes using it are withdrawn. SDP auto-bind does not make use of an LSP until it is installed as “usable”.
- If the BFD session is up on the active path and the LSP is installed as “usable” in the TTM, and if the LSP switches from its current active path to a new path, the system triggers a new BFD bootstrap using LSP ping for the new path, and waits for a maximum of 10 s for the BFD session to come up on the new path before switching traffic to it. If the BFD session does not come up on the new path after 10 s, the system switches to the new path anyway and install the LSP as “not-usable” in the TTM. This is the only scenario where a switch of the active path can be delayed due to BFD transition state.
- If the BFD session is down on the active path and the LSP was already installed as “not-usable” in the TTM, then the system immediately switches to the new path without waiting for BFD to become operationally up.
- If BFD is disabled, MPLS installs the LSP as “usable” in the TTM. The **failure-action** configuration is ignored. LSP ping and LSP trace are still able to test an LSP when BFD is disabled.

Note: BFD session state is never used to trigger a switch of the active path when **failure-action down** is configured.

2.1.4.5.2 Failure-action failover

The **failure-action failover** command is supported for point-to-point RSVP LSPs (except mesh point-to-point and one-hop point-to-point auto-LSPs because these do not have a secondary path). When failure action failover is configured, the system triggers a failover from the currently active path to the secondary path, the next-best preference secondary path, or the secondary-standby path of an LSP when an LSP BFD session configured at the LSP level transitions from an up state to a down state. Unlike **failure-action failover-or-down**, this failure action does not affect how LSP paths are programmed in the data path and only runs LSP BFD on the active path.

The LSP is always marked as usable in the TTM, regardless of the BFD session state and BFD traps that are generated when the BFD state machine transitions. If BFD is enabled and failure-action failover is configured, the following conditions apply.

- It is possible to bring the LSP up regardless the current BFD session state.
- If the BFD session transitions from up to down, the current path immediately switches to the next-best preference standby path.
- If MBB is triggered, then this occurs immediately on the primary path, regardless the BFD session state.
- If the operator is concerned about detecting data path failures that may not be detected by the control plane, Nokia recommends that the revert timer be set to its maximum value.
- 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 switches to is up and forwarding. The system relies on the normal control plane mechanisms.

Table 4 describes how the system behaves if a user changes the failure-action while BFD is down. The LSP remains on the current path unless (or until) the control plane takes action or the revert timer expires.

Table 4 **Changes to the Failure Action while BFD is Down**

Action Combination (old action/new action)	Tunnel flag in TTM
None/Down	as unusable
None/Failover	as usable
Down/None	as usable
Down/Failover	as usable
Failover/None	as usable
Failover/Down	as unusable

2.1.4.5.3 LSP Active Path Failover Triggers

The active path of an LSP is switched to an alternative path in the following cases:

- the active path goes into degraded state due to FRR or soft preemption
- the active path is degraded due to the BFD session going from up to down (only applicable if the failure action is set to **failover** or **failover-or-down**)
- reverting from a secondary or standby path to the primary path (with or without a reverter time configured)
- switching between secondary or standby paths due to path preference

- switching between secondary or standby paths due to the **tools perform router mpls switch-path** or **force-switch-path** commands
- switching due to an MBB on the active path where the old and new path have the same **bfd-enable** configuration

Table 5 describes path switchover events depending on the failure action configuration.

Table 5 **Path Switchover Triggers based on BFD Failure Action Configuration**

BFD failure-action configuration	Old active path		New active path	Switchover to new path
	bfd-enable configuration at LSP or path	BFD session state	bfd-enable configuration at LSP or path	
no failure-action **failure-action failover**	Any	Any	Any	Switch immediately without checking the BFD session state on new path.
failure-action down	BFD enabled	BFD session up	BFD enabled	Wait for a maximum of 10 seconds for the BFD session to come up on the new path before switching. If the BFD session does not come up on the new path after 10 seconds, switch anyway.
			BFD disabled	Switch immediately without checking the BFD session state on new path.
		BFD session down	BFD enabled	Switch immediately without checking the BFD session state on new path.
			BFD disabled	Switch immediately without checking the BFD session state on new path.

Table 5 **Path Switchover Triggers based on BFD Failure Action Configuration (Continued)**

BFD failure-action configuration	Old active path		New active path	Switchover to new path
	bfd-enable configuration at LSP or path	BFD session state	bfd-enable configuration at LSP or path	
	BFD disabled	—	BFD enabled	Wait for a maximum of 10 seconds for the BFD session to come up on the new path before switching. If the BFD session does not come up on the new path after 10 seconds, switch anyway.
			BFD disabled	Switch immediately without checking the BFD session state on new path.

For **failure-action failover-or-down**, a path is in the degraded state if it has BFD enabled and the BFD session is not up. Switching between primary, standby, and secondary paths of the LSP will follow rules of best path selection algorithm, for example, a non-degraded path is better than a degraded path and a degraded primary is better than a degraded standby or secondary path. Since the BFD degraded state affects LSP active path selection, waiting for BFD to come up on new path is already accounted for and these cases have been excluded from Table 6.

Switching to an MBB path requires waiting for the BFD session to come up on the new MBB path. These cases are described in Table 6. This applies to MBB on both active and inactive paths to reduce the toggling of a BFD degraded state on the path.

Table 6 **MBB Path Switching with failure-action failover-or-down**

BFD failure-action configuration	Old path		New MBB path	Switching to new path
	bfd-enable configuration at LSP or path	BFD session state	bfd-enable configuration at LSP or path	
failure-action failover-or-down	BFD enabled	BFD session up	BFD enabled	Wait for a maximum of “*w*” seconds for the BFD session to come up on the new path before switching. If the BFD session does not come up on the new path after “*w*” seconds, switch anyway. Where *w* is the BFD **wait-for-up-timer** from the context where BFD is enabled.
			BFD disabled	This case is not applicable because the MBB path has same BFD configuration as existing path.
	BFD enabled	BFD session down	BFD enabled	Switch immediately without checking the BFD session state on new path.
			BFD disabled	This case is not applicable because the MBB path has same BFD configuration as existing path.
	BFD disabled	—	BFD enabled	This case is not applicable because the MBB path has the same BFD configuration as existing path.
			BFD disabled	Switch immediately without checking the BFD session state on new path.

2.1.4.6 MPLS/RSVP on Broadcast Interface

The MPLS/RSVP on Broadcast Interface feature allows MPLS/RSVP to distinguish neighbors from one another when the outgoing interface is a broadcast interface connecting to multiple neighbors over a broadcast domain. More specifically, in the case where a BFD session towards a specific neighbor on the broadcast domain goes down, the consecutive actions (for example, FRR switchover) only concerns the LSPs of the affected neighbor. Previously, the actions would have been taken on the LSPs of all neighbors over the outgoing interface.

2.1.5 MPLS Facility Bypass Method of MPLS Fast Re-Route (FRR)

The MPLS facility bypass method of MPLS Fast Re-Route (FRR) functionality is extended to the ingress node.

The behavior of an LSP at an ingress LER with both fast reroute and a standby LSP path configured is as follows:

- When a downstream detour becomes active at a point of local repair (PLR):

  The ingress LER switches to the standby LSP path. If the primary LSP path is repaired subsequently at the PLR, the LSP switches back to the primary path. If the standby goes down, the LSP is switched back to the primary, even though it is still on the detour at the PLR. If the primary goes down at the ingress while the LSP is on the standby, the detour at the ingress is cleaned up and for one-to-one detours a “path tear” is sent for the detour path. In other words, the detour at the ingress does not protect the standby. If and when the primary LSP is again successfully re-signaled, the ingress detour state machine is restarted.
- When the primary fails at the ingress:

  The LSP switches to the detour path. If a standby is available then LSP would switch to standby on expiration of **hold-timer**. If **hold-timer** is disabled then switchover to standby would happen immediately. On successful global revert of primary path, the LSP would switch back to the primary path.
- Admin groups are not taken into account when creating detours for LSPs.

2.1.6 Manual Bypass LSP

SR OS implements dynamic bypass tunnels as defined in RFC 4090, *Fast Reroute Extensions to RSVP-TE for LSP Tunnels*. When an LSP is signaled and the local protection flag is set in the session_attribute object and/or the FRR object in the path message indicates that facility backup is desired, the PLR establishes a bypass tunnel to provide node and link protection. The bypass tunnel is selected if a bypass LSP that merges in a downstream node with the protected LSP exists, and if this LSP satisfies the constraints in the FRR object.

With the manual bypass feature, an LSP can be preconfigured from a PLR that is used exclusively for bypass protection. When a Path message for a new LSP requests bypass protection, the node first checks if a manual bypass tunnel satisfying the path constraints exists. If one is found, it is selected. If no manual bypass tunnel is found, the router dynamically signals a bypass LSP in the default behavior. Users can disable the dynamic bypass creation on a per node basis using the CLI.

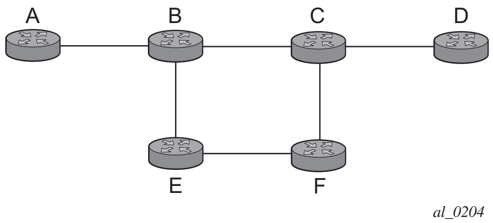
A maximum of 1000 associations of primary LSP paths can be made with a single manual bypass by default. The **max-bypass-associations** integer command allows to increase or decrease the number of associations. If dynamic bypass creation is disabled on the node, it is recommended to configure additional manual bypass LSPs to handle the required number of associations.

Refer to Configuring Manual Bypass Tunnels for configuration information.

2.1.6.1 PLR Bypass LSP Selection Rules

The PLR uses rules to select a bypass LSP among multiple manual and dynamic bypass LSPs at the time of establishment of the primary LSP path or when searching for a bypass for a protected LSP which does not have an association with a bypass tunnel: Figure 3 shows an example of bypass tunnel nodes.

Figure 3 **Bypass Tunnel Nodes**



The rules are:

1. The MPLS task in the PLR node checks if an existing manual bypass satisfies the constraints. If the path message for the primary LSP path indicated node protection desired, which is the default LSP FRR setting at the head end node, MPLS task searches for a node-protect bypass LSP. If the path message for the primary LSP path indicated link protection desired, then it searches for a link-protect bypass LSP.
2. If multiple manual bypass LSPs satisfying the path constraints exist, it prefers a manual-bypass terminating closer to the PLR over a manual bypass terminating further away. If multiple manual bypass LSPs satisfying the path constraints terminate on the same downstream node, it selects one with the lowest IGP path cost or if in a tie, picks the first one available.
3. If none satisfies the constraints and dynamic bypass tunnels have not been disabled on PLR node, then the MPLS task in the PLR checks if any of the already established dynamic bypasses of the requested type satisfy the constraints.
4. If none do, then the MPLS task asks CSPF to check if a new dynamic bypass of the requested type, node-protect or link-protect, can be established.
5. If the path message for the primary LSP path indicated node protection desired, and no manual bypass was found after Step 1, and/or no dynamic bypass LSP was found after one attempt of performing Step 3, the MPLS task repeats Steps 1 to 3 looking for a suitable link-protect bypass LSP. If none are found, the primary LSP has no protection and the PLR node must clear the “local protection available” flag in the IPv4 address sub-object of the RRO starting in the next Resv refresh message it sends upstream. Node protection continues to be attempted using a background re-evaluation process.
6. If the path message for the primary LSP path indicated link protection desired, and no manual bypass was found after step 1, and/or no dynamic bypass LSP was found after performing Step 3, the primary LSP has no protection and the PLR node must clear the “local protection available” flag in the IPv4 address sub-object of the RRO starting in the next RESV refresh message it sends upstream. The PLR will not search for a node-protect’ bypass LSP in this case.
7. If the PLR node successfully makes an association, it must set the “local protection available” flag in the IPv4 address sub-object of the RRO starting in the next RESV refresh message it sends upstream.
8. For all primary LSP that requested FRR protection but are not currently associated with a bypass tunnel, the PLR node on reception of RESV refresh on the primary LSP path repeats Steps 1 to 7.

If the user disables dynamic-bypass tunnels on a node while dynamic bypass tunnels were activated and were passing traffic, traffic loss will occur on the protected LSP. Furthermore, if no manual bypass exist that satisfy the constraints of the protected LSP, the LSP will remain without protection.

If the user configures a bypass tunnel on node B and dynamic bypass tunnels have been disabled, LSPs which have been previously signaled and which were not associated with any manual bypass tunnel, for example, none existed, are associated with the manual bypass tunnel if suitable. The node checks for the availability of a suitable bypass tunnel for each of the outstanding LSPs every time a RESV message is received for these LSPs.

If the user configures a bypass tunnel on node B and dynamic bypass tunnels have not been disabled, LSPs which have been previously signaled over dynamic bypass tunnels will not automatically be switched into the manual bypass tunnel even if the manual bypass is a more optimized path. The user will have to perform a make before break at the head end of these LSPs.

If the manual bypass goes into the down state in node B and dynamic bypass tunnels have been disabled, node B (PLR) will clear the “protection available” flag in the RRO IPv4 sub-object in the next RESV refresh message for each affected LSP. It will then try to associate each of these LSPs with one of the manual bypass tunnels that are still up. If it finds one, it will make the association and set again the “protection available” flag in the next RESV refresh message for each of these LSPs. If it could not find one, it will keep checking for one every time a RESV message is received for each of the remaining LSPs. When the manual bypass tunnel is back UP, the LSPs which did not find a match are associated back to this tunnel and the protection available flag is set starting in the next RESV refresh message.

If the manual bypass goes into the down state in node B and dynamic bypass tunnels have not been disabled, node B will automatically signal a dynamic bypass to protect the LSPs if a suitable one does not exist. Similarly, if an LSP is signaled while the manual bypass is in the down state, the node will only signal a dynamic bypass tunnel if the user has not disabled dynamic tunnels. When the manual bypass tunnel is back into the UP state, the node will not switch the protected LSPs from the dynamic bypass tunnel into the manual bypass tunnel.

2.1.6.2 FRR Facility Background Evaluation Task

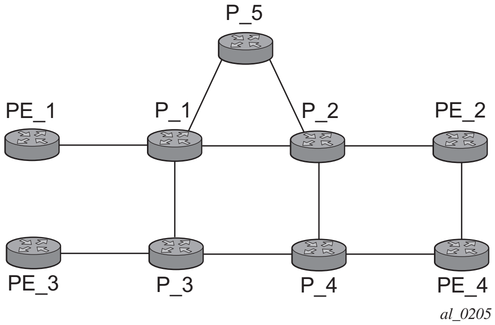
The MPLS Fast Re-Route (FRR) feature implements a background task to evaluate Path State Block (PSB) associations with bypass LSP. The following is the task evaluation behavior.

- For PSBs that have facility FRR enabled but no bypass association, the task triggers a FRR protection request.
- For PSBs that have requested node-protect bypass LSP but are currently associated with a link-protect bypass LSP, the task triggers a node-protect FRR request.

- For PSBs that have LSP statistics enabled but the statistic index allocation failed, the task re-attempts index allocation.

The MPLS FRR background task thus enables PLRs to be aware of the missing node protection and lets them regularly probe for a node-bypass. Figure 4 shows an example of FRR node protection.

Figure 4 **FRR Node-Protection Example**



The following describes an LSP scenario where:

- LSP 1: from PE_1 to PE_2, with CSPF, FRR facility node-protect enabled
- P_1 protects P_2 with bypass-nodes P_1 -P_3 - P_4 - PE_4 -PE_2
- If P_4 fails, P_1 tries to establish the bypass-node three times
- When the bypass-node creation fails, P_1 will protect link P_1-P_2
- P_1 protects the link to P_2 through P_1 - P_5 - P_2
- P_4 returns online

LSP 1 has requested node protection, but due to the lack of an available path, it can only obtain link protection. Therefore, every 60 seconds the PSB background task triggers the PLR for LSP 1 to search for a new path that can provide node protection. Once P_4 is back online and such a path is available, a new bypass tunnel is signaled and LSP 1 gets associated with this new bypass tunnel.

2.1.7 Uniform FRR Failover Time

The failover time during FRR consists of a detection time and a switchover time. The detection time corresponds to the time it takes for the RSVP control plane protocol to detect that a network IP interface is down or that a neighbor/next-hop over a network IP interface is down. The control plane can be informed of an interface down event when event is due to a failure in a lower layer such in the physical layer. The control plane can also detect the failure of a neighbor/next-hop on its own by running a protocol such as Hello, Keep-Alive, or BFD.

The switchover time is measured from the time the control plane detects the failure of the interface or neighbor/next-hop to the time the XCMs or IOMs completes the reprogramming of all the impacted ILM or service records in the data path. This includes the time it takes for the control plane to send a down notification to all XCMs or IOMs to request a switch to the backup NHLFE.

Uniform Fast-Reroute (FRR) failover enables the switchover of MPLS and service packets from the outgoing interface of the primary LSP path to that of the FRR backup LSP within the same amount of time regardless of the number of LSPs or service records. This is achieved by updating Ingress Label Map (ILM) records and service records to point to the backup Next-Hop Label to Forwarding Entry (NHLFE) in a single operation.

2.1.8 Automatic Bandwidth Allocation for RSVP LSPs

2.1.8.1 Enabling and Disabling Auto-Bandwidth Allocation on an LSP

This section discusses an auto-bandwidth hierarchy configurable in the **config**>**router**>**mpls**>**lsp** context.

Adding auto-bandwidth at the LSP level starts the measurement of LSP bandwidth described in Measurement of LSP Bandwidth and allows auto-bandwidth adjustments to take place based on the triggers described in Periodic Automatic Bandwidth Adjustment.

When an LSP is first established, the bandwidth reserved along its primary path is controlled by the bandwidth parameter in the **config**>**router**>**mpls**>**lsp**>**primary** context, whether or not the LSP has auto-bandwidth enabled, while the bandwidth reserved along a secondary path is controlled by the bandwidth parameter in the **config**>**router**>**mpls**>**lsp**>**secondary** context. When auto-bandwidth is enabled and a trigger occurs, the system will attempt to change the bandwidth of the LSP to a value between **min-bandwidth** and **max-bandwidth**, which are configurable values in the **lsp**>**auto-bandwidth** context. **min-bandwidth** is the minimum bandwidth that **auto-bandwidth** can signal for the LSP, and **max-bandwidth** is the maximum bandwidth that can be signaled. The user can set the **min-bandwidth** to the same value as the primary path bandwidth but the system will not enforce this restriction. The system will allow:

- no **min-bandwidth** to be configured. In this case, the implicit minimum is 0 Mb/s.
- no **max-bandwidth** to be configured, as long as overflow-triggered auto-bandwidth is not configured. In this case, the implicit maximum is infinite (effectively 100 Gb/s).
- the configured primary path bandwidth to be outside the range of **min-bandwidth** to **max-bandwidth**
- **auto-bandwidth** parameters can be changed at any time on an operational LSP; in most cases, the changes have no immediate impact, but subsequent sections will describe some exceptions

All of the auto-bandwidth adjustments discussed are performed using MBB procedures.

Auto bandwidth can be added to an operational LSP at any time (without the need to shut down the LSP or path), but no bandwidth change occurs until a future trigger event. Auto bandwidth may also be removed from an operational LSP at any time and this causes an immediate MBB bandwidth change to be attempted using the configured primary path bandwidth.

A change to the configured bandwidth of an auto-bandwidth LSP has no immediate effect. The change only occurs if the LSP/path goes down (due to failure or administrative action) and comes back up, or if auto-bandwidth is removed from the LSP. The operator can force an auto-bandwidth LSP to be resized immediately to an arbitrary bandwidth using the appropriate tools commands.

2.1.8.2 Autobandwidth on LSPs with Secondary or Secondary Standby Paths

Autobandwidth is supported for LSPs that have secondary or secondary standby paths. A secondary path is only initialized at its configured bandwidth when it is established, and the bandwidth is adjusted only when the secondary path becomes active.

This description makes use of the following terminology:

- current_BW: the last known reserved bandwidth for the LSP; may be the value of a different path from the currently active path
- operational BW: the last known reserved BW for a given path, as recorded in the MIB
- configured BW: the bandwidth explicitly configured for the LSP path by the user in CLI
- active path: the path (primary or secondary) the LSP currently uses to forward traffic
- signaled BW: the new BW value signaled during an MBB

A secondary or standby secondary path is initially signaled with its configured bandwidth. Setup for the secondary path is triggered only when the active path goes down or becomes degraded (for example, due to FRR or preemption). An auto-BW triggered bandwidth adjustment (auto bandwidth MBB) only takes place on the active path. For example, if an auto-BW adjustment occurs on the primary path, which is currently active, no adjustment is made at that time to the secondary path since that path is not active.

When the active path changes, the current_bw is updated to the operational bandwidth of the newly active path. While the auto-BW MBB on the active path is in progress, a statistics sample could be triggered, and this would be collected in the background. Auto-bandwidth computations will use the current_bw of the newly active path. In case the statistics sample collection results in a bandwidth adjustment, the in-progress auto-BW MBB is restarted. If after five attempts, the auto-BW MBB fails, the current_bw and secondary operational BW remain unchanged.

For a secondary or standby secondary path, if the active path for an LSP changes (without the LSP going down), an auto-BW MBB is triggered for the new active path. The bandwidth used to signal the MBB is the operational bandwidth of the previous active path. If the MBB fails, it will retry with a maximum of five attempts. The reserved bandwidth of the newly active path will therefore be 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 setup. If the LSP goes down, all statistics and counters are cleared, so the previous path operational bandwidth is lost. That is, the operational BW of a path is not persistent across LSP down events. In this case, there is no immediate bandwidth adjustment on the secondary path.

The following algorithm is used to determine the signaled bandwidth on a newly active path:

1. For a path that is operationally down, signaled_bw = config_bw.

2. For the active path, if an auto-BW MBB adjustment is in progress, signaled_bw = previous path operational BW for the first five attempts. For the remaining attempts, the signaled BW = operational BW.
3. For an MBB on the active path (other than an auto-BW MBB), MBB signaled BW = operational BW.
4. For an MBB on the inactive path, MBB signaled BW = configured BW.

If the primary path is not the currently active path and it has not gone down, then any MBB uses the configured BW for the primary path. However, if the configured BW is changed for a path that is currently not active, then a config change MBB is not triggered.

If the standby is SRLG enabled, and the active path is the standby, and the primary comes up, this will immediately trigger a delayed retry MBB on the standby. If the delayed retry MBB fails, immediate reversion to the primary occurs regardless of the retry timer.

When the system reverts from a secondary standby or secondary path to the primary path, a Delayed Retry MBB is attempted to bring bandwidth of the standby path back to its configured bandwidth. Delayed Retry MBB is attempted once, and if it fails, the standby is torn down. A Delayed Retry MBB has highest priority among all MBBs, so it will take precedence over any other MBB in progress on the standby path (for example, Config change or Preemption).

The system will carry-over the last signaled BW of the LSP over multiple failovers. For example, if an LSP is configured with auto-BW for some time, and adjusts its currently reserved bandwidth for the primary, and Monitor mode is then enabled, BW adjustment on the primary ceases, but the BW remains reserved at the last adjusted value. Next, the LSP fails over to a secondary or secondary standby. The secondary will inherit the last reserved BW of the primary, but then disable further adjustment as long as monitoring mode is enabled.

The system's ability to carry-over the last signaled BW across failovers has the following limitations:

- Case 1: If the LSP fails over from path1 to path2 and the AutoBW MBB on path2 is successful, the last signaled BW is carried over when the LSP reverts back to path1 or fails over to a new path3. This may trigger an AutoBW MBB on the new active path to adjust its bandwidth to last signaled BW.
- Case 2: If the LSP fails over from path1 to path2 and the AutoBW MBB on path2 is still in progress and the LSP reverts back to path1 or fails over to a new path3, the last signaled BW is carried over to the new active path (path1 or path3) and this may result in an AutoBW MBB on that path.

- Case 3: If the LSP fails over from path1 to path2 and the AutoBW MBB on path2 fails (after 5 retry attempts), the last signaled BW from when path1 was active is lost. Therefore, when the LSP reverts back to path1 or fails over to a new path3, the original signaled BW from path1 is not carried over. However the signaled bandwidth of path2 is carried over to the new active path (path1 or path3) and may trigger an AutoBW on that path.

2.1.8.3 Measurement of LSP Bandwidth

Automatic adjustment of RSVP LSP bandwidth based on measured traffic rate into the tunnel requires the LSP to be configured for egress statistics collection at the ingress LER. The following CLI shows an example:

```
config router mpls lsp name
      egress-statistics
      accounting-policy 99
      collect-stats
      no shutdown
exit
```

All LSPs configured for accounting, including any configured for auto-bandwidth based on traffic measurements, must reference the same accounting policy. An example configuration of such an accounting-policy is shown below: in the CLI example below.

```
config log
     accounting-policy 99
     collection-interval 5
         record combined-mpls-lsp-egress
     exit
exit
```

The record **combined-mpls-lsp-egress** command in the accounting policy has the effect of recording both egress packet and byte counts and bandwidth measurements based on the byte counts if auto-bandwidth is enabled on the LSP.

When egress statistics are enabled the CPM collects stats from of all XCMs or IOMs involved in forwarding traffic belonging to the LSP (whether the traffic is currently leaving the ingress LER via the primary LSP path, a secondary LSP path, an FRR detour path or an FRR 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 vs. out-of-profile) basis. When auto-bandwidth is configured for an LSP the ingress LER calculates a traffic rate for the LSP as follows:

Average data rate of LSP[x] during interval[i] = F(x, i)—F(x, i-1)/sample interval

F(x, i) — The total number of bytes belonging to LSP[x], regardless of forwarding-class or priority, at time[i]

sample interval = time[i] — time [i-1], time[i+1] — time[i], and so on.

The sample interval is the product of sample-multiplier and the collection-interval specified in the auto-bandwidth accounting policy. A default sample-multiplier for all LSPs may be configured using the **config>router>mpls>auto-bandwidth-defaults** command but this value can be overridden on a per-LSP basis at the **config>router>mpls>lsp>auto-bandwidth** context. The default value of sample-multiplier (the value that would result from the no auto-bandwidth-defaults command) is 1, which means the default sample interval is 300 seconds.

Over a longer period of time called the adjust interval the router keeps track of the maximum average data rate recorded during any constituent sample interval. The adjust interval is the product of adjust-multiplier and the collection-interval specified in the auto-bandwidth accounting-policy. A default adjust-multiplier for all LSPs may be configured using the **config>router>mpls>auto-bandwidth-multiplier** command but this value can be overridden on a per-LSP basis at the **config>router>mpls>lsp>auto-bandwidth** context. The default value of adjust-multiplier (the value that would result from the no auto-bandwidth-multiplier command) is 288, which means the default adjust interval is 86400 seconds or 24 hours. The system enforces the restriction that adjust-multiplier is equal to or greater than sample-multiplier. It is recommended that the adjust-multiplier be an integer multiple of the sample-multiplier.

The collection-interval in the auto-bandwidth accounting policy can be changed at any time, without disabling any of the LSPs that rely on that policy for statistics collection.

The sample-multiplier (at the **mpls>auto-bandwidth** level or the **lsp>auto-bandwidth** level) can be changed at any time. This will have no effect until the beginning of the next sample interval. In this case the adjust-interval does not change and information about the current adjust interval (such as the remaining adjust-multiplier, the maximum average data rate) is not lost when the sample-multiplier change takes effect.

The system allows adjust-multiplier (at the **mpls** level or the **lsp>auto-bandwidth** level) to be changed at any time as well but in this case the new value shall have no effect until the beginning of the next adjust interval.

Byte counts collected for LSP statistics include layer 2 encapsulation (Ethernet headers and trailers) and therefore average data rates measured by this feature include Layer 2 overhead as well.

2.1.8.4 Passive Monitoring of LSP Bandwidth

The system offers the option to measure the bandwidth of an RSVP LSP (see Measurement of LSP Bandwidth) without taking any action to adjust the bandwidth reservation, regardless of how different the measured bandwidth is from the current reservation. Passive monitoring is enabled using the **config>router>mpls>lsp>auto-bandwidth>monitor-bandwidth** command.

The **show>router>mpls>lsp detail** command can be used to view the maximum average data rate in the current adjust interval and the remaining time in the current adjust interval.

2.1.8.5 Periodic Automatic Bandwidth Adjustment

Automatic bandwidth allocation is supported on any RSVP LSP that has MBB enabled. MBB is enabled in the **config**>**router**>**mpls**>**lsp** context using the **adaptive** command. If the **monitor-bandwidth** command is enabled in **config**>**router**>**mpls**>**lsp**>**auto-bandwidth** context, the LSP is not resignaled to adjust its bandwidth to the calculated values.

If an eligible RSVP LSP is configured for auto-bandwidth, by entering auto-bandwidth at the **config**>**router**>**mpls**>**lsp** context, then the ingress LER decides every adjust interval whether to attempt auto-bandwidth adjustment. The following parameters are defined:

- current_bw — The currently reserved bandwidth of the LSP; this is the operational bandwidth that is already maintained in the MIB.
- measured_bw — The maximum average data rate in the current adjust interval.
- signaled_bw — The bandwidth that is provided to the CSPF algorithm and signaled in the SENDER_TSPEC and FLOWSPEC objects when an auto-bandwidth adjustment is attempted.
- min — The configured min-bandwidth of the LSP.
- max — The configured max-bandwidth of the LSP.
- up% — The minimum difference between measured_bw and current_bw, expressed as a percentage of current_bw, for increasing the bandwidth of the LSP.
- up — The minimum difference between measured_bw and current_bw, expressed as an absolute bandwidth relative to current_bw, for increasing the bandwidth of the LSP. This is an optional parameter; if not defined the value is 0.

- down% — The minimum difference between current_bw and measured_bw, expressed as a percentage of current_bw, for decreasing the bandwidth of the LSP.
- down — The minimum difference between current_bw and measured_bw, expressed as an absolute bandwidth relative to current_bw, for decreasing the bandwidth of the LSP. This is an optional parameter; if not defined the value is 0.

At the end of every adjust interval the system decides if an auto-bandwidth adjustment should be attempted. The heuristics are as follows:

- If the measured bandwidth exceeds the current bandwidth by more than the percentage threshold and also by more than the absolute threshold then the bandwidth is re-signaled to the measured bandwidth (subject to min and max constraints).
- If the measured bandwidth is less than the current bandwidth by more than the percentage threshold and also by more than the absolute threshold then the bandwidth is re-signaled to the measured bandwidth (subject to min and max constraints).
- If the current bandwidth is greater than the max bandwidth then the LSP bandwidth is re-signaled to max bandwidth, even if the thresholds have not been triggered.
- If the current bandwidth is less than the min bandwidth then the LSP bandwidth is re-signaled to min bandwidth, even if the thresholds have not been triggered.

Changes to min-bandwidth, max-bandwidth and any of the threshold values (up, up%, down, down%) are permitted at any time on an operational LSP but the changes have no effect until the next auto-bandwidth trigger (for example, adjust interval expiry).

If the measured bandwidth exceeds the current bandwidth by more than the percentage threshold and also by more than the absolute threshold then the bandwidth is re-signaled to the measured bandwidth (subject to min and max constraints).

The adjust-interval and maximum average data rate are reset whether the adjustment succeeds or fails. If the bandwidth adjustment fails (for example, CSPF cannot find a path) then the existing LSP is maintained with its existing bandwidth reservation. The system does not retry the bandwidth adjustment (for example, per the configuration of the LSP retry-timer and retry-limit).

2.1.8.6 Overflow-Triggered Auto-Bandwidth Adjustment

For cases where the measured bandwidth of an LSP has increased significantly since the start of the current adjust interval it may be desirable for the system to preemptively adjust the bandwidth of the LSP and not wait until the end of the adjust interval.

The following parameters are defined:

- current_bw — The currently reserved bandwidth of the LSP.
- sampled_bw — The average data rate of the sample interval that just ended.
- measured_bw — The maximum average data rate in the current adjust interval.
- signaled_bw — The bandwidth that is provided to the CSPF algorithm and signaled in the SENDER_TSPEC and FLOWSPEC objects when an auto-bandwidth adjustment is attempted.
- max — The configured max-bandwidth of the LSP.
- %_threshold — The minimum difference between sampled_bw and current_bw, expressed as a percentage of the current_bw, for counting an overflow event.
- min_threshold — The minimum difference between sampled_bw and current_bw, expressed as an absolute bandwidth relative to current_bw, for counting an overflow event. This is an optional parameter; if not defined the value is 0.

When a sample interval ends it is counted as an overflow if:

- The sampled bandwidth exceeds the current bandwidth by more than the percentage threshold and by more than the absolute bandwidth threshold (if defined).
- When the number of overflow samples reaches a configured limit, an immediate attempt is made to adjust the bandwidth to the measured bandwidth (subject to the min and max constraints).

If the bandwidth adjustment is successful then the adjust-interval, maximum average data rate and overflow count are all reset. If the bandwidth adjustment fails then the overflow count is reset but the adjust-interval and maximum average data rate continue with current values. It is possible that the overflow count will once again reach the configured limit before the end of adjust-interval is reached and this will once again trigger an immediate auto-bandwidth adjustment attempt.

The overflow configuration command fails if the max-bandwidth of the LSP has not been defined.

The threshold limit can be changed on an operational auto-bandwidth LSP at any time and the change should take effect at the end of the current sample interval (for example, if the user decreases the overflow limit to a value lower than the current overflow count then auto-bandwidth adjustment will take place as soon as the sample interval ends). The threshold values can also be changed at any time (for example, %_threshold and min_threshold) but the new values will not take effect until the end of the current sample interval.

2.1.8.7 Manually-Triggered Auto-Bandwidth Adjustment

Manually-triggered auto-bandwidth adjustment feature is configured with the **tools>perform>router>mpls adjust-autobandwidth** [**lsp** *lsp-name* [**force** [**bandwidth** *mbps*]]] command to attempt immediate auto-bandwidth adjustment for either one specific LSP or all active LSPs. If the LSP is not specified then the system assumes the command applies to all LSPs. If an LSP name is provided then the command applies to that specific LSP only and the optional **force** parameter (with or without a bandwidth) can be used.

If **force** is not specified (or the command is not LSP-specific) then measured_bw is compared to current_bw and bandwidth adjustment may or may not occur

If **force** is specified and a bandwidth is not provided then the threshold checking is bypassed but the min and max bandwidth constraints are still enforced.

If **force** is specified with a bandwidth (in Mb/s) then signaled_bw is set to this bandwidth. There is no requirement that the bandwidth entered as part of the command fall within the range of min-bandwidth to max-bandwidth.

The adjust-interval, maximum average data rate and overflow count are not reset by the manual auto-bandwidth command, whether or not the bandwidth adjustment succeeds or fails. The overflow count is reset only if the manual auto-bandwidth adjustment is successful.

2.1.8.8 Operational Bandwidth Carryover between Active Paths

SR OS supports carrying over of the operational bandwidth (for example, the last successfully signaled bandwidth) of an LSP path to the next active path following a switchover. The new active path can be a secondary or a primary path. The bandwidth is not lost even when the previously active path fails. The last successfully signaled bandwidth is known as the last adjusted bandwidth.

This feature is enabled using the **configure router mpls lsp auto-bandwidth use-last-adj-bw** command.

When enabled, secondary paths are initially signaled with the last adjusted bandwidth of the primary, and not the configured bandwidth. If signaling a secondary at this bandwidth fails after some number of retries, then the path fails rather than falling back to using the configured bandwidth. The number of retries of secondary paths at the last adjusted bandwidth is configured using the **secondary-retry-limit** command under **use-last-adj-bw**.

A shutdown of the primary or any configuration change events that cause a switch to a secondary, uses the last adjusted bandwidth. The user can toggle **use-last-adj-bw** at any time; this does not require an administrative shutdown of auto bandwidth, however, the new value is not used until the next path switchover.

Note: The last adjusted bandwidth value is reset on a shutdown of MPLS, the LSP, or autobandwidth.

If the revert timer is enabled, the primary is re-signaled before the revert timer expires with its configured bandwidth. An auto-bandwidth MBB using the last adjusted bandwidth of the secondary occurs immediately on switching back when the revert timer expires. If the system switches to a new path while an auto-bandwidth MBB is in progress on the previously active path, then the bandwidth used to signal the new path is the new value that was being attempted on the old path (rather than the last adjusted bandwidth). This means that the new path establishes with the most up to date bandwidth for the LSP (provided sufficient network resources are available) rather than a potentially out of date bandwidth.

2.1.9 LSP Failure Codes

Table 7 lists the MPLS LSP path failure codes and their meanings. These failure codes are indicated in the FailureCode output field of the **show router mpls lsp path detail** command, as well as in the TIMETRA MPLS MIB.

Table 7 **LSP Failure Codes**

LSP Failure Code (Value)	Meaning
noError (0)	Indicates no errors for this LSP.

Table 7 **LSP Failure Codes (Continued)**

LSP Failure Code (Value)	Meaning
admissionControlError (1)	An RSVP admission control failure occurred at some point along the path of an LSP. This is recorded as a result of a PathErr message.
noRouteToDestination (2)	No route could be found toward the requested destination.
trafficControlSystemError (3)	An error in the traffic control system due to an unsupported traffic parameter, for example a bad FLOWSPEC, TSPEC or ADSPEC value.
routingError (4)	There is a problem with the route defined for the LSP, for example the ERO is truncated.
noResourcesAvailable (5)	Insufficient system or protocol resources are available to complete the request, for example, out of memory or out of resources such as NHLFE indexes or labels. This error code is also used for RSVP packet decode failures such as. bad object length or unknown sub-object.
badNode (6)	Indicates a bad node in the path hop list at head-end or ERO at transit.
routingLoop (7)	A routing loop was detected for the LSP path.
labelAllocationError (8)	Unable to allocate a label for the LSP path.
badL3PID (9)	The router has received a PathErr with the error code “Routing problem” and the error value “Unsupported L3PID.” This indicates that a downstream LSR does not support the protocol type “L3PID”.
tunnelLocallyRepaired (10)	A PLR has triggered a local repair at some point along the path of the LSP.
unknownObjectClass (11)	A downstream LSR rejected an RSVP message because it contained an Unknown object class - Error code 13 as defined in RFC 2205, *Resource ReSerVation Protocol (RSVP) - Version 1 Functional Specification*.
unknownCType (12)	A downstream LSR rejected an RSVP message due to an Unknown object C-type - Error code 14 as defined in RFC 2205.
noEgressMplsInterface (13)	An egress MPLS interface could not be found for the LSP path.
noEgressRsvpInterface (14)	An egress RSVP interface could not be found for the LSP path.

Table 7 **LSP Failure Codes (Continued)**

LSP Failure Code (Value)	Meaning
looseHopsInFRRLsp (15)	The path calculated for the FRR enabled LSP contains loose hops.
unknown (16)	Indicates an error not covered by one of the other known errors for this LSP.
retryExceeded (17)	The retry limit for the LSP path has been exceeded.
noCspfRouteOwner (18)	No IGP instance was found that has a route to the LSP destination.
noCspfRouteToDestination (19)	CSPF was unable to find a route to the requested destination that satisfies all of the constraints.
hopLimitExceeded (20)	The hop-limit for the LSP path has been exceeded.
looseHopsInManualBypassLsp (21)	A manual bypass LSP contains loose hops.
emptyPathInManualBypassLsp (22)	A manual bypass LSP uses an empty path.
lspFlowControlled (23)	The router initiated flow control for path messages for paths that have not yet been established.
srlgSecondaryNotDisjoint (24)	The secondary path is not SRLG disjoint from the primary path.
srlgPrimaryCspfDisabled (25)	An SRLG disjoint path could not be found for the secondary because CSPF is disabled on the primary.
srlgPrimaryPathDown (26)	An SRLG disjoint path could not be found for the secondary because the primary is down.
localLinkMaintenance (27)	A TE link (RSVP interface) local to this LSR or on a remote LSR used by the LSP is in TE graceful shutdown. The link that has been gracefully shutdown is also identified.
unexpectedCtObject (28)	A downstream LSR does not recognize something about the content of the diffserv class type object.
unsupportedCt (29)	A downstream LSR does not support the signaled Diffserv class type.
invalidCt (30)	Indicates the signaled diffserv class type is invalid, for example it is 0.
invCtAndSetupPri (31)	The combination of signaled Diffserv class type and setup priority does not map to a valid Diffserv TE class.

Table 7 **LSP Failure Codes (Continued)**

LSP Failure Code (Value)	Meaning
invCtAndHoldPri (32)	The combination of signaled diffserv class type and hold priority does not map to a valid Diffserv TE class.
invCtAndSetupAndHoldPri (33)	The combination of signaled Diffserv class type and setup priority and hold priority does not map to a valid Diffserv TE class.
localNodeMaintenance (34)	The local LSR or a remote LSR used by the LSP is in TE graceful shutdown due to maintenance The LSR that s shutdown is also identified.
softPreemption (35)	The LSP path is under soft pre-emption.
p2mpNotSupported (36)	An LSR does not support P2MP LSPs.
badXro (37)	An LSR for the LSP encountered a badly formed exclude route object, for example a sub-object is missing or unrecognized.
localNodeInXro (38)	The Exclude Route Object includes the local node.
routeBlockedByXro (39)	The Exclude Route Object prevents the LSP path from being established at all.
xroTooComplex (40)	The Exclude Route Object contains too many entries or is too complex to calculate a path. If an SR OS router receives an XRO with more than 5 sub-objects then it will be rejected.
rsvpNotSupported (41)	Maps to SubErrorCode 8 for ErrorCode 24 (Routing error) from RFC 3209. An LSR will send ErrorCode=24, SubErrorCode=8 when it receives PATH for P2MP LSP but P2MP is not supported on that router.
conflictingAdminGroups (42)	The specified admin groups contradict for example the same group is both included and excluded.
nodeInIgpOverload (43)	An LSR along the path of the LSP has advertised the ISIS overload state.
srTunnelDown (44)	An SR tunnel is admin or operationally down.
fibAddFailed (45)	An LSP path could not be added to the FIB for example if IOM programming fails for an SR-TE tunnel.
labelStackExceeded (46)	The label stack depth for an SR-TE LSP exceeds the max-sr-labels.
pccDown (47)	The PCC or the PCEP channel to the PCC is down.

Table 7 **LSP Failure Codes (Continued)**

LSP Failure Code (Value)	Meaning
pccError (48)	An error has been received from the PCC related to this LSP. Such errors relate to processing requests, message objects, or TLVs.
pceDown (49)	The Path Computation Element or PCEP channel is down.
pceError (50)	An error has been received from the PCE related to this LSP. Such errors relate to processing requests, message objects, or TLVs.
pceUpdateWithEmptyEro (51)	MPLS received an update from PCE with an empty ERO.
pceInitLspDisabled (52)	The related **config>router>mpls>pce-initiated-lsp** context for this LSP type is disabled.
adminDown (53)	A related MPLS path is disabled.
srlgPathWithSidHops (59)	Configuration conflicts with the use of a path with hops consisting of SID labels.

2.1.10 Labeled Traffic Statistics

SR OS provides a wide range of capabilities for collecting statistics of labeled traffic. This section provides an overview of these capabilities.

2.1.10.1 Interface Statistics

By default, the system continuously collects statistics (packet and octet counts) of MPLS traffic on ingress and egress of MPLS interfaces. These statistics can be accessed, for example, using the **show**>**router**>**mpls**>**interface statistics** command.

In addition, the system can provide auxiliary statistics (packet and octet counts) for a specific type of labeled traffic on ingress and egress of MPLS interfaces. The **config**>**router**>**mpls**>**aux-stats** command accesses these statistics and also specifies which types of labeled traffic should be counted. The **sr** keyword refers to any type of MPLS-SR traffic (such as SR-OSPF, SR-ISIS, SR-TE). After being enabled and configured, auxiliary statistics can be viewed, monitored, and cleared. The two types of statistics (global or default MPLS statistics and auxiliary statistics) are independent; clearing one counter does not affect the values of the other counter.

For both types of statistics, implicit null on ingress is not regarded as labeled traffic and octet counts include L2 headers and trailers.

Segment Routing traffic statistics have a dependency with the ability to account for dark bandwidth in IGP-TE advertisements.

2.1.10.2 Traffic Statistics for Stacked Tunnels

The nature of MPLS allows for LSPs, owned by a given protocol, to be tunneled into an LSP that is owned by another protocol. Typical examples of this capability are LDP over RSVP-TE, SR over RSVP-TE, and LDP over SR-TE. Also, in a variety of constructs (SR-TE LSPs, SR Policies) SR OS uses hierarchical NHLFEs where a single (top) NHLFE that models the forwarding actions towards the next hop, can be referenced by one or more (inner) NHLFEs that model the forwarding actions for the rest of the end-to-end path.

SR OS enables collecting the traffic statistics from the majority of all supported types of tunnels. In cases where statistics collection is enabled on multiple labels of the stack, SR OS provides the capability to collect traffic statistics on two labels of the MPLS stack. Any label needs to be processed (as part of ILM or NHLFE processing) for statistics to be collected. For example, a node acting as an LSR for an RSVP-TE LSP (that transports an LDP LSP) can collect statistics for the RSVP-TE LSP but does not collect stats for the LDP LSP. A node acting as an LER for that same RSVP-TE LSP is, however, able to collect statistics for the LDP LSP.

To control whether statistics are collected on one or two labels, use the following command:

configure>**system**>**ip**>**mpls**>**label-stack-statistics-count** *label-stack-id*

This command does not enable statistics collection. It only controls on how many labels, out of those that have statistics collection enabled, statistics collection is effectively performed.

If the MPLS label stack represents more than two stacked tunnels, the system collects statistics on the outermost (top) label for which statistics collection is enabled (if above value is 1 or 2), and collects statistics on the innermost (bottom) label for which statistics collection is enabled (if above value is 2).

2.1.10.3 Traffic Statistics Details and Scale

For RSVP-TE and LDP, statistics are provided per forwarding class and as “**in-profile**” or “**out-of-profile**”. For all other labeled constructs, statistics are provided regardless of the forwarding class and the QoS profile. Altogether, labeled constructs share 128k statistic indices (on ingress and on egress independently). Statistics with FC and QoS profile consume 16 indices.

2.1.10.4 RSVP-TE and MPLS-TP Traffic Statistics

See RSVP-TE LSP Statistics and P2MP RSVP-TE LSP Statistics for information about RSVP-TE and MPLS-TP traffic statistics.

2.1.10.5 MPLS Forwarding Policy Statistics

See Statistics for more information about MPLS forwarding policy statistics.

2.1.10.6 gRPC-based RIB API Statistics

Refer to “Traffic Statistics” in the *7450 ESS, 7750 SR, 7950 XRS, and VSR Router Configuration Guide* for more information about gRPC-based RIB API statistics.

2.1.10.7 Segment Routing Statistics

Refer to “Segment Routing Traffic Statistics” in the *7450 ESS, 7750 SR, 7950 XRS, and VSR Unicast Routing Protocols Guide* for more information about segment routing statistics.

2.1.10.8 SR-TE LSP Statistics

See SR-TE LSP Traffic Statistics for more information about SR-TE LSP statistics.

2.1.10.9 SR Policy Statistics

See Statically-Configured Segment Routing Policies for more information about SR policy statistics.

2.2 RSVP

The Resource Reservation Protocol (RSVP) is a network control protocol used by a host to request specific qualities of service from the network for particular application data streams or flows. RSVP is also used by routers to deliver quality of service (QoS) requests to all nodes along the path(s) of the flows and to establish and maintain state to provide the requested service. RSVP requests generally result in resources reserved in each node along the data path. MPLS leverages this RSVP mechanism to set up traffic engineered LSPs. RSVP is not enabled by default and must be explicitly enabled.

RSVP requests resources for simplex flows. It requests resources only in one direction (unidirectional). Therefore, RSVP treats a sender as logically distinct from a receiver, although the same application process may act as both a sender and a receiver at the same time. Duplex flows require two LSPs, to carry traffic in each direction.

RSVP is not a routing protocol. RSVP operates with unicast and multicast routing protocols. Routing protocols determine where packets are forwarded. RSVP consults local routing tables to relay RSVP messages.

RSVP uses two message types to set up LSPs, PATH and RESV. Figure 5 depicts the process to establish an LSP.

- The sender (the ingress LER (ILER)), sends PATH messages toward the receiver, (the egress LER (eLER)) to indicate the FEC for which label bindings are desired. PATH messages are used to signal and request label bindings required to establish the LSP from ingress to egress. Each router along the path observes the traffic type.

  PATH messages facilitate the routers along the path to make the necessary bandwidth reservations and distribute the label binding to the router upstream.
- The eLER sends label binding information in the RESV messages in response to PATH messages received.
- The LSP is considered operational when the ILER receives the label binding information.

Figure 5 **Establishing LSPs**

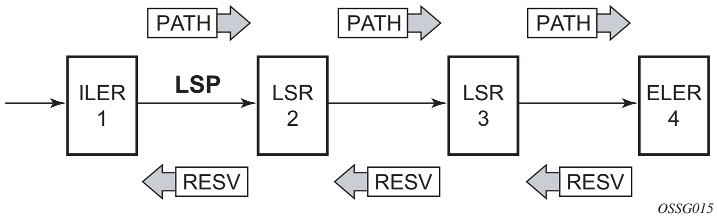


Figure 6 **LSP Using RSVP Path Set Up**

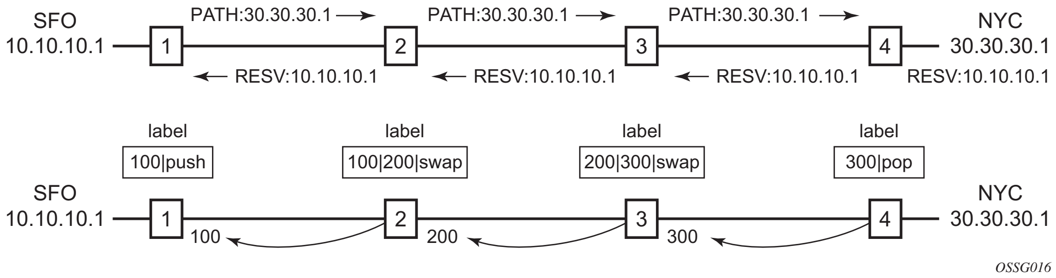


Figure 6 displays an example of an LSP path set up using RSVP. The ingress label edge router (ILER 1) transmits an RSVP path message (path: 30.30.30.1) downstream to the egress label edge router (eLER 4). The path message contains a label request object that requests intermediate LSRs and the eLER to provide a label binding for this path.

In addition to the label request object, an RSVP PATH message can also contain a number of optional objects:

- Explicit route object (ERO) — When the ERO is present, the RSVP path message is forced to follow the path specified by the ERO (independent of the IGP shortest path).
- Record route object (RRO) — Allows the ILER to receive a listing of the LSRs that the LSP tunnel actually traverses.
- A session attribute object controls the path set up priority, holding priority, and local-rerouting features.

Upon receiving a path message containing a label request object, the eLER transmits a RESV message that contains a label object. The label object contains the label binding that the downstream LSR communicates to its upstream neighbor. The RESV message is sent upstream towards the ILER, in a direction opposite to that followed by the path message. Each LSR that processes the RESV message carrying a label object uses the received label for outgoing traffic associated with the specific LSP. When the RESV message arrives at the ingress LSR, the LSP is established.

2.2.1 Using RSVP for MPLS

Hosts and routers that support both MPLS and RSVP can associate labels with RSVP flows. When MPLS and RSVP are combined, the definition of a flow can be made more flexible. Once an LSP is established, the traffic through the path is defined by the label applied at the ingress node of the LSP. The mapping of label to traffic can be accomplished using a variety of criteria. The set of packets that are assigned the same label value by a specific node are considered to belong to the same FEC which defines the RSVP flow.

For use with MPLS, RSVP already has the resource reservation component built-in which makes it ideal to reserve resources for LSPs.

2.2.1.1 RSVP Traffic Engineering Extensions for MPLS

RSVP has been extended for MPLS to support automatic signaling of LSPs. To enhance the scalability, latency, and reliability of RSVP signaling, several extensions have been defined. Refresh messages are still transmitted but the volume of traffic, the amount of CPU utilization, and response latency are reduced while reliability is supported. None of these extensions result in backward compatibility problems with traditional RSVP implementations.

2.2.1.2 Hello Protocol

The Hello protocol detects the loss of a neighbor node or the reset of a neighbor's RSVP state information. In standard RSVP, neighbor monitoring occurs as part of RSVP's soft-state model. The reservation state is maintained as cached information that is first installed and then periodically refreshed by the ingress and egress LSRs. If the state is not refreshed within a specified time interval, the LSR discards the state because it assumes that either the neighbor node has been lost or its RSVP state information has been reset.

The Hello protocol extension is composed of a hello message, a hello request object and a hello ACK object. Hello processing between two neighbors supports independent selection of failure detection intervals. Each neighbor can automatically issue hello request objects. Each hello request object is answered by a hello ACK object.

2.2.1.3 MD5 Authentication of RSVP Interface

When enabled on an RSVP interface, authentication of RSVP messages operates in both directions of the interface.

A node maintains a security association with its neighbors for each authentication key. The following items are stored in the context of this security association:

- The HMAC-MD5 authentication algorithm.
- Key used with the authentication algorithm.
- Lifetime of the key. A key is user-generated key using a third party software/ hardware and enters the value as static string into CLI configuration of the RSVP interface. The key will continue to be valid until it is removed from that RSVP interface.
- Source Address of the sending system.
- Latest sending sequence number used with this key identifier.

The RSVP sender transmits an authenticating digest of the RSVP message, computed using the shared authentication key and a keyed-hash algorithm. The message digest is included in an Integrity object which also contains a Flags field, a Key Identifier field, and a Sequence Number field. The RSVP sender complies to the procedures for RSVP message generation in RFC 2747, *RSVP Cryptographic Authentication*.

An RSVP receiver uses the key together with the authentication algorithm to process received RSVP messages.

When a PLR node switches the path of the LSP to a bypass LSP, it does not send the Integrity object in the RSVP messages over the bypass tunnel. If an integrity object is received from the MP node, then the message is discarded since there is no security association with the next-next-hop MP node.

The MD5 implementation does not support the authentication challenge procedures in RFC 2747.

2.2.1.4 Configuring Authentication using Keychains

The use of authentication mechanism is recommended to protect against malicious attack on the communications between routing protocol neighbors. These attacks could aim to either disrupt communications or to inject incorrect routing information into the systems routing table. The use of authentication keys can help to protect the routing protocols from these types of attacks.

Within RSVP, authentication must be explicitly configured through the use of the authentication keychain mechanism. This mechanism allows for the configuration of authentication keys and allows the keys to be changed without affecting the state of the protocol adjacencies.

To configure the use of an authentication keychain within RSVP, use the following steps:

1. Configure an authentication keychain within the **config>system>security** context. The configured keychain must include at least on valid key entry, using a valid authentication algorithm for the RSVP protocol.
2. Associate the configure authentication keychain with RSVP at the interface level of the CLI, this is done with the **auth-keychain** *name* command.

For a key entry to be valid, it must include a valid key, the current system clock value must be within the begin and end time of the key entry, and the algorithm specified in the key entry must be supported by the RSVP protocol.

The RSVP protocol supports the following algorithms:

- clear text password
- HMAC-MD5
- HMC-SHA-1

Error handling:

- If a keychain exists but there are no active key entries with an authentication type that is valid for the associated protocol then inbound protocol packets will not be authenticated and discarded, and no outbound protocol packets should be sent.
- If keychain exists but the last key entry has expired, a log entry is raised indicating that all keychain entries have expired. The RSVP protocol requires that the protocol not revert to an unauthenticated state and requires that the old key is not to be used, therefore, once the last key has expired, all traffic is discarded.

2.2.2 Reservation Styles

LSPs can be signaled with explicit reservation styles. A reservation style is a set of control options that specify a number of supported parameters. The style information is part of the LSP configuration. SR OS supports two reservation styles:

- Fixed Filter (FF) — The Fixed Filter (FF) reservation style specifies an explicit list of senders and a distinct reservation for each of them. Each sender has a dedicated reservation that is not shared with other senders. Each sender is identified by an IP address and a local identification number, the LSP ID. Because each sender has its own reservation, a unique label and a separate LSP can be constructed for each sender-receiver pair. For traditional RSVP applications, the FF reservation style is ideal for a video distribution application in which each channel (or source) requires a separate pipe for each of the individual video streams.
- Shared Explicit (SE) — The Shared Explicit (SE) reservation style creates a single reservation over a link that is shared by an explicit list of senders. Because each sender is explicitly listed in the RESV message, different labels can be assigned to different sender-receiver pairs, thereby creating separate LSPs.

  If FRR option is enabled for the LSP and selects the facility FRR method at the head-end node, only the SE reservation style is allowed. Furthermore, if a PLR node receives a path message with fast-reroute requested with facility method and the FF reservation style, it will reject the reservation. The one-to-one detour method supports both FF and SE styles.

2.2.2.1 RSVP Message Pacing

When a flood of signaling messages arrive because of topology changes in the network, signaling messages can be dropped which results in longer set up times for LSPs. RSVP message pacing controls the transmission rate for RSVP messages, allowing the messages to be sent in timed intervals. Pacing reduces the number of dropped messages that can occur from bursts of signaling messages in large networks.

2.2.3 RSVP Overhead Refresh Reduction

The RSVP refresh reduction feature consists of the following capabilities implemented in accordance to RFC 2961, *RSVP Refresh Overhead Reduction Extensions*:

- RSVP message bundling — This capability is intended to reduce overall message handling load. The system supports receipt and processing of bundled message only, but no transmission of bundled messages.

- Reliable message delivery — This capability consists of sending a message-id and returning a message-ack for each RSVP message. It can be used to detect message loss and support reliable RSVP message delivery on a per hop basis. It also helps reduce the refresh rate since the delivery becomes more reliable.
- Summary refresh — This capability consists of refreshing multiples states with a single message-id list and sending negative ACKs (NACKs) for a message_id which could not be matched. The summary refresh capability reduce the amount of messaging exchanged and the corresponding message processing between peers. It does not however reduce the amount of soft state to be stored in the node.

These capabilities can be enabled on a per-RSVP-interface basis are referred to collectively as “refresh overhead reduction extensions”. When the refresh-reduction is enabled on a system RSVP interface, the node indicates this to its peer by setting a refresh-reduction- capable bit in the flags field of the common RSVP header. If both peers of an RSVP interface set this bit, all the above three capabilities can be used. Furthermore, the node monitors the settings of this bit in received RSVP messages from the peer on the interface. As soon as this bit is cleared, the node stops sending summary refresh messages. If a peer did not set the “refresh-reduction-capable” bit, a node does not attempt to send summary refresh messages.

The RSVP Overhead Refresh Reduction is supported with both RSVP P2P LSP path and the S2L path of an RSVP P2MP LSP instance over the same RSVP interface.

2.2.4 RSVP Graceful Restart Helper

The **gr-helper** command enables the RSVP Graceful Restart Helper feature.

The RSVP-TE Graceful Restart helper mode allows the SR OS based system (the helper node) to provide another router that has requested it (the restarting node) a grace period, during which the system will continue to use RSVP sessions to neighbors requesting the grace period. This is typically used when another router is rebooting its control plane but its forwarding plane is expected to continue to forward traffic based on the previously available Path and Resv states.

The user can enable Graceful Restart helper on each RSVP interface separately. When the GR helper feature is enabled on an RSVP interface, the node starts inserting a new Restart_Cap Object in the Hello packets to its neighbor. The restarting node does the same and indicates to the helper node the desired Restart Time and Recovery Time.

The GR Restart helper consists of a couple of phases. Once it loses Hello communication with its neighbor, the helper node enters the Restart phase. During this phase, it preserves the state of all RSVP sessions to its neighbor and waits for a new Hello message.

Once the Hello message is received indicating the restarting node preserved state, the helper node enters the recovery phase in which it starts refreshing all the sessions that were preserved. The restarting node will activate all the stale sessions that are refreshed by the helper node. Any Path state that did not get a Resv message from the restarting node once the Recovery Phase time is over is considered to have expired and is deleted by the helper node causing the proper Path Tear generation downstream.

The duration of the restart phase (recovery phase) is equal to the minimum of the neighbor's advertised Restart Time (Recovery Time) in its last Hello message and the locally configured value of the max-restart (max-recovery) parameter.

When GR helper is enabled on an RSVP interface, its procedures apply to the state of both P2P and P2MP RSVP LSP to a neighbor over this interface.

2.2.5 Enhancements to RSVP Control Plane Congestion Control

The RSVP control plane makes use of a global flow control mechanism to adjust the rate of Path messages for unmapped LSP paths sent to the network under congestion conditions. When a Path message for establishing a new LSP path or retrying an LSP path that failed is sent out, the control plane keeps track of the rate of successful establishment of these paths and adjusts the number of Path messages it sends per second to reflect the success ratio.

In addition, an option to enable an exponential back-off retry-timer is available. When an LSP path establishment attempt fails, the path is put into retry procedures and a new attempt is performed at the expiry of the user-configurable retry-timer. By default, the retry time is constant. The exponential back-off timer procedures will double the value of the user configurable retry-timer value at every failure of the attempt to adjust to the potential network congestion that caused the failure. An LSP establishment fails if no Resv message was received and the Path message retry-timer expired, or a PathErr message was received before the timer expired.

Three enhancements to this flow-control mechanism to improve congestion handling in the rest of the network are supported.

The first enhancement is the change to the LSP path retry procedure. If the establishment attempt failed due to a Path message timeout and no Resv was received, the next attempt is performed at the expiry of a new LSP path initial retry-timer instead of the existing retry-timer. While the LSP path initial retry-timer is still running, a refresh of the Path message using the same path and the same LSP-id is performed according to the configuration of the refresh-timer. Once the LSP path initial retry-timer expires, the ingress LER then puts this path on the regular retry-timer to schedule the next path signaling using a new computed path by CSPF and a new LSP-id.

The benefits of this enhancement is that the user can now control how many refreshes of the pending PATH state can be performed before starting a new retry-cycle with a new LSP-id. This is all done without affecting the ability to react faster to failures of the LSP path, which will continue to be governed by the existing retry-timer. By configuring the LSP path initial retry-timer to values that are larger than the retry-timer, the ingress LER will decrease the probability of overwhelming a congested LSR with new state while the previous states installed by the same LSP are lingering and will only be removed after the refresh timeout period expires.

The second enhancement consists of applying a jitter +/- 25% to the value of the retry-timer similar to how it is currently done for the refresh timer. This will further decrease the probability that ingress LER nodes synchronize their sending of Path messages during the retry-procedure in response to a congestion event in the network.

The third enhances the RSVP flow control mechanism by taking into account new parameters: outstanding CSPF requests, Resv timeouts and Path timeouts.

2.2.6 BFD for RSVP-TE

BFD will notify RSVP-TE if the BFD session goes down, in addition to notifying other configured BFD enabled protocols (for example, OSPF, IS-IS, and PIM). This notification will then be used by RSVP-TE to begin the reconvergence process. This greatly accelerates the overall RSVP-TE response to network failures.

All encapsulation types supporting IPv4 and IPv6 are supported because all BFD packets are carried in IPv4 and IPv6 packets; this includes Frame Relay and ATM.

BFD is supported on the following interfaces:

- Ethernet (Null, Dot1Q & QinQ)
- Spoke SDPs
- LAG interfaces

The following interfaces are supported only on the 7750 SR and 7450 ESS:

- VSM interfaces
- POS interfaces (including APS)
- Channelized interfaces (PPP, HDLC, FR, and ATM) on ASAP (priority 1) and channelized MDAs (priority 2) including link bundles and IMA

2.2.7 RSVP-TE LSP Statistics

This feature provides the following counters:

- Per forwarding class forwarded in-profile packet count
- Per forwarding class forwarded in-profile byte count
- Per forwarding class forwarded out of profile packet count
- Per forwarding class forwarded out of profile byte count

The counters are available for an RSVP LSP at the egress datapath of an ingress LER and at the ingress datapath of an egress LER. No LSR statistics are provided.

2.2.8 P2MP RSVP-TE LSP Statistics

This feature provides the following counters for a RSVP P2MP LSP instance:

- Per forwarding class forwarded in-profile packet count.
- Per forwarding class forwarded in-profile byte count.
- Per forwarding class forwarded out of profile packet count.
- Per forwarding class forwarded out of profile byte count.

The above counters are provided for the following LSR roles:

- At ingress LER, a set of per P2MP LSP instance counters for packets forwarded to the P2MP LSP instance without counting the replications is provided. In other words, a packet replicated over multiple branches of the same P2MP LSP instance will count once as long as at least one LSP branch forwarded it.
- At BUD LSR and egress LER, per ILM statistics are provided. These counters will include all packets received on the ILM, whether they match a L2/L3 MFIB record or not. ILM stats will work the same way as for a P2P LSP. In other words, they will count all packets received on the primary ILM, including packets received over the bypass LSP.

When MBB is occurring for an S2L path of an RSVP P2MP LSP, paths of the new and old S2L will both receive packets on the egress LER. Both packets are forwarded to the fabric and outgoing PIM/IGMP interfaces until the older path is torn down by the ingress LER. In this case, packet duplication should be counted.

- No branch LSR statistics are provided.
- The P2MP LSP statistics share the same pool of counters and stat indices the P2P LSP share on the node. Each P2P/P2MP RSVP LSP or LDP FEC consumes one stat index for egress stats and one stat index for ingress stats.
- The user can retrieve the above counters in four different ways:
  - In CLI display of the output of the show command applied to a specific instance, or a specific template instance, of a RSVP P2MP.
  - In CLI display of the output of the monitor command applied to a specific instance, or a specific template instance, of a RSVP P2MP.
  - Via an SNMP interface by querying the MIB.
  - Via an accounting file if statistics collection with the default or user specified accounting policy is enabled for the MPLS LSP stats configuration contexts.
- OAM packets that are forwarded using the LSP encapsulation, for example, P2MP LSP Ping and P2MP LSP Trace, are also included in the above counters.

The user can determine if packets are dropped for a given branch of a P2MP RSVP LSP by comparing the egress counters at the ingress LER with the ILM counters at the egress LER or BUD LSR.

Octet counters are for the entire frame and so include the label stack and the L2 header and padding similar to the existing P2P RSVP LSP and LDP FEC counters. As such, ingress and egress octet counters for an LSP may slightly differ if the type of interface or encapsulation is different (POS, Ethernet NULL, Ethernet Dot1.Q).

2.2.8.1 Configuring RSVP P2MP LSP Egress Statistics

At ingress LER, the configuration of the egress statistics is under the MPLS P2MP LSP context when carrying multicast packets over a RSVP P2MP LSP in the base routing instance. This is the same configuration as the one already supported with P2P RSVP LSP.

```
config
     router
        [no] mpls
            [no] lsp lsp-name p2mp-lsp
                 [no] egress-statistics
                      accounting-policy policy-id
                      no accounting-policy
```

```
                        [no] collect-stats
                   [no] shutdown
```

If there are no stat indices available when the user performs the ‘no shutdown’ command for the egress statistics node, the command fails.

The configuration is in the P2MP LSP template when the RSVP P2MP LSP is used as an I-PMSI or S-PMSI in multicast VPN or in VPLS/B-VPLS.

```
config
     router
          [no] mpls
               lsp-template template-name p2mp
               no lsp-template template-name
                     [no] egress-statistics
                      accounting-policy policy-id
                      no accounting-policy
                      [no] collect-stats
```

If there are no stat indices available at the time an instance of the P2MP LSP template is signaled, no stats are allocated to the instance, but the LSP is brought up. In this case, an operational state of out-of-resources is shown for the egress stats in the show output of the P2MP LSP S2L path.

2.2.8.2 Configuring RSVP P2MP LSP Ingress Statistics

When the ingress LER signals the path of the S2L sub-LSP, it includes the name of the LSP and that of the path in the Session Name field of the Session Attribute object in the Path message. The encoding is as follows:

Session Name: *lsp-name::path-name*, where lsp-name component is encoded as follows:

1. P2MP LSP via user configuration for L3 multicast in global routing instance: “LspNameFromConfig”
2. P2MP LSP as I-PMSI or S-PMSI in L3 mVPN: templateName-SvcId-mTTmIndex
3. P2MP LSP as I-PMSI in VPLS/B-VPLS: templateName-SvcId-mTTmIndex

The ingress statistics CLI configuration allows the user to match either on the exact name of the P2MP LSP as configured at the ingress LER or on a context which matches on the template name and the service-id as configured at the ingress LER.

```
config
     router
          [no] mpls
                  ingress-statistics
```

```
[no] lsp lsp-name sender sender-address
       accounting-policy policy-id
       no accounting-policy
       [no] collect-stats
       [no] shutdown

[no] p2mp-template-lsp rsvp-session-name
SessionNameString sender sender-address
       accounting-policy policy-id
       no accounting-policy
       [no] collect-stats
       max-stats integer<1-8192 | max, default max>
       no max-stats
[no] shutdown
```

When the matching is performed on a context, the user must enter the RSVP session name string in the format "*templateName-svcId*" to include the LSP template name as well as the mVPN VPLS/B-VPLS service ID as configured at the ingress LER. In this case, one or more P2MP LSP instances signaled by the same ingress LER could be associated with the ingress statistics configuration. In this case, the user is provided with CLI parameter **max-stats** to limit the maximum number of stat indices which can be assigned to this context. If the context matches more than this value, the additional request for stat indices from this context is rejected.

The rules when configuring an ingress statistics context based on template matching are the following:

1. **max-stats** once allocated can be increased but not decreased unless the entire ingress statistics context matching a template name is deleted.
2. In order to delete ingress statistics context matching a template name, a shutdown is required.
3. An accounting policy cannot be configured or de-configured until the ingress statistics context matching a template name is shutdown.
4. After deleting an accounting policy from an ingress statistics context matching a template name, the policy is not removed from the log until a 'no shut' is performed on the ingress statistics context.

If there are no stat indices available at the time the session of the P2MP LSP matching a template context is signaled and the session state installed by the egress LER, no stats are allocated to the session.

Furthermore, the assignment of stat indices to the LSP names that match the context will also be not deterministic. The latter is due to the fact that a stat index is assigned and released following the dynamics of the LSP creation or deletion by the ingress LER. For example, a multicast stream crosses the rate threshold and is moved to a newly signaled S-PMSI dedicated to this stream. Later on, the same steam crosses the threshold downwards and is moved back to the shared I-PMSI and the P2MP LSP corresponding to the S-PMSI is deleted by the ingress LER.

2.2.8.3 Configuring Implicit Null

The implicit null label option allows a router egress LER to receive MPLS packets from the previous hop without the outer LSP label. The operation of the previous hop is referred to as penultimate hop popping (PHP).

This option is signaled by the egress LER to the previous hop during the LSP signaling with RSVP control protocol. In addition, the egress LER can be configured to receive MPLS packet with the implicit null label on a static LSP.

The user can configure your router to signal the implicit null label value over all RSVP interfaces and for all RSVP LSPs for which this node is the egress LER using the **implicit-null-label** command in the **config**>**router**>**rsvp** context.

The user must shut down RSVP before being able to change the implicit null configuration option.

The user can also override the RSVP level configuration for a specific RSVP interface:

config>**router**>**rsvp**>**if**>**implicit-null-label** {**enable** | **disable**}

All LSPs for which this node is the egress LER and for which the path message is received from the previous hop node over this RSVP interface will signal the implicit null label. This means that if the egress LER is also the merge-point (MP) node, then the incoming interface for the path refresh message over the bypass dictates if the packet will use the implicit null label or not; the same applies to a 1-to-1 detour LSP.

By default, an RSVP interface inherits the RSVP level configuration. The user must shut down the RSVP interface before being able to change the implicit null configuration option.

Note: The RSVP interface must be shutdown regardless of whether the new value for the interface is the same or different than the one it is currently using.

The egress LER does not signal the implicit null label value on P2MP RSVP LSPs. However, the PHP node can honor a Resv message with the label value set to the implicit null value when the egress LER is a third party implementation.

The implicit null label option is also supported on a static label LSP. The following commands can be used to cause the node to push or to swap to an implicit null label on the MPLS packet:

config>**router**>**mpls**>**static-lsp**>**push implicit-null-label nexthop** *ip-address*

config>**router**>**mpls**>**if**>**label-map**>**swap implicit-null-label nexthop** *ip-address*

2.2.9 Using Unnumbered Point-to-Point Interface in RSVP

This feature introduces the use of unnumbered IP interface as a Traffic Engineering (TE) link for the signaling of RSVP P2P LSP and P2MP LSP.

An unnumbered IP interface is identified uniquely on a router in the network by the tuple {router-id, ifIndex}. Each side of the link assigns a system-wide unique interface index to the unnumbered interface. ISIS, OSPF, RSVP, and OAM modules will use this tuple to advertise the link information, signal LSP paths over this unnumbered interface, or send and respond to an MPLS echo request message over an unnumbered interface.

The interface borrowed IP address is used exclusively as the source address for IP packets that are originated from the interface and needs to be configured to an address different from system interface for the FRR bypass LSP to come up at the ingress LER.

The borrowed IP address for an unnumbered interface is configured using the following CLI command with a default value set to the system interface address:

config>router>if>unnumbered [*ip-int-name* | *ip-address*]

The support of unnumbered TE link in IS-IS consists of adding a new sub-TLV of the extended IS reachability TLV, which encodes the Link Local and Link Remote Identifiers as defined in RFC 5307.

The support of unnumbered TE link in OSPF consists of adding a new sub-TLV, which encodes the same Link Local and Link Remote Identifiers in the Link TLV of the TE area opaque LSA and sends the local Identifier in the Link Local Identifier TLV in the TE link local opaque LSA as per RFC 4203.

The support of unnumbered TE link in RSVP implements the signaling of unnumbered interfaces in ERO/RRO as per RFC 3477 and the support of IF_ID RSVP_HOP object with a new Ctype as per Section 8.1.1 of RFC 3473. The IPv4 Next/Previous Hop Address field is set to the borrowed IP interface address.

The unnumbered IP is advertised by IS-IS TE and OSPF TE, and CSPF can include them in the computation of a path for a P2P LSP or for the S2L of a P2MP LSP. This feature does not, however, support defining an unnumbered interface a hop in the path definition of an LSP.

A router creates an RSVP neighbor over an unnumbered interface using the tuple {router-id, ifIndex}. The router-id of the router that advertised a given unnumbered interface index is obtained from the TE database. As a result, if TE is disabled in IS-IS or OSPF, a non-CSPF LSP with the next-hop for its path is over an unnumbered interface will not come up at the ingress LER since the router-id of the neighbor that has the next-hop of the path message cannot be looked up. In this case, the LSP path will remain in operationally down state with a reason noRouteToDestination. If a PATH message was received at the LSR in which TE was disabled and the next-hop for the LSP path is over an unnumbered interface, a PathErr message is sent back to the ingress LER with the *Routing Problem* error code of 24 and an error value of 5 “No route available toward destination”.

All MPLS features available for numbered IP interfaces are supported, with the exception of the following:

- Configuring a router-id with a value other than system.
- Signaling of an LSP path with an ERO based a loose/strict hop using an unnumbered TE link in the path hop definition.
- Signaling of one-to-one detour LSP over unnumbered interface.
- Unnumbered RSVP interface registration with BFD.
- RSVP Hello and all Hello related capabilities such as Graceful-restart helper.
- The user SRLG database feature. The user-srlg-db option under MPLS allows the user to manually enter the SRLG membership of any link in the network in a local database at the ingress LER. The user cannot enter an unnumbered interface into this database and as such, all unnumbered interfaces are considered as having no SRLG membership if the user enabled the user-srlg-db option.

This feature also extends the support of lsp-ping, p2mp-lsp-ping, lsp-trace, and p2mp-lsptrace to P2P and P2MP LSPs that have unnumbered TE links in their path.

2.2.9.1 Operation of RSVP FRR Facility Backup over Unnumbered Interface

When the Point-of-Local Repair (PLR) node activates the bypass LSP by sending a PATH message to refresh the path state of protected LSP at the Merge-Point (MP) node, it must use an *IPv4 tunnel sender address* in the sender template object that is different than the one used by the ingress LER in the PATH message. These are the procedures specified in RFC 4090 that are followed in the SR OS implementation.

The router uses the address of the outgoing interface of the bypass LSP as the *IPv4 tunnel sender address* in the sender template object. This address is different from the system interface address used in the sender template of the protected LSP by the ingress LER and so, there are no conflicts when the ingress LER acts as a PLR.

When the PLR is the ingress LER node and the outgoing interface of the bypass LSP is unnumbered, it is required that the user assigns to the interface a borrowed IP address that is different from the system interface. If not, the bypass LSP will not come up.

In addition, the PLR node will include the IPv4 RSVP_HOP object (C-Type=1) or the IF_ID RSVP_HOP object (C-Type=3) in the PATH message if the outgoing interface of the bypass LSP is numbered or unnumbered respectively.

When the MP node receives the PATH message over the bypass LSP, it will create the merge-point context for the protected LSP and associate it with the existing state if any of the following is satisfied:

- Change in C-Type of the RSVP_HOP object, or
- C-Type is IF_ID RSVP_HOP and did not change but IF_ID TLV is different, or
- Change in IPv4 Next/Previous Hop Address in RSVP_HOP object regardless of the C-Type value.

These procedures at PLR and MP nodes are followed in both link-protect and node-protect FRR. If the MP node is running a pre-Release 11.0 implementation, it will reject the new IF_ID C-Type and will drop the PATH over bypass. This will result in the protected LSP state expiring at the MP node, which will tear down the path. This is the case in general when node-protect FRR is enabled and the MP node does not support unnumbered RSVP interface.

2.3 MPLS Transport Profile

MPLS can be used to provide a network layer to support packet transport services. In some operational environments, it is desirable that the operation and maintenance of such an MPLS based packet transport network follow operational models typical in traditional optical transport networks (for example, SONET/SDH), while providing additional OAM, survivability and other maintenance functions targeted at that environment.

MPLS-TP defines a profile of MPLS targeted at transport applications. This profile defines the specific MPLS characteristics and extensions required to meet transport requirements, while retaining compliance to the standard IETF MPLS architecture and label switching paradigm. The basic requirements are architecture for MPLS-TP are described by the IETF in RFC 5654, RFC 5921, and RFC 5960, in order to meet two objectives:

1. To enable MPLS to be deployed in a transport network and operated in a similar manner to existing transport technologies.
2. To enable MPLS to support packet transport services with a similar degree of predictability to that found in existing transport networks.

In order to meet these objectives, MPLS-TP has a number of high level characteristics:

- It does not modify the MPLS forwarding architecture, which is based on existing pseudowire and LSP constructs. Point-to-point LSPs may be unidirectional or bi-directional. Bi-directional LSPs must be congruent (that is, co-routed and follow the same path in each direction). The system supports bidirectional co-routed MPLS-TP LSPs.
- There is no LSP merging.
- OAM, protection, and forwarding of data packets can operate without IP forwarding support. When static provisioning is used, there is no dependency on dynamic routing or signaling.
- LSP and pseudowire monitoring is only achieved through the use of OAM and does not rely on control plane or routing functions to determine the health of a path. For example, LDP hello failures do not trigger protection.
- MPLS-TP can operate in the absence of an IP control plane and IP forwarding of OAM traffic. MPLS-TP is only supported on static LSPs and PWs.

The system supports MPLS-TP on LSPs and PWs with static labels. MPLS-TP is not supported on dynamically signaled LSPs and PWs. MPLS-TP is supported for Epipe, Apipe, and Cpipe VLLs, and Epipe spoke SDP termination on IES, VPRN and VPLS. Static PWs may use SDPs that use either static MPLS-TP LSPs or RSVP-TE LSPs.

The following MPLS-TP OAM and protection mechanisms, defined by the IETF, are supported:

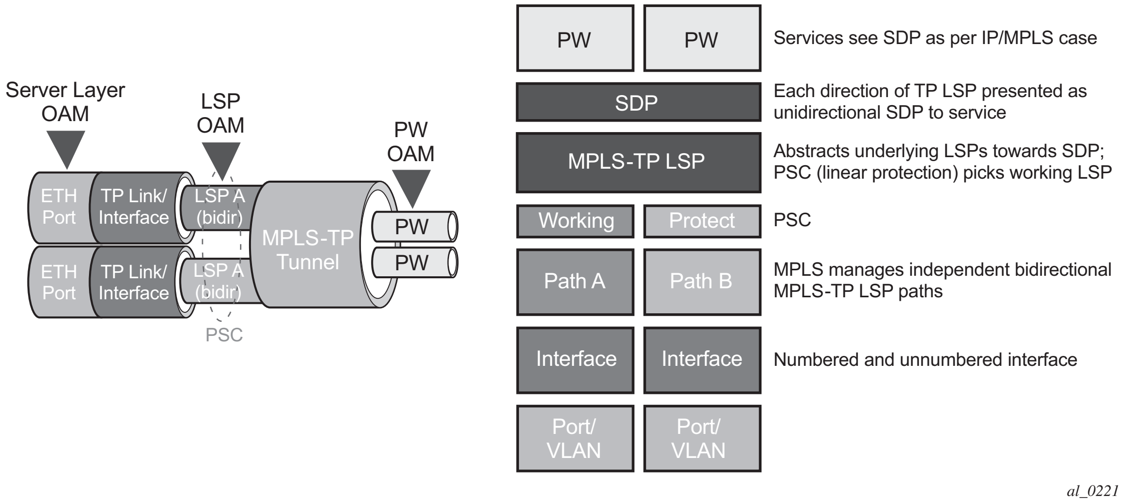
- MPLS-TP Generic Associated Channel for LSPs and PWs (RFC 5586)
- MPLS-TP Identifiers (RFC 6370)
- Proactive CC, CV, and RDI using BFD for LSPs (RFC 6428)
- On-Demand CV for LSPs and PWs using LSP Ping and LSP Trace (RFC 6426)
- 1-for-1 Linear protection for LSPs (RFC 6378)
- Static PW Status Signaling (RFC 6478)

The system can play the role of an LER and an LSR for static MPLS-TP LSPs, and a PE/T-PE and an S-PE for static MPLS-TP PWs. It can also act as a S-PE for MPLS-TP segments between an MPLS network that strictly follows the transport profile, and an MPLS network that supports both MPLS-TP and dynamic IP/MPLS.

2.3.1 MPLS-TP Model

Figure 7 shows a high level functional model for MPLS-TP in SR OS. LSP A and LSP B are the working and protect LSPs of an LSP tunnel. These are modeled as working and protect paths of an MPLS-TP LSP in SR OS. MPLS-TP OAM runs in-band on each path. 1:1 linear protection coordinates the working and protect paths, using a protection switching coordination protocol (PSC) that runs in-band on each path over a Generic Associated Channel (G-ACh) on each path. Each path can use either an IP numbered, IP unnumbered, or MPLS-TP unnumbered (that is, non-IP) interface.

Figure 7 **MPLS-TP Model**


Server Layer OAM
LSP OAM
PW OAM
ETH Port
TP Link/ Interface
LSP A (bidir)
ETH Port
TP Link/ Interface
LSP A (bidir)
PSC
MPLS-TP Tunnel
PW
PW
PW
PW
Services see SDP as per IP/MPLS case
SDP
Each direction of TP LSP presented as unidirectional SDP to service
MPLS-TP LSP
Abstracts underlying LSPs towards SDP; PSC (linear protection) picks working LSP
Working
Protect
PSC
Path A
Path B
MPLS manages independent bidirectional MPLS-TP LSP paths
Interface
Interface
Numbered and unnumbered interface
Port/ VLAN
Port/ VLAN
al_0221

All MPLS-TP LSPs are bidirectional co-routed, as detailed in RFC5654. That is, the forward and backward directions follow the same route (in terms of links and nodes) across the network. Both directions are set up, monitored and protected as a single entity. Therefore, both ingress and egress directions of the same LSP segment are associated at the LER and LSR and use the same interface (although this is not enforced by the system).

In the above model, an SDP can use one MPLS-TP LSP. This abstracts the underlying paths towards the overlying services, which are transported on pseudowires. Pseudowires are modeled as spoke SDPs and can also use MPLS-TP OAM. PWs with static labels may use SDPs that, in turn, use either signaled RSVP-TE LSPs or one static MPLS-TP LSP.

2.3.2 MPLS-TP Provider Edge and Gateway

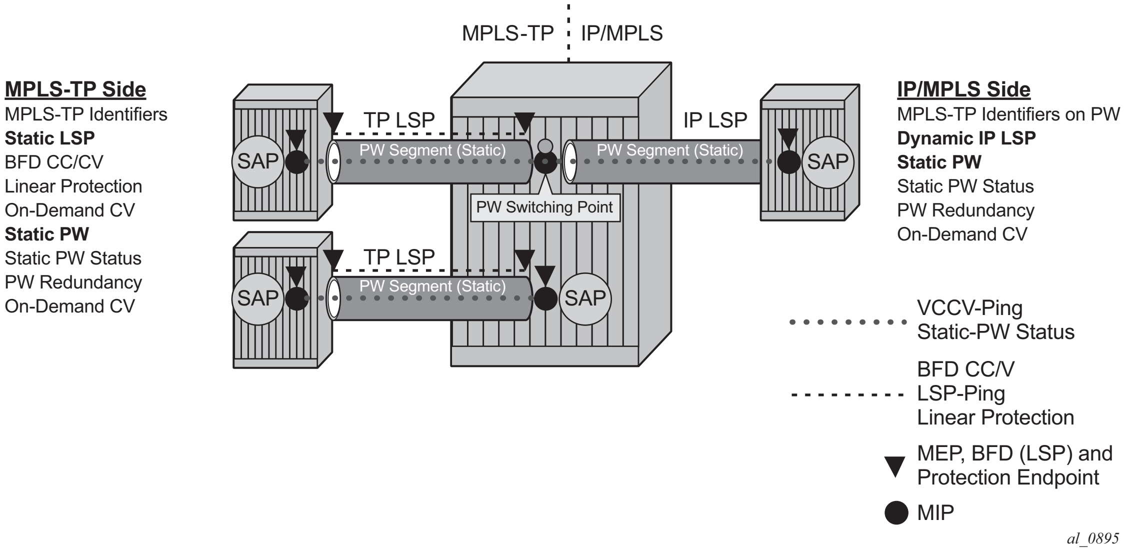
This section describes some example roles for the system in an MPLS-TP network.

2.3.2.1 VLL Services

The system may use MPLS TP LSPs, and PWs, to transport point to point virtual leased line services. The router may play the role of a terminating PE or switching PE for VLLs. Epipe, Apipe, and Cpipe VLLs are supported.

Figure 8 illustrates the use of the system as a T-PE for services in an MPLS-TP domain, and as a S-PE for services between an MPLS-TP domain and an IP/MPLS domain. Static PWs with MPLS-TP identifiers, originating in the MPLS-TP network, are transported over static MPLS-TP LSPs. These either terminate on a local SAP on the system, or are switched to another PW segment across the IP/MPLS network. The PW segment in the IP/MPLS network may have static labels or be signaled using T-LDP.

Figure 8 **MPLS-TP Provider Edge and Gateway, VLL Services**



2.3.2.2 Spoke SDP Termination

Figure 9 and Figure 10 illustrate the model for spoke SDP termination on VPLS and IES/VPRN services, respectively. Similar to the VLL case, the static MPLS-TP PW may terminate on an interface belonging to the service on the router at the border between the MPLS-TP and IP/MPLS networks, or be switched to another PW segment to be terminated on a remote PE.

Figure 9 **MPLS-TP Provider Edge and Gateway, Spoke SDP Termination on VPLS**

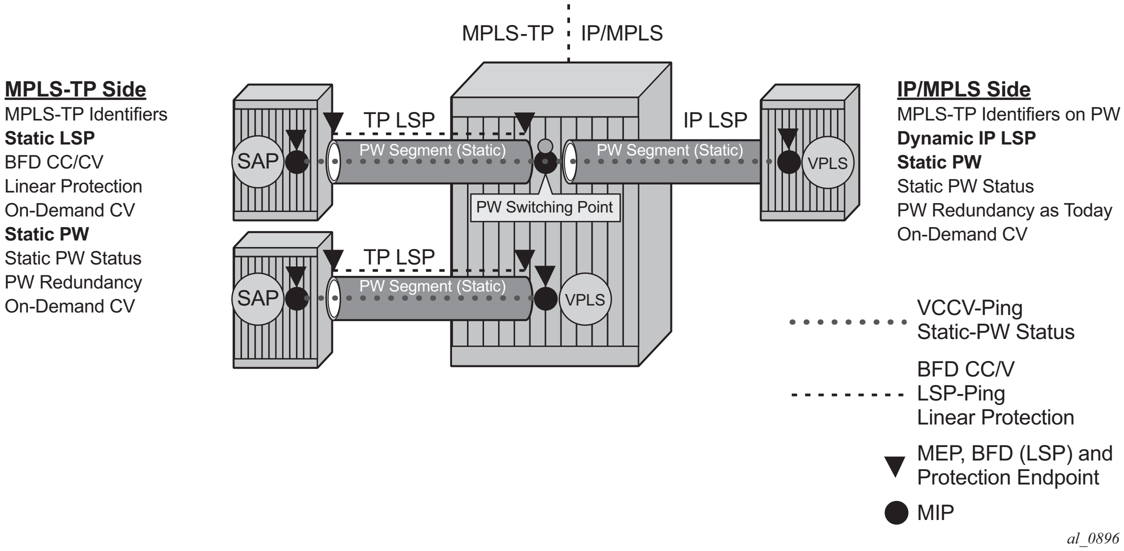
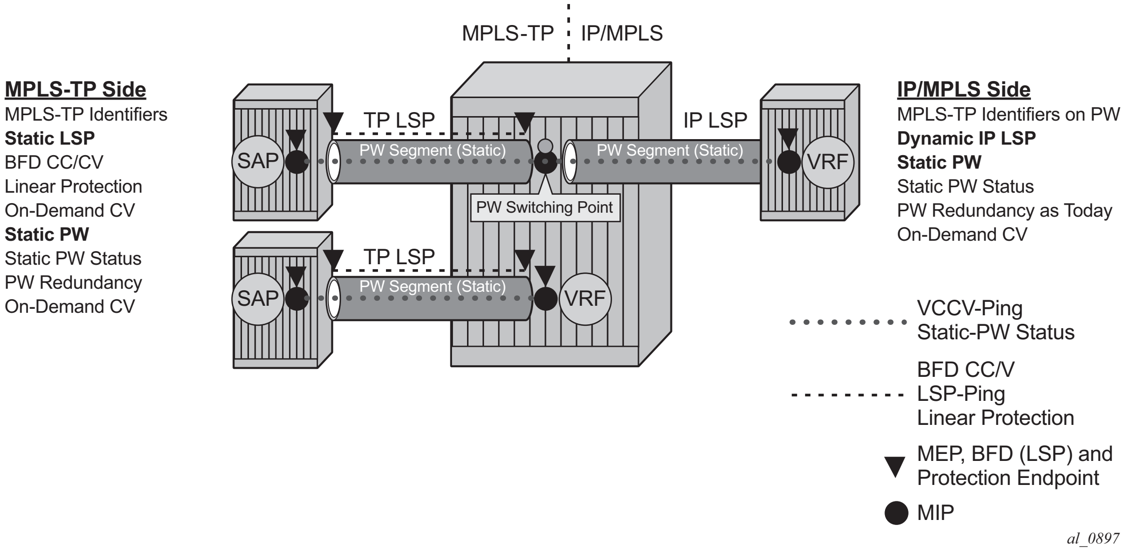


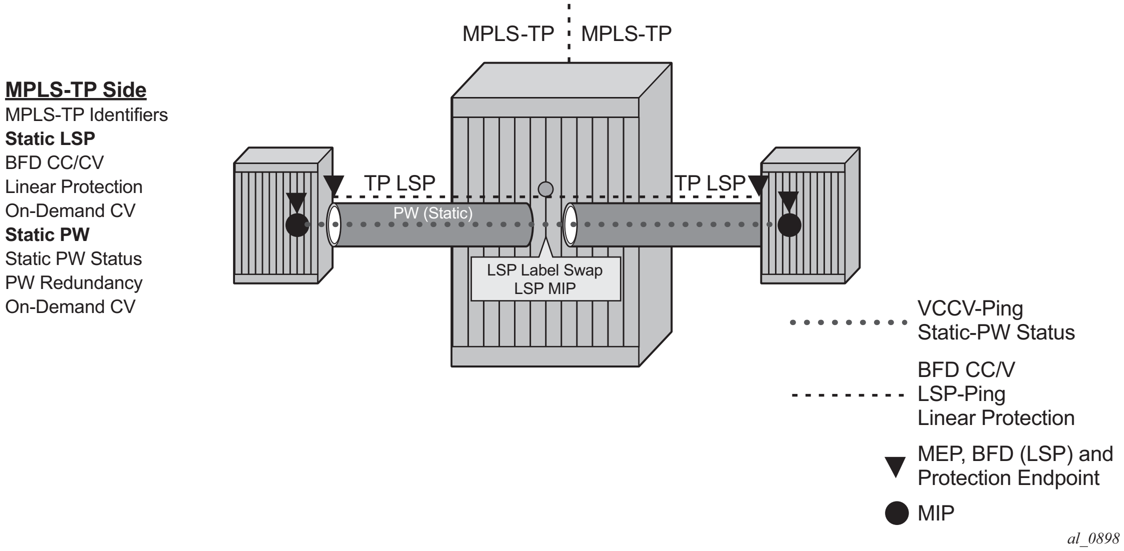
Figure 10 **MPLS-TP Provider Edge and Gateway, Spoke SDP Termination on IES/VPRN**



2.3.3 MPLS-TP LSR

The SR OS MPLS-TP LSR model is illustrated in MPLS-TP LSR. The system is able to swap a statically configured LSP label on an ingress path to a statically configured LSP label on an egress path. Bidirectional co-routed MPLS TP LSPs are supported by configuring the forward and reverse paths of the LSP to use the same ports on ingress and egress.

Figure 11 **MPLS-TP LSR**



2.3.4 Detailed Descriptions of MPLS-TP

2.3.4.1 MPLS-TP LSPs

SR OS supports the configuration of MPLS-TP tunnels, which comprise a working and, optionally, a protect LSP. In SR OS, a tunnel is referred to as an LSP, while an MPLS-TP LSP is referred to as a path. It is then possible to bind an MPLS-TP tunnel to an SDP.

MPLS-TP LSPs (that is, paths) with static labels are supported. MPLS-TP is not supported for signaled LSPs.

Both bidirectional associated (where the forward and reverse directions of a bidirectional LSP are associated at a given LER, but may take different routes through the intervening network) and bidirectional co-routed (where the forward and reverse directions of the LSP are associated at each LSR, and take the same route through the network) are possible in MPLS-TP. However, only bidirectional co-routed LSPs are supported.

It is possible to configure MPLS-TP identifiers associated with the LSP, and MPLS-TP OAM parameters on each LSP of a tunnel. MPLS-TP protection is configured for a tunnel at the level of the protect path level. Both protection and OAM configuration is managed via templates, in order to simplify provisioning for large numbers of tunnels.

The router may play the role of either an LER or an LSR.

2.3.4.2 MPLS-TP on Pseudowires

MPLS-TP is supported on PWs with static labels. The provisioning model supports RFC6370-style PW path identifiers for MPLS-TP PWs.

MPLS-TP PWs reuse the static PW provisioning model of previous SR OS releases. Including the use of the PW-switching key work to distinguish an S-PE. Therefore, the primary distinguishing feature for an MPLS-TP PW is the ability to configure MPLS-TP PW path identifiers, and to support MPLS-TP OAM and static PW status signaling.

The system can perform the role of a T-PE or an S-PE for a PW with MPLS-TP.

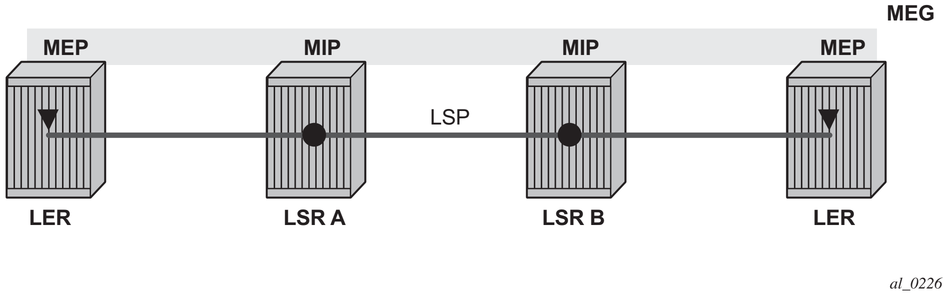
A spoke SDP with static PW labels and MPLS-TP identifiers and OAM capabilities can use an SDP that uses either an MPLS-TP tunnel, or that uses regular RSVP-TE LSPs. The control word is supported for all MPLS-TP PWs.

2.3.5 MPLS-TP Maintenance Identifiers

MPLS-TP is designed for use both with, and without, a control plane. MPLS-TP therefore specifies a set of identifiers that can be used for objects in either environment. This includes a path and maintenance identifier architecture comprising Node, Interface, PW and LSP identifiers, Maintenance Entity Groups (MEGs), Maintenance End Points (MEPs) and Maintenance Intermediate Points (MIPs). These identifiers are specified in RFC6370.

MPLS-TP OAM and protection switching operates within a framework that is designed to be similar to existing transport network maintenance architectures. MPLS-TP introduces concept of maintenance domains to be managed and monitored. In these, Maintenance Entity Group End Points (MEPs) are edges of a maintenance domain. OAM of a maintenance level must not leak beyond corresponding MEP and so MEPs typically reside at the end points of LSPs and PWs. Maintenance Intermediate Points (MIPS) define intermediate nodes to be monitored. Maintenance Entity Groups (MEGs) comprise all the MEPs and MIPs on an LSP or PW.

Figure 12 **MPLS-TP Maintenance Architecture**



Both IP-compatible and ICC (ITU-T carrier code) based identifiers for the above objects are specified in the IETF, but only the IP-compatible identifiers defined in RFC6370 are supported.

SR OS supports the configuration of the following node and interface related identifiers:

- Global_ID: this is similar to the global ID that can be configured for Dynamic MS-PWs. However, in MPLS-TP this should be set to the AS# of the node. If not explicitly configured, then it assumes the default value of 0. In SR OS, the source Global ID for an MPLS-TP Tunnel is taken to be the Global ID configured at the LER. The destination Global ID is optional in the tunnel configuration. If it is not configured, then it is taken as the same as the source Global ID.
- Node_ID: This is a 32-bit value assigned by the operator within the scope of the Global_ID. The system supports the configuration of an IPv4 formatted address <a.b.c.d> or an unsigned 32-bit integer for the MPLS-TP Node ID at each node. The node ID must be unique within the scope of the global ID, but there is no requirement for it to be a valid routable IP address. Indeed, a node-id can represent a separate IP-compatible addressing space that may be separate

from the IP addressing plan of the underlying network. If no node ID is configured, then the node ID is taken to be the system interface IPv4 address of the node. When configuring a tunnel at an LER, either an IPv4 or an unsigned integer Node ID can be configured as the source and destination identifiers, but both ends must be of the same type.

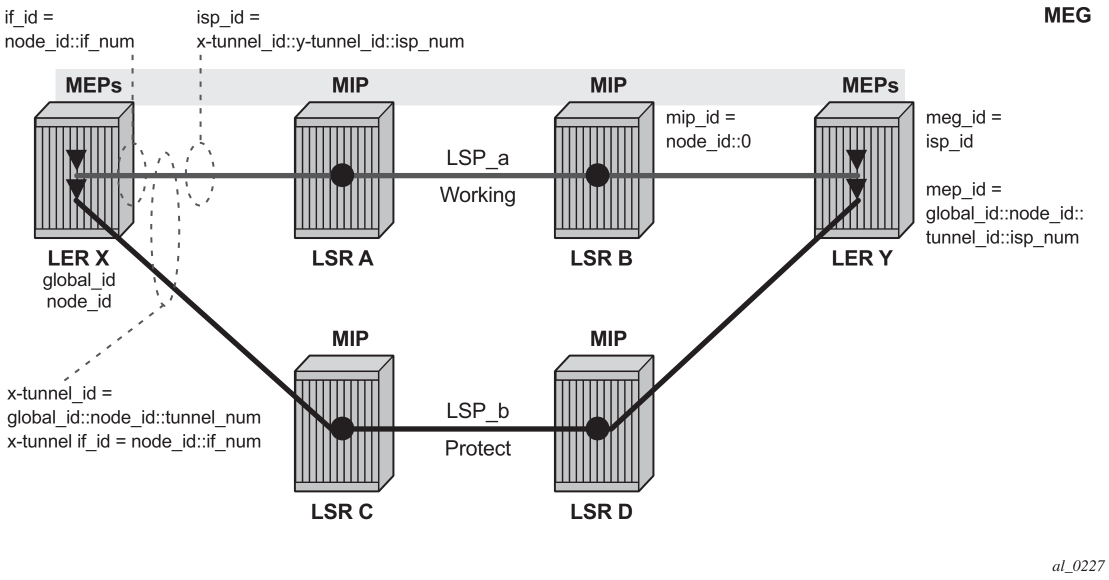
- IF_ID: This is an MPLS-TP section layer identifier at the MPLS interface level. On the router, this is used to provide an identifier for the LSP-Trace DSMAP when an IP identifier is not available. The IF_ID is a 64-bit identifier of an MPLS-TP interface on a node that is unique within the scope of a Global_ID. It is composed of the Node_ID and the IF_Num. The IF_Num is a node-wide unique identifier for an MPLS-TP interface. On the router, this is primarily used for supporting the DSMAP TLV in LSP Trace using MPLS-TP identifiers with unnumbered MPLS-TP interfaces.

Statically configured LSPs are identified using GMPLS-compatible identifiers with the addition of a Tunnel_Num and LSP_Num. As in RSVP-TE, tunnels represent, for example, a set of working and protect LSPs. These are GMPLS-compatible because GMPLS chosen by the IETF as the control plane for MPLS-TP LSPs, although this is not supported in Release 11.0 of the software. PWs are identified using a PW Path ID which has the same structure as FEC129 AII Type 2.

SR OS derives the identifiers for MEPs and MIPs on LSPs and PWs based on the configured identifiers for the MPLS-TP Tunnel, LSP or PW Path ID, for use in MPLS-TP OAM and protection switching, as per RFC6370.

The information models for LSPs and PWs are illustrated in Figure 13 and Figure 14. The figures use the terminology defined in RFC6370.

Figure 13 **MPLS-TP LSP and Tunnel Information Model**

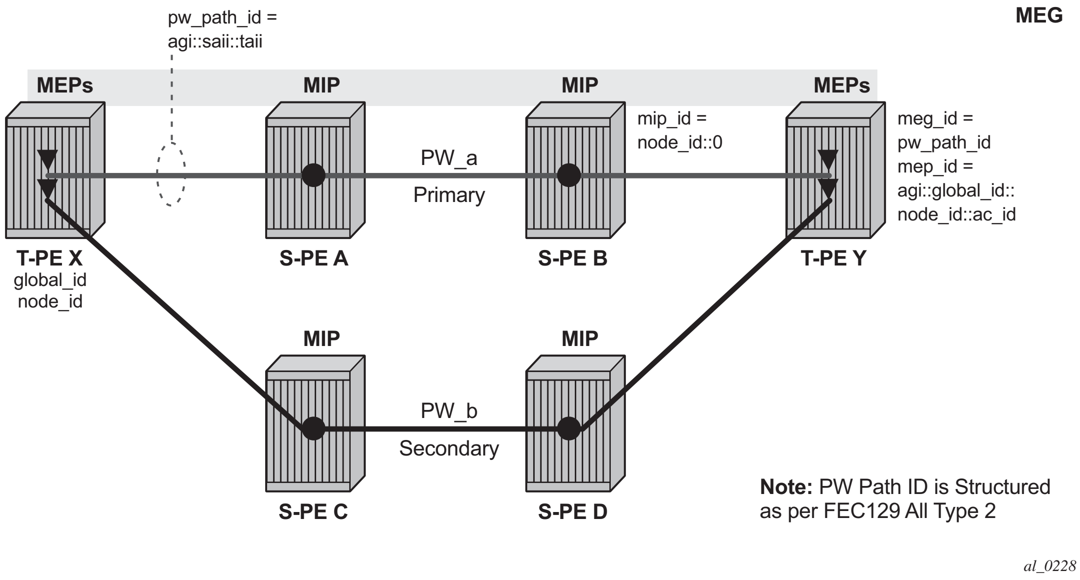


The MPLS-TP Tunnel ID and LSP ID are not to be confused with the RSVP-TE tunnel id implemented on the router system. Table 8 shows how these map to the X and Y ends of the tunnel shown in Figure 13 for the case of co-routed bidirectional LSPs.

Table 8 **Mapping from RSVP-TE to MPLS-TP Maintenance Identifiers**

RSVP-TE Identifier	MPLS-TP Maintenance Identifier
Tunnel Endpoint Address	Node ID (Y)
Tunnel ID (X)	Tunnel Num (X)
Extended Tunnel ID	Node ID (X)
Tunnel Sender Address	Node ID (X)
LSP ID	LSP Num

Figure 14 **MPLS-TP PW Information Model**

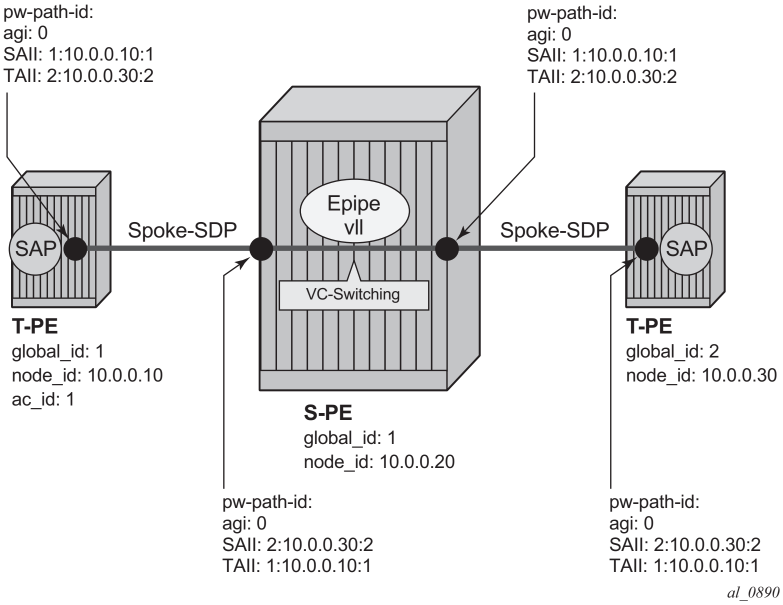


In the PW information model shown in Figure 14, the MS-PW is identified by the PW Path ID that is composed of the full AGI:SAII:TAII. The PW Path ID is also the MEP ID at the T-PEs, so a user does not have to explicitly configure a MEP ID; it is automatically derived by the system. For MPLS-TP PWs with static labels, although the PW is not signaled end-to-end, the directionality of the SAII and TAII is taken to be the same as for the equivalent label mapping message that is from downstream to upstream. This is to maintain consistency with signaled pseudowires using FEC 129.

On the system, an S-PE for an MS-PW with static labels is configured as a pair of spoke SDPs bound together in an VLL service using the VC-switching command. Therefore, the PW Path ID configured at the spoke SDP level at an S-PE must contain the Global-ID, Node-ID and AC-ID at the far end T-PEs, not the local S-PE. The ordering of the SAII:TAII in the PW Path ID where static PWs are used should be consistent with the direction of signaling of the egress label to a spoke SDP forming that segment, if that label were signaled using T-LDP (in downstream unsolicited mode). VCCV Ping will check the PW ID in the VCCV Ping echo request message against the configured PW Path ID for the egress PW segment.

Figure 15 shows an example of how the PW Path IDs can be configured for a simple two-segment MS-PW.

Figure 15 **Example usage of PW Identifiers**

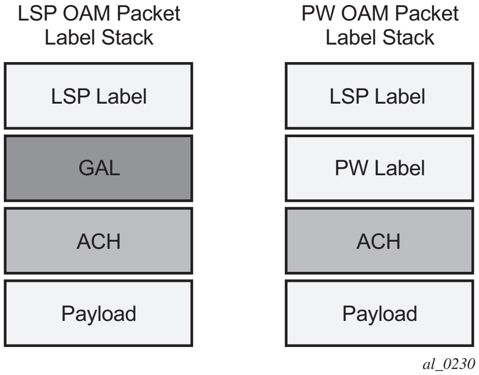


2.3.5.1 Generic Associated Channel

MPLS-TP requires that all OAM traffic be carried in-band on both directions of an LSP or PW. This is to ensure that OAM traffic always shares fate with user data traffic. This is achieved by using an associated control channel on an LSP or PW, similar to that used today on PWs. This creates a channel, which is used for OAM, protection switching protocols (for example, LSP linear protection switching coordination), and other maintenance traffic., and is known as the Generic Associated Channel (G-ACh).

RFC5586 specifies mechanisms for implementing the G-ACh, relying on the combination of a reserved MPLS label, the Generic-ACH Label (GAL), as an alert mechanism (value=13) and Generic Associated Channel Header (G-ACH) for MPLS LSPs, and using the Generic Associated Channel Header, only, for MPLS PWs (although the GAL is allowed on PWs). The purpose of the GAL is to indicate that a G-ACH resides at the bottom of the label stack, and is only visible when the bottom non-reserved label is popped. The G-ACH channel type is used to indicate the packet type carried on the G-ACh. Packets on a G-ACh are targeted to a node containing a MEP by ensuring that the GAL is pushed immediately below the label that is popped at the MEP (for example, LSP endpoint or PW endpoint), so that it can be inspected as soon as the label is popped. A G-ACh packet is targeted to a node containing a MIP by setting the TTL of the LSP or PW label, as applicable, so that it expires at that node, in a similar manner to the SR OS implementation of VCCV for MS-PWs.

Figure 16 **Label for LSP and PW G-ACh Packets**



The system supports the G-ACh on static pseudowires and static LSPs.

2.3.5.2 MPLS-TP Operations, Administration and Maintenance (OAM)

This section details the MPLS-TP OAM mechanisms that are supported.

2.3.5.2.1 On-Demand Connectivity Verification (CV) using LSP-Ping

MPLS–TP supports mechanisms for on demand CC/CV as well as route tracing for LSPs and PWs. These are required to enable an operator to test the initial configuration of a transport path, or to assist with fault isolation and diagnosis. On demand CC/CV and route tracing for MPLS-TP is based on LSP-Ping and is described in RFC6426. Three possible encapsulations are specified in that RFC:

- IP encapsulation, using the same label stack as RFC 8029, or encapsulated in the IPv4 G-ACh channel with a GAL/ACH
- and non-IP encapsulation with GAL/ACH for LSPs and ACH for PWs.

In IP-encapsulation, LSP-Ping packets are sent over the MPLS LSP for which OAM is being performed and contain an IP/UDP packet within them. The On-demand CV echo response message is sent on the reverse path of the LSP, and the reply contains IP/UDP headers followed by the On-demand CV payload.

In non-IP environments, LSP ping can be encapsulated with no IP/UDP headers in a G-ACh and use a source address TLV to identify the source node, using forward and reverse LSP or PW associated channels on the same LSP or PW for the echo request and reply packets. In this case, no IP/UDP headers are included in the LSP-Ping packets.

The routers support the following encapsulations:

- IP encapsulation with ACH for PWs (as per VCCV type 1).
- IP encapsulation without ACH for LSPs using labeled encapsulation
- Non-IP encapsulation with ACH for both PWs and LSPs.

LSP Ping and VCCV Ping for MPLS-TP use two new FEC sub-types in the target FEC stack in order to identify the static LSP or static PW being checked. These are the Static LSP FEC sub-type, which has the same format as the LSP identifier described above, and the Static PW FEC sub-type,. These are used in-place of the currently defined target FEC stack sub-TLVs.

In addition, MPLS-TP uses a source/destination TLV to carry the MPLS-TP global-id and node-id of the target node for the LSP ping packet, and the source node of the LSP ping packet.

LSP Ping and VCCV-Ping for MPLS-TP can only be launched by the LER or T-PE. The replying node therefore sets the TTL of the LSP label or PW label in the reply packet to 255 to ensure that it reaches the node that launched the LSP ping or VCCV Ping request.

Downstream Mapping Support

RFC 8029 specifies four address types for the downstream mapping TLV for use with IP numbered and unnumbered interfaces, as listed in Table 9:

Table 9 **Downstream Mapping (RFC 8029)**

Type #	Address Type	K Octets	Reference
1	IPv4 Numbered	16	RFC 8029
2	IPv4 Unnumbered	16	
3	IPv6 Numbered	40	
4	IPv6 Unnumbered	28	

RFC 6426 adds address type 5 for use with Non IP interfaces, including MPLS-TP interfaces. In addition, this RFC specifies that type 5 must be used when non-IP ACH encapsulation is used for LSP Trace.

It is possible to send and respond to a DSMAP/DDMAP TLV in the LSP Trace packet for numbered IP interfaces as per RFC8029. In this case, the echo request message contains a downstream mapping TLV with address type 1 (IPv4 address) and the IPv4 address in the DDMAP/DSMAP TLV is taken to be the IP address of the IP interface that the LSP uses. The LSP trace packet therefore contains a DSMAP TLV in addition to the MPLS-TP static LSP TLV in the target FEC stack.

DSMAP/DDMAP is not supported for pseudo wires.

2.3.5.2.2 Proactive CC, CV and RDI

Proactive Continuity Check (CC) is used to detect a loss of continuity defect (LOC) between two MEPs in a MEG. Proactive Connectivity Verification (CV) is used to detect an unexpected connectivity defect between two MEPs (for example, mis-merging or mis-connection), as well as unexpected connectivity within the MEG with an unexpected MEP. This feature implements both functions using proactive generation of OAM packets by the source MEP that are processed by the peer sink MEP. CC and CV packets are always sent in-band such that they fate share with user traffic, either on an LSP, PW or section and are used to trigger protection switching mechanisms.

Proactive CC/CV based on bidirectional forwarding detection (BFD) for MPLS-TP is described in RFC6428. BFD packets are sent using operator configurable timers and encapsulated without UDP/IP headers on a standardized G-ACh channel on an LSP or PW. CC packets simply consist of a BFD control packet, while CV packets also include an identifier for the source MEP in order that the sink MEP can detect if it is receiving packets from an incorrect peer MEP, indicating a mis-connectivity defect. Other defect types (including period mis-configuration defect) should be supported.

When a supported defect is detected, an appropriate alarm is generated (for example, log, SNMP trap) at the receiving MEP and all traffic on the associated transport path (LSP or PW) is blocked. This is achieved using linear protection for CC defects, and by blocking the ingress data path for CV defects. The system supports both a CC-only mode and a combine CC/CV mode, as defined in RFC6428.

When an LSP with CV is first configured, the LSP is held in the CV defect state for 3.5 seconds after the first valid CV packet is received.

Figure 17 **BFD Used for Proactive CC on MPLS-TP LSP**

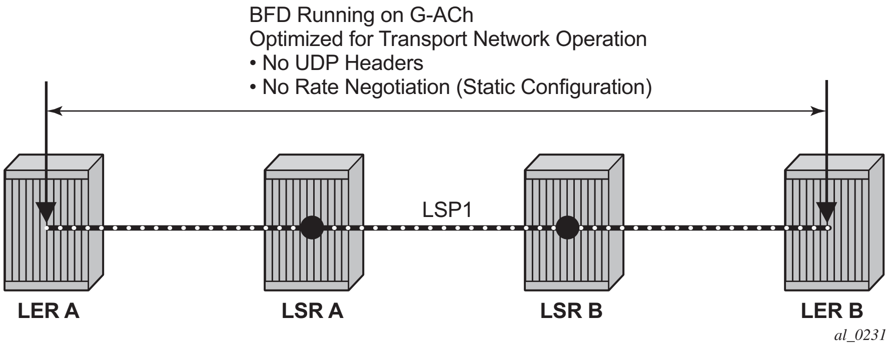
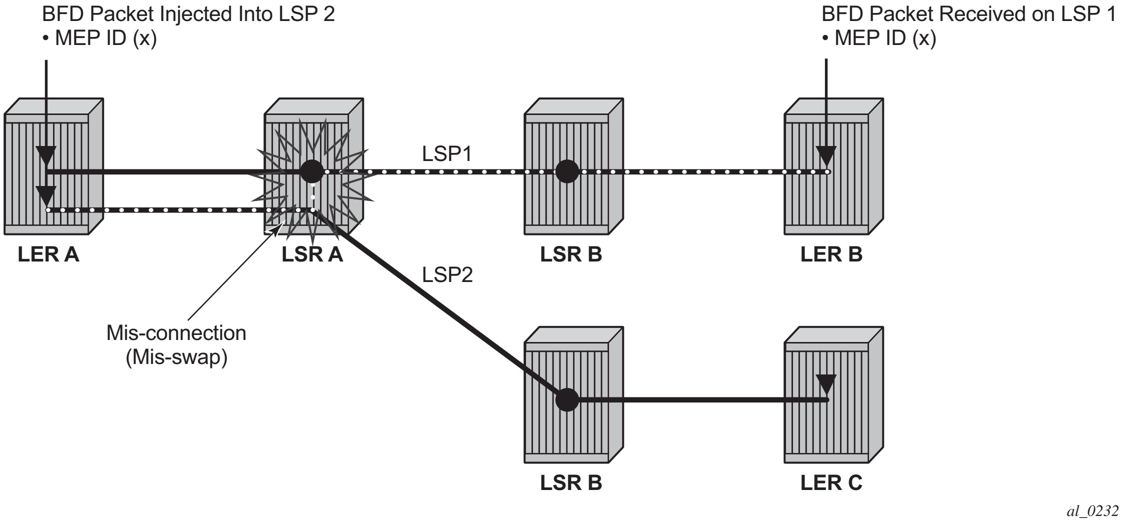

BFD Running on G-ACh
Optimized for Transport Network Operation
• No UDP Headers
• No Rate Negotiation (Static Configuration)
LSP1
LER A
LSR A
LSR B
LER B
al_0231

Figure 18 **BFD Used for Proactive CV on MPLS-TP LSP**



Linear protection switching of LSPs (see below) is triggered based on a CC or CV defect detected by BFD CC/CV.

RFC6428 defines two BFD session modes: Coordinated mode, in which the session state on both directions of the LSP is coordinated and constructed from a single, bidirectional BFD session, and independent mode, in which two independent sessions are bound together at a MEP. Coordinated mode is supported.

BFD is supported on MPLS-TP LSPs. When BFD_CV detects a mis-connectivity on an LSP, the system will drop all incoming non-OAM traffic with the LSP label (at the LSP termination point) instead of forwarding it to the associated SAP or PW segment.

The following GACh channel types are supported for the combined CC/CV mode:

- 0x22 for BFD CC with no IP encapsulation
- 0x23 for BFD CV

The following G-ACh channel types are used for the CC-only mode:

- 0x07

2.3.5.2.3 BFD-based RDI

RDI provides a mechanism whereby the source MEP can be informed of a downstream failure on an LSP, and can either raise an alarm, or initiate a protection switching operation. In the case of BFD based CC/CV, RDI is communicated using the BFD diagnostic field in BFD CC/CV messages. The following diagnostic codes are supported:

1 - Control Detection Time Expired

9 - mis-connectivity defect

2.3.5.3 PW Control Channel Status Notifications (Static Pseudowire Status Signaling)

MPLS-TP introduces the ability to support a full range of OAM and protection / redundancy on PWs for which no dynamic T-LDP control plane exists. Static PW status signaling is used to advertise the status of a PW with statically configured labels by encapsulating the PW status TLV in a G-ACh on the PW. This mechanism enables OAM message mapping and PW redundancy for such PWs, as defined in RFC6478. This mechanism is known as control channel status signaling in SR OS.

PW control channel status notifications use a similar model to T-LDP status signaling. That is, in general, status is always sent to the nearest neighbor T-PE or S-PE and relayed to the next segment by the S-PE. To achieve this, the PW label TTL is set to 1 for the G-ACh packet containing the status message.

Control channel status notifications are disabled by default on a spoke SDP. If they are enabled, then the default refresh interval is set to zero (although this value should be configurable in CLI). That is, when a status bit changes, three control channel status packets are sent consecutively at one-second intervals, and then the transmitter will fall silent. If the refresh timer interval is non-zero, then status messages will continue to be sent at that interval. The system supports the configuration of a refresh timer of 0, or from 10-65535 seconds. The recommended value is 600 seconds.

The system supports the optional acknowledgment of a PW control channel status message.

In order to constrain the CPU resources consumed processing control channel status messages, the system implements a credit-based mechanism. If a user enables control channel status on a PW[n], then a certain number of credits *c_n* are consumed from a CPM-wide pool of max_credit credits. The number of credits consumed is inversely proportional to the configured refresh timer (the first three messages at 1 second interval do not count against the credit). If the current_credit ≤ 0, then control channel status signaling cannot be configured on a PW (but the PW can still be configured and no shutdown).

If a PE with a non-zero refresh timer configured does not receive control channel status refresh messages for 3.5 time the specified timer value, then by default it will time out and assume a PW status of zero.

A trap is generated if the refresh timer times-out.

If PW redundancy is configured, the system will always consider the literal value of the PW status; a time-out of the refresh timer will not impact the choice of the active transit object for the VLL service. The result of this is that if the refresh timer times-out, and a given PW is currently the active PW, then the system will not fail-over to an alternative PW if the status is zero and some lower-layer OAM mechanism; for example, BFD has not brought down the LSP due to a connectivity defect. It is recommended that the PW refresh timer be configured with a much longer interval than any proactive OAM on the LSP tunnel, so that the tunnel can be brought down before the refresh timer expires if there is a CC defect.

A unidirectional continuity fault on a RSVP TE LSP may not result in the LSP being brought down before the received PW status refresh timer expires. It is therefore recommended that either bidirectional static MPLS-TP LSPs with BFD CC, or additional protection mechanisms; for example, FRR be used on RSVP-TE LSPs carrying MPLS-TP PWs. This is particularly important in active/standby PW dual homing configurations, where the active / standby forwarding state or operational state of every PW in the redundancy set must be accurately reflected at the redundant PE side of the configuration.

A PW with a refresh timer value of zero is always treated as having not expired.

The system implements a hold-down timer for control-channel-status PW-status bits in order to suppress bouncing of the status of a PW. For a specific spoke SDP, if the system receives 10 PW-status *change* events in 10 seconds, the system will *hold-down* the spoke SDP on the local node with the last received non-zero PW-status bits for 20 seconds. It will update the local spoke with the most recently received PW-status. This hold down timer is not persistent across shutdown/no-shutdown events.

2.3.5.4 PW Control Channel Status Request Mechanism

The system implements an optional PW control channel status request mechanism. This enhances the existing control channel status mechanism so that a peer that has *stale* PW status for the far-end of a PW can request that the peer PE send a static PW status update. Accurate and current information about the far end status of a PW is important for proper operation of PW redundancy. This mechanism ensures a consistent view of the control plane is maintained, as far as possible, between peer nodes. It is not intended to act as a continuity check between peer nodes.

2.3.5.5 Pseudowire Redundancy and Active / Standby Dual Homing

PW redundancy is supported for static MPLS-TP pseudowires. However, instead of using T-LDP status signaling to signal the forwarding state of a PW, control channel status signaling is used.

The following PW redundancy scenarios must be supported:

- MC-LAG and MC-APS with single and multi-segment PWs interconnecting the PEs.
- MS-PW (S-PE) Redundancy between VLL PEs with single-homed CEs.

- Dual-homing of a VLL service into redundant IES or VPRN PEs, with active/standby PWs.
- Dual-homing of a VLL service into a VPLS with active/standby PWs.

Active/standby dual-homing into routed VPLS is not supported in for MPLS-TP PWs. This is because it relies on PW label withdrawal of the standby PW in order to take down the VPLS instance, and hence the associated IP interface. Instead, it is possible to enable BGP multi-homing on a routed VPLS that has MPLS-TP PWs as spokes, and for the PW status of each spoke SDP to be driven (using control channel status) from the active or standby forwarding state assigned to each PW by BGP.

It is possible to configure inter-chassis backup (ICB) PWs as static MPLS-TP PWs with MPLS-TP identifiers. Only MPLS-TP PWs are supported in the same endpoint. That is, PWs in an endpoint must either be all MPLS-TP, or none of them must be MPLS-TP. This implies that an ICB used in an endpoint for which other PWs are MPLS TP must also be configured as an MPLS-TP PW.

A failover to a standby pseudowire is initiated based on the existing supported methods (for example, failure of the SDP).

2.3.5.6 Lock Instruct and Loopback for MPLS-TP Pseudowires

On the 7750 SR and 7450 ESS, the MPLS-TP supports lock instruct and loopback for PWs, including the ability to:

- administratively lock a spoke SDP with MPLS-TP identifiers
- divert traffic to and from an external device connected to a SAP
- create a data path loopback on the corresponding PW at a downstream S-PE or T-PE that was not originally bound to the spoke SDP being tested
- forward test traffic from an external test generator into an administratively locked PW, while simultaneously blocking the forwarding of user service traffic

MPLS-TP provides the ability to conduct test service throughput for PWs, through the configuration of a loopback on an administratively locked pseudowire. To conduct a service throughput test, an administrative lock is applied at each end of the PW. A test service that contains the SAP connected to the external device is used to inject test traffic into the PW. Lock request messaging is not supported.

A lock can be applied using the CLI or NMS. The forwarding state of the PW can be either active or standby.

After the PW is locked it can be put into loopback mode (for two way tests) so the ingress data path in the forward direction is cross connected to the egress data path in the reverse direction of the PW. The loopback can be configured through the CLI or NMS.

The PW loopback is created at the PW level, so everything under the PW label is looped back. This distinguishes a PW loopback from a service loopback, where only the native service packets are looped back.

The following MPLS-TP loopback configuration is supported:

- An MPLS-TP loopback can be created for an epipe, cpipe or apipe VLL.
- Test traffic can be inserted at an epipe, cpipe or apipe VLL endpoint or at an epipe spoke-sdp termination on a VPLS interface.

For more information about configuring lock instruct and loopback for MPLS-TP Pseudowires see, the *7450 ESS, 7750 SR, 7950 XRS, and VSR Services Overview Guide* and the *7450 ESS, 7750 SR, 7950 XRS, and VSR Layer 2 Services and EVPN Guide: VLL, VPLS, PBB, and EVPN*.

2.3.5.7 MPLS-TP LSP Protection

Linear 1-for-1 protection of MPLS-TP LSPs is supported, as defined in RFC. This applies only to LSPs (not PWs).

This is supported edge-to-edge on an LSP, between two LERs, where normal traffic is transported either on the working LSP or on the protection LSP using a logical selector bridge at the source of the protected LSP.

At the sink LER of the protected LSP, the LSP that carries the normal traffic is selected, and that LSP becomes the working LSP. A protection switching coordination (PSC) protocol coordinates between the source and sink bridge, which LSP is used, as working path and protection path. The PSC protocol is always carried on a G-ACh on the protection LSP.

The system supports single-phased coordination between the LSP endpoints, in which the initiating LER performs the protection switchover to the alternate path and informs the far-end LER of the switch.

Bidirectional protection switching is achieved by the PSC protocol coordinating between the two end points to determine which of the two possible paths (that is the working or protect path), transmits user traffic at any given time.

It is possible to configure non-revertive or revertive behavior. For non-revertive, the LSP will not switch back to the working path when the PSC switchover requests end, while for revertive configurations, the LSP always returns back to the working path when the switchover requests end.

The following figures illustrate the behavior of linear protection in more detail.

Figure 19 **Normal Operation**

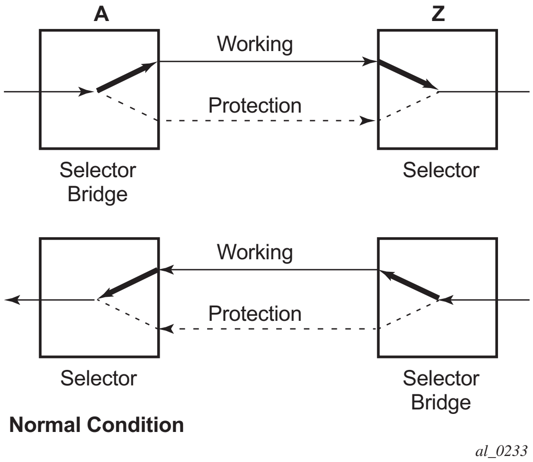
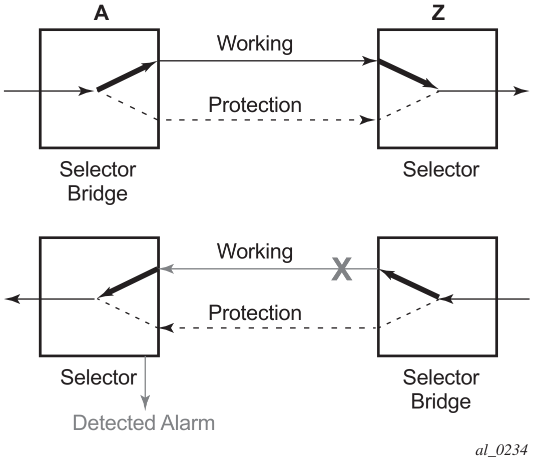


Figure 20 **Failed Condition**



In normal condition, user data packets are sent on the working path on both directions, from A to Z and Z to A.

A defect in the direction of transmission from node Z to node A impacts the working connection Z-to-A, and initiates the detection of a defect at the node A.

Figure 21 **Failed Condition - Switching at A**

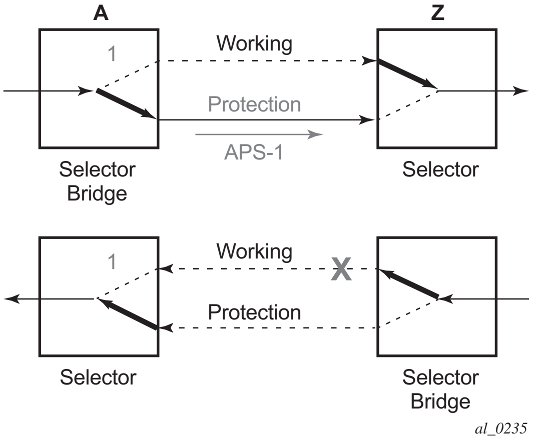
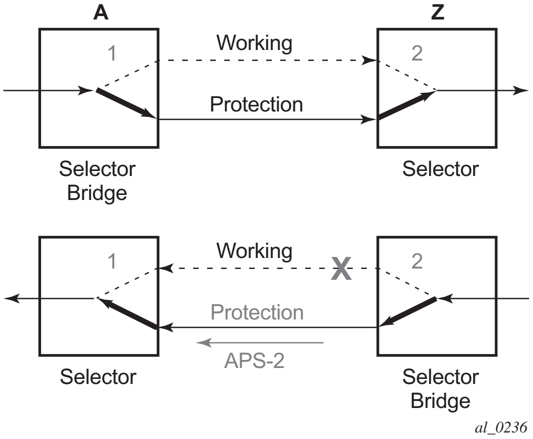


Figure 22 **Failed Condition - Switching at Z**



The unidirectional PSC protocol initiates protection switching: the selector bridge at node A is switched to protection connection A-to-Z and the selector at node A switches to protection connection Z to-A. The PSC packet, sent from node A to node Z, requests a protection switch to node Z.

After node Z validates the priority of the protection switch request, the selector at node Z is switched to protection connection A-to-Z and the selector bridge at the node Z is switched to protection connection Z-to-A. The PSC packet, sent from node Z to node A, is used as acknowledge, informing node A about the switching.

If BFD CC or CC/CV OAM packets are used to detect defects on the working and protection paths, they are inserted on both working and protection paths. Packets are sent whether or not the path is selected as the currently active path. Linear protection switching is also triggered on receipt of an AIS with the LDI bit set.

The following operator commands are supported:

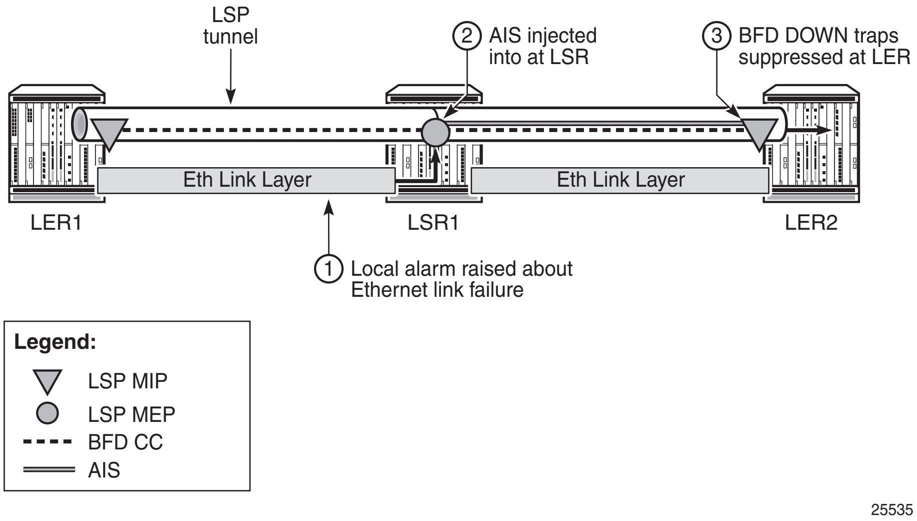
- Forced Switch
- Manual Switch
- Clear

2.3.6 Alarm Indication Signal (AIS)

When a MEP at a server layer (such as a link layer with respect to a given LSP) detects a failure, the server MEP notifies a co-located client layer of the condition. The client layer then generates Alarm Indication Signal (AIS) packets downstream in the client layer. These fault OAM messages are generated by intermediate nodes where a client LSP is switched, as per RFC 6427. This means that AIS packets are only inserted at an LSP MIP. AIS is used by the receiving MEP to suppress client layer traps caused by the upstream server layer failure; for example, if BFD CC is running on the LSP, then AIS will suppress the generation of multiple traps due to loss of CC.

Figure 23 illustrates an example of the operation of AIS in MPLS-TP.

Figure 23 **Example of AIS in MPLS-TP**


LSP tunnel
2 AIS injected into at LSR
3 BFD DOWN traps suppressed at LER
Eth Link Layer
Eth Link Layer
LER1
LSR1
LER2
1 Local alarm raised about Ethernet link failure
Legend:
LSP MIP
LSP MEP
BFD CC
AIS
25535

In the example, a failure of the Ethernet link layer between PE1 and LSR1 is detected at LSR1, which raises a local trap. LSPs transiting the LSR may be running CC OAM, such as BFD, and have AIS packets injected into them at LSR1. These AIS messages are received by the corresponding downstream MEP and processed. The failure of the Ethernet link between PE1 and LSR1 means that CC OAM on the LSPs is not received by the MEPs at PE2. Normally, this would cause multiple traps to be raised at PE2, but the reception of AIS causes PE2 to suppress the local generation of traps related to the failed LSP.

For traps to be suppressed successfully, the AIS message must arrive and be processed at the far-end PE or LER in sufficient time for the initial alarm to be suppressed. Therefore, the router implements a 2.5 secs hold-down timer for such traps on MPLS-TP LSPs.

Fault management for MPLS-TP, including AIS, is specified in RFC 6427.

The router supports:

- receiving and processing of AIS messages at LSP MEPs (at the LER)
- generation of AIS messages at LSP MIPs (at the LSR) in response to a failure of the ingress link
- suppression of SNMP traps indicating changes in the state of a BFD session, which result from the failure of the LSP data path upstream of a receiving LER; these traps would otherwise be sent to the 5620 SAM
- suppression of any BFD state machine Up/Down changes that occur while AIS is being received; there is no buffering or storage of state machine changes that occur during this period. This suppression only applies to Up/Down state change traps; other traps that would be expected are observed as normal.
- inclusion of the Link Down Indication (LDI) in an AIS message. This triggers a switchover of LSP linear protection if used on the LSP.
- insertion of AIS in the downstream direction of the transit path if a unidirectional fault is detected at an LSR. This suppresses CC traps at the downstream LER. However, the BFD session will still go down, causing RDI to be sent upstream in BFD, which will cause an alarm at the upstream LER.

2.3.7 Configuring MPLS-TP

This section describes the steps required to configure MPLS-TP.

2.3.7.1 Configuration Overview

The following steps must be performed in order to configure MPLS-TP LSPs or PWs.

At the router LER and LSR:

1. Create an MPLS-TP context, containing nodal MPLS-TP identifiers. This is configured under **config>router>mpls>mpls-tp**.
2. Ensure that a sufficient range of labels is reserved for static LSPs and PWs. This range is configured under **config>router>mpls-labels>static-label-range**.
3. Ensure that a range of tunnel identifiers is reserved for MPLS-TP LSPs under **config>router>mpls-mpls-tp>tp-tunnel-id-range**.
4. A user may optionally configure MPLS-TP interfaces, which are interfaces that do not use IP addressing or ARP for next hop resolution. These can only be used by MPLS-TP LSPs.

At the router LER, configure:

1. OAM Templates. These contain generic parameters for MPLS-TP proactive OAM. An OAM template is configured under **config>router>mpls>mpls-tp>oam-template**.
2. BFD templates. These contain generic parameters for BFD used for MPLS-TP LSPs. A BFD template is configured under **config>router>bfd>bfd-template**.
3. Protection templates. These contain generic parameters for MPLS-TP 1-for-1 linear protection. A protection template is configured under **config>router>mpls>mpls-tp>protection-template**.
4. MPLS-TP LSPs are configured under **config>router>mpls>lsp mpls-tp**
5. Pseudowires using MPLS-TP are configured as spoke SDPs with static PW labels.

At an LSR, a use must configure an LSP transit-path under **config>router>mpls>mpls-tp>transit-path**.

The following sections describe these configuration steps in more detail.

2.3.7.2 Node-Wide MPLS-TP Parameter Configuration

Generic MPLS-TP parameters are configured under **config>router>mpls>mpls-tp**. If a user configures **no mpls**, normally the entire MPLS configuration is deleted. However, in the case of MPLS-TP a check that there is no other MPLS-TP configuration; for example, services or tunnels using MPLS-TP on the node, is performed.

The MPLS-TP context is configured as follows:

```
config
   router
      mpls
         [no] mpls-tp
            . . .
            [no] shutdown
```

MPLS-TP LSPs may be configured if the MPLS-TP context is administratively down (shutdown), but they will remain down until the MPLS-TP context is configured as administratively up. No programming of the data path for an MPLS-TP path occurs until the following are all true:

- MPLS-TP context is **no shutdown**
- MPLS-TP LSP context is **no shutdown**
- MPLS-TP Path context is **no shutdown**

A **shutdown** of MPLS-TP will therefore bring down all MPLS-TP LSPs on the system.

The MPLS-TP context cannot be deleted if MPLS-TP LSPs or SDPs exist on the system.

2.3.7.3 Node-Wide MPLS-TP Identifier Configuration

MPLS-TP identifiers are configured for a node under the following CLI tree:

```
config
   router
     mpls
       mpls-tp
          global-id <global-id>
          node-id {<ipv4address> | | <1.. .4,294,967,295>}
          [no] shutdown
          exit
```

The default value for the global-id is 0. This is used if the global-id is not explicitly configured. If a user expects that inter domain LSPs are configured, then it is recommended that the global ID should be set to the local ASN of the node, as configured under **config>system**. If two-byte ASNs are used, then the most significant two bytes of the global-id are padded with zeros.

The default value of the node-id is the system interface IPv4 address. The MPLS-TP context cannot be administratively enabled unless at least a system interface IPv4 address is configured because MPLS requires that this value is configured.

These values are used unless overridden at the LSP or PW end-points, and apply only to static MPLS-TP LSPs and PWs.

In order to change the values, **config>router>mpls>mpls-tp** must be in the shutdown state. This will bring down all of the MPLS-TP LSPs on the node. New values are propagated to the system when a **no shutdown** is performed.

2.3.7.4 Static LSP and Pseudowire (VC) Label and Tunnel Ranges

The SR OS reserves a range of labels for use by static LSPs, and a range of labels for use by static pseudowires (SVCs) that is LSPs and pseudowires with no dynamic signaling of the label mapping. These are configured as follows:

```
config
   router
      mpls-labels
         [no] static-label max-lsp-labels <number>
static-svc-label <number>
```

<number>: indicates the maximum number of labels for the label type.

The minimum label value for the static LSP label starts at 32 and expands all the way to the maximum number specified. The static VC label range is contiguous with this. The dynamic label range exists above the static VC label range (the label ranges for the respective label type are contiguous). This prevents fragmentation of the label range.

The MPLS-TP tunnel ID range is configured as follows:

```
config
   router
      mpls
         mpls-tp
            [no] tp-tunnel-id-range <start-id> <end-id>
```

The tunnel ID range referred to here is a contiguous range of RSVP-TE Tunnel IDs is reserved for use by MPLS TP, and these IDs map to the MPLS-TP Tunnel Numbers. There are some cases where the dynamic LSPs may have caused fragmentation to the number space such that contiguous range {max-min} is not available. In these cases, the command will fail.

There is no default value for the tunnel id range, and it must be configured to enable MPLS-TP.

If a configuration of the tunnel ID range fails, then the system will give a reason. This could be that the initially requested range, or the change to the allocated range, is not available that is tunnel IDs in that range have already been allocated by RSVP-TE. Allocated Tunnel IDs are visible using a show command.

Changing the LSP or static VC label ranges does not require a reboot.

The static label ranges for LSPs, above, apply only to static LSPs configured using the CLI tree for MPLS-TP specified in this section. Different scalability constraints apply to static LSPs configured using the following CLI introduced in earlier SR OS releases:

config>router>mpls>static-lsp

config>router>mpls>if>label-map

The scalability applying to labels configured using this CLI is enforced as follows:

- A maximum of 1000 static LSP names may be configured with a PUSH operation.
- A maximum of 1000 LSPs with a POP or SWAP operation may be configured.

These two limits are independent of one another, giving a combined limit of 1000 PUSH and 1000 POP/SAP operations configured on a node.

The static LSP and VC label spaces are contiguous. Therefore, the dimensioning of these label spaces requires careful planning by an operator as increasing the static LSP label space impacts the start of the static VC label space, which may already-deployed

2.3.7.5 Interface Configuration for MPLS-TP

It is possible for MPLS-TP paths to use both numbered IP numbered interfaces that use ARP/static ARP, or IP unnumbered interfaces. MPLS-TP requires no changes to these interfaces. It is also possible to use a new type of interface that does not require any IP addressing or next-hop resolution.

RFC 7213 provides guidelines for the usage of various Layer 2 next-hop resolution mechanisms with MPLS-TP. If protocols such as ARP are supported, then they should be used. However, in the case where no dynamic next hop resolution protocol is used, it should be possible to configure a unicast, multicast or broadcast next-hop MAC address. The rationale is to minimize the amount of configuration required for upstream nodes when downstream interfaces are changes. A default multicast MAC address for use by MPLS-TP point-to-point LSPs has been assigned by IANA (Value: 01-00-5e-90-00-00). This value is configurable on the router to support interoperability with third-party implementations that do not default to this value, and this no default value is implemented on the router.

In order to support these requirements, a new interface type, known as an unnumbered MPLS-TP interface is introduced. This is an unnumbered interface that allows a broadcast or multicast destination MAC address to be configured. An unnumbered MPLS-TP interface is configured using the **unnumbered-mpls-tp** keyword, as follows:

```
config
   router
      interface <if-name> [unnumbered-mpls-tp]
         port <port-id>[:encap-val]
         mac <local-mac-address>
         static-arp <remote-mac-addr>
         //ieee-address needs to support mcast and bcast
                  exit
```

The **remote-mac-address** may be any unicast, broadcast of multicast address. However, a broadcast or multicast remote-mac-address is only allowed in the **static-arp** command on Ethernet unnumbered interfaces when the **unnumbered-mpls-tp** keyword has been configured. This also allows the interface to accept packets on a broadcast or any multicast MAC address. If a packet is received with a unicast destination MAC address, then it is checked against the configured <local-mac-address> for the interface, and dropped if it does not match. When an interface is of type **unnumbered-mpls-tp**, only MPLS-TP LSPs are allowed on that interface; other protocols are blocked from using the interface.

An unnumbered MPLS-TP interface is assumed to be point-to-point, and therefore users must ensure that the associated link is not broadcast or multicast in nature if a multicast or broadcast remote MAC address is configured.

The following is a summary of the constraints of an unnumbered MPLS-TP interface:

- It is unnumbered and may borrow/use the system interface address
- It prevents explicit configuration of a borrowed address
- It prevents IP address configuration
- It prevents all protocols except mpls
- It prevents deletion if an MPLS-TP LSP is bound to the Interface

MPLS-TP is only supported over Ethernet ports. The system will block the association of an MPLS-TP LSP to an interface whose port is non-Ethernet.

If required, the IF_Num is configured under a MEP context under the MPLS interface. The **mpls-tp-mep** context is created under the interface as shown below. The *if-num* parameter, when concatenated with the Node ID, forms the IF_ID (as per RFC 6370), which is the identifier of this MEP. It is possible to configure this context whether the interface is IP numbered, IP unnumbered, or MPLS-TP unnumbered:

```
config
   router
     mpls
      interface <ip-int-name>
          mpls-tp-mep
            [no] ais-enable
            [no] if-num <if-num>
            [no] if-num-validation [enable | disable]
                    ...
 exit
```

The **if-num-validation** command is used to enable or disable validation of the if-num in LSP Trace packet against the locally configured if-num for the interface over which the LSP Trace packet was received at the egress LER. This is because some implementations do not perform interface validation for unnumbered MPLS-TP interfaces and instead set the if-num in the DSMAP TLV to 0. The default is enabled.

AIS insertion is configured using the **ais-enable** command under the **mpls-tp-mep** context on an MPLS interface.

2.3.7.6 LER Configuration for MPLS-TP

2.3.7.6.1 LSP and Path Configuration

MPLS-TP tunnels are configured using the **mpls-tp** LSP type at an LER under the LSP configuration, using the following CLI tree:

```
config
   router
      mpls
         lsp <xyz> [bypass-only | p2mp-lsp | mpls-tp <src-tunnel-num>]
            to node-id {<a.b.c.d> | <1.. .4,294,967,295>}
            dest-global-id <global-id>
             dest-tunnel-number <tunnel-num>
             [no] working-tp-path
                lsp-num <lsp-num>
                in-label <in-label>
                out-label <out-label> out-link <if-name>
                           [next-hop <ipv4-address>]
                [no] mep
                   [no] bfd-enable [cc | cc-cv]
                   [no] bfd-trap-suppression
            [no] oam-template <name>
            [no] shutdown
            exit
        [no] shutdown
        exit
    [no] protect-tp-path
        lsp-num <lsp-num>
        in-label <in-label>
        out-label <out-label> out-link <if-name>
                  [next-hop <ipv4-address> ]
        [no] mep
            [no] bfd-enable [cc | cc-cv]
            [no] bfd-trap-suppression
            [no] oam-template <name>
            [no] protection-template <name>
            [no] shutdown
            exit
        [no] shutdown
        exit
```

<*if-name*> could be numbered or unnumbered interface using an Ethernet port.

<*src-tunnel-num*> is a mandatory create time parameter for mpls-tp tunnels, and has to be assigned by the user based on the configured range of tunnel ids. The *src-global-id* used for the LSP ID is derived from the node-wide *global-id* value configured under config>router>mpls>mpls-tp. A tunnel can not be brought up unless the *global-id* is configured.

The from address of an LSP to be used in the tunnel identifier is taken to be the local node's node-id/global-id, as configured under config>router>mpls>mpls-tp. If that is not explicitly configured, either, then the default value of the system interface IPv4 address is used

The **to node-id** address may be entered in 4-octet IPv4 address format or unsigned 32-bit format. This is the far-end node-id for the LSP, and does do need to be routable IP addresses.

The **from** and **to** addresses are used as the from and to node-id in the MPLS-TP Tunnel Identifier used for the MEP ID.

Each LSP consists of a working-tp-path and, optionally, a protect-tp-path. The protect-tp-path provides protection for the working-tp-path is 1:1 linear protection is configured (see below). Proactive OAM, such as BFD, is configured under the MEP context of each path. Protection for the LSP is configured under the protect-tp-path MEP context.

The *to* global-id is an optional parameter. If it is not entered, then the destination global ID takes the default value of 0. Global ID values of 0 are allowed and indicate that the node's configured Global ID should be used. If the local global ID value is 0, then the remote **to** global ID must also be 0. The *to* global ID value cannot be changed if an LSP is in use by an SDP.

The *to* tunnel number is an optional parameter. If it is not entered, then it is taken to be the same value as the source tunnel number.

LSPs are assumed to be bidirectional and co-routed. Therefore, the system will assume that the incoming interface is the same as the out-link.

The next-hop *ip-address* can only be configured if the out-link if-name refers to a numbered IP interface. In this case, the system will determine the interface to use to reach the configured next-hop, but will check that the user-entered value for the out-link corresponds to the link returned by the system. If they do not correspond, then the path will not come up. If a user changes the physical port referred to in the interface configuration, BFD, if configured on the LSP, will go down. Users must ensure that an LSP is moved to a different interface with a different port configuration in order to change the port that it uses. This is enforced by blocking the next-hop configuration for an unnumbered interface.

There is no check made that a valid ARP entry exists before allowing a path to be un shut. Therefore, a path is only held down if BFD is down. If static ARP is not configured for the interface, then it is assumed that dynamic ARP is used. The result is that if BFD is not configured, a path can come up before ARP resolution has completed for an interface. If BFD is not used, then it is recommended that the connectivity of the path is explicitly checked using on-demand CC/CV prior to sending user traffic on it.

The following is a list of additional considerations for the configuration of MPLS-TP LSPs and paths:

- The **working-tp-path** must be configured before the **protect-tp-path**.
- Likewise, the **protect-tp-path** must be deleted first before the **working-tp-path**.
- The *lsp-num* parameter is optional. The default values are 1 for the working-tp-path and 2 for protect-tp-path.
- The **mep** context must be deleted before a path can be deleted.
- An MPLS interface needs to be created under **config>router>mpls>interface** before using/specifying the out-label/out-link in the Forward path for an MPLS-TP LSP. Creation of the LSP fails if the corresponding mpls interface does not exist even though the specified router interface may be valid.
- The system programs the MPLS-TP LSP information upon a **no shutdown** command of the TP-Path only on the very first **no shutdown**. The **working-tp-path** is programmed as the primary and the **protect-tp-path** is programmed as the backup.

- The system will not deprogram the IOM on an admin shutdown of the MPLS-TP path. Traffic will gracefully move to the other TP-Path if valid, as determined by the proactive MPLS-TP OAM. This should not result in traffic loss. However it is recommended that the user does moves traffic to the other TP-Path through a tools command before performing the **admin shutdown** command of an Active TP-Path.
- Deletion of the out-label/out-link sub-command under the MPLS-TP Path is not allowed once configured. These can only be modified.
- MPLS allows the deletion of an **admin shutdown** TP-Path. This causes MPLS to deprogram the corresponding TP-Path forwarding information from IOM. This can cause traffic loss for certain users that are bound to the MPLS-TP LSP.
- MPLS will not deprogram the IOM on a specific interface admin shut/clear unless the interface is a System Interface. However, if MPLS informs the TP-OAM module that the MPLS interface has gone down, then it triggers a switch to the standby tp-path if the associated interface went down and if it is valid.
- If a MEP is defined and shut down, the corresponding path is also operationally down. The MEP admin state is applicable only when a MEP is created from an MPLS-TP path.
- It is not mandatory to configure BFD or protection on an MPLS-TP path in order to bring the LSP up.
- If **bfd-enable cc** is configured, then CC-only mode using ACh channel 0x07 is used. If **bfd-enable cc_v** is configured, then BFD CC packets use channel 0x22 and CV packets use channel 0x23.
- Under the MEP context, the **bfd-trap-suppression** command allows the reception of AIS packets on the path to suppress BFD Down traps if a BFD session goes down on that path.

The protection template is associated with an LSP as a part of the MEP on the protect path. If only a working path is configured, then the protection template is not configured.

BFD cannot be enabled under the MEP context unless a named BFD template is configured.

2.3.7.6.2 Support for Downstream Mapping Information

In order to validate the downstream mapping for an LSP, a node sending a DSMAP TLV must include the incoming and (optionally) outgoing IF_Num values for the interfaces that it expects the LSP to transit. Additionally, it will include the out-label for the LSP in the Label TLV for the DSMAP in the echo request message.

The incoming and outgoing if-num values correspond to the incoming and outgoing interfaces transited by an LSP at the next hop LER and LSR are configured using the **dsmap** command, as follows:

```
config
   router
      mpls
         lsp
            working-tp-path
               mep
                 dsmap <in-if-num>[:<out-if-num>]

config
   router
      mpls
         lsp
            protect-tp-path
               mep
                  dsmap <in-if-num>[:<out-if-num>]


config
   router
      mpls
         mpls-tp
            transit-path
               forward-path
                  mip
                     dsmap <in-if-num>[:<out-if-num>]
                     exit
                reverse-path
                   mip
                      dsmap <in-if-num>[:<out-if-num>]
                      exit
```

A node sending a DSMAP TLV includes these **in-if-num** and **out-if-num** (if configured) values. Additionally, it includes the out-label for the LSP in the Label TLV for the DSMAP in the echo request message.

2.3.7.6.3 Proactive CC/CV (using BFD) Configuration

Generally applicable proactive OAM parameters are configured using templates.

Proactive CC and CV uses BFD parameters such as Tx/Rx timer intervals, multiplier and other session/fault management parameters which are specific to BFD. These are configured using a BFD Template. The BFD Template may be used for non-MPLS-TP applications of BFD, and therefore contains the full set of possible configuration parameters for BFD. Only a sub-set of these may be used for any given application.

Generic MPLS-TP OAM and fault management parameters are configured in the OAM Template.

Named templates are referenced from the MPLS-TP Path MEP configuration, so different parameter values are possible for the working and protect paths of a tunnel.

The BFD Template is configured as follows:

```
config
   router
      bfd
         [no] bfd-template <name>
            [no] transmit-interval <transmit-interval>
            [no] receive-interval <receive-interval>
            [no] echo-receive <echo-interval>
            [no] multiplier <multiplier>
            [no] type <cpm-np>
            exit
```

The parameters are as follows:

- **transmit-interval** *transmit-interval* and the **rx** *receive-interval*: These are the transmit and receive timers for BFD packets. If the template is used for MPLS-TP, then these are the timers used by CC packets. Values are in ms: 10 ms to 100 000 ms, with 1ms granularity. Default 10ms for CPM3 or better, 1 sec for other hardware. For MPLS-TP CV packets, a transmit interval of 1 s is always used.
- **multiplier** *multiplier*: Integer 3 to 20. Default: 3. This parameter is ignored for MPLS-TP combined cc-v BFD sessions, and the default of 3 used, as per RFC6428.
- **echo-receive** *echo-interval*: Sets the minimum echo receive interval (in ms), for a session. Values: 100 ms to 100 000 ms. Default: 100. This parameter is not used by a BFD session for MPLS-TP.
- **type cpm-np**: This selects the CPM network processor as the local termination point for the BFD session. This is enabled by default.

If the BFD timer values as shown above are changed in a template, any BFD sessions on MEPs to which that template is bound will try to renegotiate their timers to the new values.

Caution: The BFD implementations in some MPLS-TP peer nodes may not be able handle renegotiation, as allowed by Section 3.7.1 of RFC6428, and may take the BFD session down. This can result in undesired behavior, such as an unexpected protection switching event. We recommend that users of the system exercise caution when modifying the BFD timer values after a BFD session is up.

Commands within the BFD-template use a begin-commit model. To edit any value within the BFD template, a *begin* needs to be executed once the template context has been entered. However, a value will still be stored temporarily until the commit is issued. Once the commit is issued, values will actually be used by other modules like the mpls-tp module and BFD module.

A BFD template is referenced from the OAM template. The OAM Template is configured as follows:

```
config
   router
      mpls
         mpls-tp
            [no] oam-template <name>
               [no] bfd-template <name>
               [no] hold-time-down <interval>
               [no] hold-time-up <interval>
            exit
```

- **hold-time-down** *interval*: 0-5000 deciseconds, 10ms steps, default 0. This is equivalent to the standardized hold-off timer.
- **hold-time-up** *interval*: 0-500 centiseconds in 100ms steps, default 2 seconds This is an additional timer that can be used to reduce BFD bouncing.
- **bfd-template** *name*: This is the named BFD template to use for any BFD sessions enabled under a MEP for which the OAM template is configured.

An OAM template is then applied to a MEP as described above.

2.3.7.6.4 Protection templates and Linear Protection Configuration

Protection templates defines the generally applicable protection parameters for an MPLS-TP tunnel. Only linear protection is supported, and so the application of a named template to an MPLS-TP tunnel implies that linear protection is used.

A template is configured as follows:

```
config
   router
      mpls
         mpls-tp
            protection-template <name>
              [no] revertive
               [no] wait-to-restore <interval>
               rapid-psc-timer <interval>
               slow-psc-timer <interval>
               exit
```

The allowed values are as follows:

- **wait-to-restore** *interval*: 0-720 seconds, 1 sec steps, default 300 seconds. This is applicable to revertive mode only.
- **rapid-psc-timer** *interval*: [10, 100, 1000ms]. Default 100ms
- **slow-psc-timer** *interval*: 5s-60s. Default: 5s
- **revertive**: Selects revertive behavior. Default: no revertive.

LSP Linear Protection operations are enacted using the following **tools>perform** commands.

```
tools>perform>router>mpls
            tp-tunnel
               clear {<lsp-name> | id <tunnel-id>}
               force {<lsp-name> | id <tunnel-id>}
               lockout {<lsp-name> | id <tunnel-id>}
               manual {<lsp-name> | id <tunnel-id>}
            exit
         exit
```

To minimize outage times, users should use the “mpls-tp protection command” (for example, force/manual) to switch all the relevant MPLS-TP paths before executing the following commands:

- clear router mpls interface <>
- config router mpls interface <> shut

2.3.7.7 Intermediate LSR Configuration for MPLS-TP LSPs

The forward and reverse directions of the MPLS-TP LSP Path at a transit LSR are configured using the following CLI tree:

```
config
   router
      mpls
         mpls-tp
            transit-path <path-name>
               [no] path-id {lsp-num <lsp-num> | working-path | protect-path
                  [src-global-id <global-id>]
                  src-node-id {<ipv4address> | <1.. .4,294,967,295>}
                  src-tunnel-num <tunnel-num>
                  [dest-global-id <global-id>]
                  dest-node-id {<ipv4address> | <1.. .4,294,967,295>}
                  [dest-tunnel-num <tunnel-num>]}

               forward-path
                  in-label <in-label> out-label <out-label>
                       out-link <if-name> [next-hop <ipv4-next-hop>]
               reverse-path
                  in-label <in-label> out-label <out-label>
                      [out-link <if-name> [next-hop <ipv4-next-hop>]
```

```
                    [no] shutdown
```

The *src-tunnel-num* and *dest-tunnel-num* are consistent with the source and destination of a label mapping message for a signaled LSP.

If *dest-tunnel-num* is not entered in CLI, the *dest-tunnel-num* value is taken to be the same as the SRC-tunnel-num value.

If any of the *global-id* values are not entered, the value is taken to be 0.

If the *src-global-id* value is entered, but the *dest-global-id* value is not entered, *dest-global-id* value is the same as the *src-global-id* value.

The *lsp-num* must match the value configured in the LER for a given path. If no explicit lsp-num is configured, then working-path or protect-path must be specified (equating to 1 or 2 in the system).

The forward path must be configured before the reverse path. The configuration of the reverse path is optional.

The LSP-ID (path-id) parameters apply with respect to the downstream direction of the forward LSP path, and are used to populate the MIP ID for the path at this LSR.

The reverse path configuration must be deleted before the forward path.

The forward-path (and reverse-path if applicable) parameters can be configured with or without the path-id, but they must be configured if MPLS-TP OAM is to be able to identify the LSR MIP.

The transit-path can be no shutdown (as long as the forward-path/reverse-path parameters have been configured properly) with or without identifiers.

The path-id and path-name must be unique on the node. There is a one to one mapping between a given path-name and path-id.

Traffic can not pass through the transit-path if the transit-path is in the **shutdown** state.

2.3.8 MPLS-TP Show Commands

2.3.8.1 Static MPLS Labels

The following new commands show the details of the static MPLS labels.

show>router>mpls-labels>label <start-label> [<end-label> [in-use | <label-owner>]]

show>router>mpls-labels>label-range

An example output is as follows:

```
*A:mlstp-dutA# show router mpls
mpls         mpls-labels
*A:mlstp-dutA# show router mpls label
label        label-range
*A:7950 XRS-20# 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           505856
=======================================================================
```

2.3.8.2 MPLS-TP Tunnel Configuration

These commands show the configuration of a given tunnel.

show>router>mpls>tp-lsp

A sample output is as follows:

```
*A:mlstp-dutA# show router mpls tp-lsp
  - tp-lsp [<lsp-name>] [status {up | down}] [from <ip-address> | to <ip-address>]
    [detail]
  - tp-lsp [<lsp-name>] path [protect | working] [detail]
  - tp-lsp [<lsp-name>] protection

 <lsp-name>           : [32 chars max] - accepts * as wildcard char
 <path>               : keyword - Display LSP path information.
 <protection>         : keyword - Display LSP protection information.
 <up | down>          : keywords - Specify state of the LSP
 <ip-address>         : a.b.c.d
 <detail>             : keyword - Display detailed information.

*A:mlstp-dutA# show router mpls tp-lsp
path
protection
to <a.b.c.d>
<lsp-name>
 "lsp-32"  "lsp-33"  "lsp-34"  "lsp-35"  "lsp-36"  "lsp-37"  "lsp-38"  "lsp-39"
 "lsp-40"  "lsp-41"
status {up | down}
from <ip-address>
```

```
detail

*A:mlstp-dutA# show router mpls tp-lsp "lsp-
"lsp-32"  "lsp-33"  "lsp-34"  "lsp-35"  "lsp-36"  "lsp-37"  "lsp-38"  "lsp-39"
"lsp-40"  "lsp-41"

*A:mlstp-dutA# show router mpls tp-lsp "lsp-32"

===============================================================================
MPLS MPLS-TP LSPs (Originating)
===============================================================================
LSP Name                           To               Tun     Protect   Adm  Opr
                                                    Id      Path
-------------------------------------------------------------------------------
lsp-32                             0.0.3.234        32      No        Up   Up
-------------------------------------------------------------------------------
LSPs : 1
===============================================================================

*A:mlstp-dutA# show router mpls tp-lsp "lsp-32" detail

===============================================================================
MPLS MPLS-TP LSPs (Originating) (Detail)
===============================================================================
-------------------------------------------------------------------------------
Type : Originating
-------------------------------------------------------------------------------
LSP Name    : lsp-32
LSP Type    : MplsTp                         LSP Tunnel ID  : 32
From Node Id: 0.0.3.233+                     To Node Id     : 0.0.3.234
Adm State   : Up                             Oper State     : Up
LSP Up Time : 0d 04:50:47                    LSP Down Time  : 0d 00:00:00
Transitions : 1                              Path Changes   : 2

DestGlobalId: 42                             DestTunnelNum  : 32
```

2.3.8.3 MPLS-TP Path configuration

This can reuse and augment the output of the current show commands for static LSPs. They should also show if BFD is enabled on a given path. If this referring to a transit path, this should also display (among others) the path-id (7 parameters) for a given transit-path-name, or the transit-path-name for a given the path-id (7 parameters)

show>router>mpls>tp-lsp>path

A sample output is as follows:

```
===============================================================================
*A:mlstp-dutA#  show router mpls tp-lsp path

===============================================================================
MPLS-TP LSP Path Information
```

```
===============================================================================
LSP Name      : lsp-32                           To            : 0.0.3.234
Admin State   : Up                               Oper State    : Up

-------------------------------------------------------------------------------
Path        NextHop           InLabel   OutLabel  Out I/F         Admin  Oper
-------------------------------------------------------------------------------
Working                       32        32        AtoB_1          Up     Down
Protect                       2080      2080      AtoC_1          Up     Up
===============================================================================
LSP Name      : lsp-33                           To            : 0.0.3.234
Admin State   : Up                               Oper State    : Up

-------------------------------------------------------------------------------
Path        NextHop           InLabel   OutLabel  Out I/F         Admin  Oper
-------------------------------------------------------------------------------
Working                       33        33        AtoB_1          Up     Down
Protect                       2082      2082      AtoC_1          Up     Up
===============================================================================
LSP Name      : lsp-34                           To            : 0.0.3.234
Admin State   : Up                               Oper State    : Up

-------------------------------------------------------------------------------
Path        NextHop           InLabel   OutLabel  Out I/F         Admin  Oper
-------------------------------------------------------------------------------
Working                       34        34        AtoB_1          Up     Down
Protect                       2084      2084      AtoC_1          Up     Up
===============================================================================
LSP Name      : lsp-35                           To            : 0.0.3.234
Admin State   : Up                               Oper State    : Up

-------------------------------------------------------------------------------
Path        NextHop           InLabel   OutLabel  Out I/F         Admin  Oper
-------------------------------------------------------------------------------
Working                       35        35        AtoB_1          Up     Down
Protect                       2086      2086      AtoC_1          Up     Up
===============================================================================
LSP Name      : lsp-36                           To            : 0.0.3.234
Admin State   : Up                               Oper State    : Up

-------------------------------------------------------------------------------
Path        NextHop           InLabel   OutLabel  Out I/F         Admin  Oper
-------------------------------------------------------------------------------
Working                       36        36        AtoB_1          Up     Down
Protect                       2088      2088      AtoC_1          Up     Up
===============================================================================
LSP Name      : lsp-37                           To            : 0.0.3.234
Admin State   : Up                               Oper State    : Up

-------------------------------------------------------------------------------
Path        NextHop           InLabel   OutLabel  Out I/F         Admin  Oper
-------------------------------------------------------------------------------
Working                       37        37        AtoB_1          Up     Down
Protect                       2090      2090      AtoC_1          Up     Up
===============================================================================
LSP Name      : lsp-38                           To            : 0.0.3.234
Admin State   : Up                               Oper State    : Up

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

```
Path        NextHop           InLabel   OutLabel  Out I/F         Admin  Oper
-------------------------------------------------------------------------------
Working                       38        38        AtoB_1          Up     Down
Protect                       2092      2092      AtoC_1          Up     Up
===============================================================================
LSP Name      : lsp-39                           To            : 0.0.3.234
Admin State   : Up                               Oper State    : Up

-------------------------------------------------------------------------------
Path        NextHop           InLabel   OutLabel  Out I/F         Admin  Oper
-------------------------------------------------------------------------------
Working                       39        39        AtoB_1          Up     Down
Protect                       2094      2094      AtoC_1          Up     Up
===============================================================================
LSP Name      : lsp-40                           To            : 0.0.3.234
Admin State   : Up                               Oper State    : Up

-------------------------------------------------------------------------------
Path        NextHop           InLabel   OutLabel  Out I/F         Admin  Oper
-------------------------------------------------------------------------------
Working                       40        40        AtoB_1          Up     Down
Protect                       2096      2096      AtoC_1          Up     Up
===============================================================================
LSP Name      : lsp-41                           To            : 0.0.3.234
Admin State   : Up                               Oper State    : Up

-------------------------------------------------------------------------------
Path        NextHop           InLabel   OutLabel  Out I/F         Admin  Oper
-------------------------------------------------------------------------------
Working                       41        41        AtoB_1          Up     Down
Protect                       2098      2098      AtoC_1          Up     Up

*A:mlstp-dutA#  show router mpls tp-lsp "lsp-32" path working

===============================================================================
MPLS-TP LSP Working Path Information
    LSP: "lsp-32"
===============================================================================
LSP Name      : lsp-32                           To            : 0.0.3.234
Admin State   : Up                               Oper State    : Up

-------------------------------------------------------------------------------
Path        NextHop           InLabel   OutLabel  Out I/F         Admin  Oper
-------------------------------------------------------------------------------
Working                       32        32        AtoB_1          Up     Down
===============================================================================
*A:mlstp-dutA#  show router mpls tp-lsp "lsp-32" path protect

===============================================================================
MPLS-TP LSP Protect Path Information
    LSP: "lsp-32"
===============================================================================
LSP Name      : lsp-32                           To            : 0.0.3.234
Admin State   : Up                               Oper State    : Up

-------------------------------------------------------------------------------
Path        NextHop           InLabel   OutLabel  Out I/F         Admin  Oper
-------------------------------------------------------------------------------
Protect                       2080      2080      AtoC_1          Up     Up
```

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

*A:mlstp-dutA#  show router mpls tp-lsp "lsp-32" path protect detail

===============================================================================
MPLS-TP LSP Protect Path Information
    LSP: "lsp-32" (Detail)
===============================================================================
LSP Name      : lsp-32                           To            : 0.0.3.234
Admin State   : Up                               Oper State    : Up

Protect path information
-------------------------------------------------------------------------------
Path Type     : Protect                          LSP Num       : 2
Path Admin    : Up                               Path Oper     : Up
Out Interface : AtoC_1                           Next Hop Addr : n/a
In Label      : 2080                             Out Label     : 2080
Path Up Time  : 0d 04:52:17                      Path Dn Time  : 0d 00:00:00
Active Path   : Yes                              Active Time   : 0d 00:52:56

MEP information
MEP State     : Up                               BFD           : cc
OAM Templ     : privatebed-oam-template          CC Status     : inService
                                                 CV Status     : unknown
Protect Templ : privatebed-protection-template   WTR Count Down: 0 seconds
RX PDU        : SF (1,1)                         TX PDU        : SF (1,1)
Defects       :
===============================================================================

*A:mlstp-dutA#  show router mpls tp-lsp "lsp-32" path working detail

===============================================================================
MPLS-TP LSP Working Path Information
    LSP: "lsp-32" (Detail)
===============================================================================
LSP Name      : lsp-32                           To            : 0.0.3.234
Admin State   : Up                               Oper State    : Up

Working path information
-------------------------------------------------------------------------------
Path Type     : Working                          LSP Num       : 1
Path Admin    : Up                               Path Oper     : Down
Down Reason   : ccFault ifDn
Out Interface : AtoB_1                           Next Hop Addr : n/a
In Label      : 32                               Out Label     : 32
Path Up Time  : 0d 00:00:00                      Path Dn Time  : 0d 00:53:01
Active Path   : No                               Active Time   : n/a

MEP information
MEP State     : Up                               BFD           : cc
OAM Templ     : privatebed-oam-template          CC Status     : outOfService
                                                 CV Status     : unknown
===============================================================================
*A:mlstp-dutA#
```

2.3.8.4 MPLS-TP Protection

The following output shows the protection configuration for a given tunnel, which path in a tunnel is currently working and which is protect, and whether the working or protect is currently active.

show>router>mpls>tp-lsp>protection

A sample output is as follows:

```
*A:mlstp-dutA#  show router mpls tp-lsp protection

===============================================================================
MPLS-TP LSP Protection Information
Legend: W-Working, P-Protect,
===============================================================================
LSP Name                      Admin Oper  Path    Ingr/Egr      Act. Rx PDU
                              State State   State Label         Path Tx PDU
-------------------------------------------------------------------------------
lsp-32                        Up    Up    W Down      32/32     No   SF (1,1)
                                          P Up      2080/2080   Yes  SF (1,1)
lsp-33                        Up    Up    W Down      33/33     No   SF (1,1)
                                          P Up      2082/2082   Yes  SF (1,1)
lsp-34                        Up    Up    W Down      34/34     No   SF (1,1)
                                          P Up      2084/2084   Yes  SF (1,1)
lsp-35                        Up    Up    W Down      35/35     No   SF (1,1)
                                          P Up      2086/2086   Yes  SF (1,1)
lsp-36                        Up    Up    W Down      36/36     No   SF (1,1)
                                          P Up      2088/2088   Yes  SF (1,1)
lsp-37                        Up    Up    W Down      37/37     No   SF (1,1)
                                          P Up      2090/2090   Yes  SF (1,1)
lsp-38                        Up    Up    W Down      38/38     No   SF (1,1)
                                          P Up      2092/2092   Yes  SF (1,1)
lsp-39                        Up    Up    W Down      39/39     No   SF (1,1)
                                          P Up      2094/2094   Yes  SF (1,1)
lsp-40                        Up    Up    W Down      40/40     No   SF (1,1)
                                          P Up      2096/2096   Yes  SF (1,1)
lsp-41                        Up    Up    W Down      41/41     No   SF (1,1)
                                          P Up      2098/2098   Yes  SF (1,1)
-------------------------------------------------------------------------------
No. of MPLS-TP LSPs: 10
===============================================================================
```

2.3.8.5 MPLS TP Node Configuration

The following output shows the Global ID, Node ID and other general MPLS-TP configurations for the node.

show>router>mpls>mpls-tp

A sample output is as follows:

```
*A:mlstp-dutA# show router mpls mpls-tp
  - mpls-tp


      oam-template    - Display MPLS-TP OAM Template information
      protection-tem* - Display MPLS-TP Protection Template information
      status          - Display MPLS-TP system configuration
      transit-path    - Display MPLS-TP Tunnel information

*A:mlstp-dutA# show router mpls mpls-tp oam-template

===============================================================================
MPLS-TP OAM Templates
===============================================================================
Template Name : privatebed-oam-template Router ID     : 1
BFD Template  : privatebed-bfd-template Hold-Down Time: 0 centiseconds
                                        Hold-Up Time  : 20 deciseconds
===============================================================================

*A:mlstp-dutA# show router mpls mpls-tp protection-template

===============================================================================
MPLS-TP Protection Templates
===============================================================================
Template Name  : privatebed-protection-template  Router ID      : 1
Protection Mode: one2one                         Direction      : bidirectional
Revertive      : revertive                       Wait-to-Restore: 300sec
Rapid-PSC-Timer: 10ms                            Slow-PSC-Timer : 5sec
===============================================================================

*A:mlstp-dutA# show router mpls mpls-tp status

===============================================================================
MPLS-TP Status
===============================================================================
Admin Status  : Up
Global ID     : 42                               Node ID       : 0.0.3.233
Tunnel Id Min : 1                                Tunnel Id Max : 4096
===============================================================================

*A:mlstp-dutA# show router mpls mpls-tp transit-path
  - transit-path [<path-name>] [detail]

 <path-name>          : [32 chars max]
 <detail>             : keyword - Display detailed information.
```

```
A:mplstp-dutC# show router mpls mpls-tp transit-path
  - transit-path [<path-name>] [detail]

 <path-name>          : [32 chars max]
 <detail>             : keyword - Display detailed information.
```

```
A:mplstp-dutC# show router mpls mpls-tp transit-path
<path-name>
 "tp-32"   "tp-33"   "tp-34"   "tp-35"   "tp-36"   "tp-37"   "tp-38"   "tp-39"
 "tp-40"   "tp-41"
detail

A:mplstp-dutC# show router mpls mpls-tp transit-path "tp-32"

===============================================================================
MPLS-TP Transit tp-32 Path Information
===============================================================================
Path Name     : tp-32
Admin State   : Up                               Oper State    : Up

------------------------------------------------------------------
Path        NextHop           InLabel   OutLabel  Out I/F
------------------------------------------------------------------
FP                            2080      2081      CtoB_1
RP                            2081      2080      CtoA_1
===============================================================================

A:mplstp-dutC# show router mpls mpls-tp transit-path "tp-32" detail

===============================================================================
MPLS-TP Transit tp-32 Path Information (Detail)
===============================================================================
Path Name     : tp-32
Admin State   : Up                               Oper State    : Up
-------------------------------------------------------------------------------
Path ID configuration
Src Global ID : 42                               Dst Global ID : 42
Src Node ID   : 0.0.3.234                        Dst Node ID   : 0.0.3.233
LSP Number    : 2                                Dst Tunnel Num: 32

Forward Path configuration
In Label      : 2080                             Out Label     : 2081
Out Interface : CtoB_1                           Next Hop Addr : n/a

Reverse Path configuration
In Label      : 2081                             Out Label     : 2080
Out Interface : CtoA_1                           Next Hop Addr : n/a
===============================================================================
A:mplstp-dutC#
```

2.3.8.6 MPLS-TP Interfaces

The following output is an example of mpls-tp specific information.

```
*A:mlstp-dutA# show router interface "AtoB_1"

===============================================================================
Interface Table (Router: Base)
===============================================================================
Interface-Name                   Adm         Opr(v4/v6)  Mode    Port/SapId
   IP-Address                                                    PfxState
```

```
-------------------------------------------------------------------------------
AtoB_1                          Down        Down/--     Network 1/2/3:1
   Unnumbered If[system]                                        n/a
-------------------------------------------------------------------------------
Interfaces : 1
```

2.3.9 MPLS-TP Debug Commands

The following command provides the debug command for an MPLS-TP tunnel:

tools>dump>router>mpls>tp-tunnel <lsp-name> [clear]

The following is a sample output:

```
A:mlstp-dutA# tools dump router mpls tp-tunnel
  - tp-tunnel <lsp-name> [clear]
  - tp-tunnel id <tunnel-id> [clear]

 <lsp-name>           : [32 chars max]
 <tunnel-id>          : [1..61440]
 <clear>              : keyword - clear stats after reading


*A:mlstp-dutA# tools dump router mpls tp-tunnel "lsp-
"lsp-32"  "lsp-33"  "lsp-34"  "lsp-35"  "lsp-36"  "lsp-37"  "lsp-38"  "lsp-39"
"lsp-40"  "lsp-41"
*A:mlstp-dutA# tools dump router mpls tp-tunnel "lsp-32"

 Idx: 1-32 (Up/Up): pgId 4, paths 2, operChg 1, Active: Protect
  TunnelId: 42::0.0.3.233::32-42::0.0.3.234::32
  PgState: Dn, Cnt/Tm: Dn 1/000 04:00:48.160 Up:3/000 00:01:25.840
  MplsMsg: tpDn 0/000 00:00:00.000, tunDn 0/000 00:00:00.000
           wpDn 0/000 00:00:00.000, ppDn 0/000 00:00:00.000
           wpDel 0/000 00:00:00.000, ppDel 0/000 00:00:00.000
           tunUp 1/000 00:00:02.070
  Paths:
   Work (Up/Dn): Lsp 1, Lbl 32/32, If 2/128 (1/2/3 : 0.0.0.0)
    Tmpl: ptc: , oam: privatebed-oam-template (bfd: privatebed-bfd-template(np)-
10 ms)
    Bfd: Mode CC state Dn/Up handle 160005/0
    Bfd-CC (Cnt/Tm): Dn 1/000 04:00:48.160 Up:1/000 00:01:23.970
    State:  Admin Up (1::1::1)  port Up , if Dn ,  operChg 2
    DnReasons: ccFault ifDn

   Protect (Up/Up): Lsp 2, Lbl 2080/2080, If 3/127 (5/1/1 : 0.0.0.0)
    Tmpl: ptc: privatebed-protection-template, oam: privatebed-oam-template (bfd:
privatebed-bfd-template(np)-10 ms)
    Bfd: Mode CC state Up/Up handle 160006/0
    Bfd-CC (Cnt/Tm): Dn 0/000 00:00:00.000 Up:1/000 00:01:25.410
    State:  Admin Up (1::1::1)  port Up , if Up ,  operChg 1

  Aps: Rx - 5, raw 3616, nok 0(), txRaw - 3636, revert Y
   Pdu: Rx - 0x1a-21::0101 (SF), Tx - 0x1a-21::0101 (SF)
   State: PF:W:L LastEvt pdu (L-SFw/R-SFw)
   Tmrs: slow
   Defects: None  Now: 000 05:02:19.130
  Seq  Event   state     TxPdu       RxPdu       Dir    Act          Time
  ===  ======  ========  ==========  ==========  =====  ====  ================
  000   start    UA:P:L    SF (0,0)    NR (0,0)  Tx-->  Work  000 00:00:02.080
  001     pdu    UA:P:L    SF (0,0)    SF (0,0)  Rx<--  Work  000 00:01:24.860
  002     pdu    UA:P:L    SF (0,0)    NR (0,0)  Rx<--  Work  000 00:01:26.860
  003     pUp        NR    NR (0,0)    NR (0,0)  Tx-->  Work  000 00:01:27.440
  004     pdu        NR    NR (0,0)    NR (0,0)  Rx<--  Work  000 00:01:28.760
  005     wDn    PF:W:L    SF (1,1)    NR (0,0)  Tx-->  Prot  000 04:00:48.160
  006     pdu    PF:W:L    SF (1,1)    NR (0,1)  Rx<--  Prot  000 04:00:48.160
  007     pdu    PF:W:L    SF (1,1)    SF (1,1)  Rx<--  Prot  000 04:00:51.080
```

The following command shows the free MPLS tunnel IDs available between two values, *start-range* and *end-range*.

tools>dump>router>mpls>free-tunnel-id <*start-range*> <*end-range*>

The following command provides a debug tool to view control-channel-status signaling packets.

```
*A:bksim1611# /debug service id 700 sdp 200:700 event-type ?{config-change |
oper-status-change | neighbor-discovery | control-channel-status}

*A:bksim1611# /debug service id 700 sdp 200:700 event-type control-channel-status

*A:bksim1611#
1 2012/08/31 09:56:12.09 EST MINOR: DEBUG #2001 Base PW STATUS SIG PKT (RX):
"PW STATUS SIG PKT (RX)::
Sdp Bind 200:700 Instance 3
    Version          : 0x0
    PW OAM Msg Type  : 0x27
    Refresh Time     : 0xa
    Total TLV Length : 0x8
    Flags            : 0x0
    TLV Type         : 0x96a
    TLV Len          : 0x4
    PW Status Bits   : 0x0
"

2 2012/08/31 09:56:22.09 EST MINOR: DEBUG #2001 Base PW STATUS SIG PKT (RX):
"PW STATUS SIG PKT (RX)::
Sdp Bind 200:700 Instance 3
    Version          : 0x0
    PW OAM Msg Type  : 0x27
    Refresh Time     : 0xa
    Total TLV Length : 0x8
    Flags            : 0x0
    TLV Type         : 0x96a
    TLV Len          : 0x4
    PW Status Bits   : 0x0
"

3 2012/08/31 09:56:29.44 EST MINOR: DEBUG #2001 Base PW STATUS SIG PKT (TX):
"PW STATUS SIG PKT (TX)::
Sdp Bind 200:700 Instance 3
```

```
Version          : 0x0
PW OAM Msg Type  : 0x27
Refresh Time     : 0x1e
Total TLV Length : 0x8
Flags            : 0x0
TLV Type         : 0x96a
TLV Len          : 0x4
PW Status Bits   : 0x0
```

2.4 Traffic Engineering

Without traffic engineering (TE), routers route traffic according to the SPF algorithm, disregarding congestion or packet types.

With TE, network traffic is routed efficiently to maximize throughput and minimize delay. TE facilitates traffic flows to be mapped to the destination through a different (less congested) path other than the one selected by the SPF algorithm.

MPLS directs a flow of IP packets along a label switched path (LSP). LSPs are simplex, meaning that the traffic flows in one direction (unidirectional) from an ingress router to an egress router. Two LSPs are required for duplex traffic. Each LSP carries traffic in a specific direction, forwarding packets from one router to the next across the MPLS domain.

When an ingress router receives a packet, it adds an MPLS header to the packet and forwards it to the next hop in the LSP. The labeled packet is forwarded along the LSP path until it reaches the destination point. The MPLS header is removed and the packet is forwarded based on Layer 3 information such as the IP destination address. The physical path of the LSP is not constrained to the shortest path that the IGP would choose to reach the destination IP address.

2.4.1 TE Metric (IS-IS and OSPF)

When the use of the TE metric is selected for an LSP, the shortest path computation after the TE constraints are applied will select an LSP path based on the TE metric instead of the IGP metric. The user configures the TE metric under the MPLS interface. Both the TE and IGP metrics are advertised by OSPF and IS-IS for each link in the network. The TE metric is part of the TE extensions of both IGP protocols.

A typical application of the TE metric is to allow CSPF to represent a dual TE topology for the purpose of computing LSP paths.

An LSP dedicated for real-time and delay sensitive user and control traffic has its path computed by CSPF using the TE metric. The user configures the TE metric to represent the delay figure, or a combined delay/jitter figure, of the link. In this case, the shortest path satisfying the constraints of the LSP path will effectively represent the shortest delay path.

An LSP dedicated for non-delay sensitive user and control traffic has its path computed by CSPF using the IGP metric. The IGP metric could represent the link bandwidth or some other figure as required.

When the use of the TE metric is enabled for an LSP, CSPF will first prune all links in the network topology that do not meet the constraints specified for the LSP path. These constraints include bandwidth, admin-groups, and hop limit. CSPF will then run an SPF on the remaining links. The shortest path among the all SPF paths is selected based on the TE metric instead of the IGP metric which is used by default. The TE metric is only used in CSPF computations for MPLS paths and not in the regular SPF computation for IP reachability.

2.4.2 Admin Group Support on Facility Bypass Backup LSP

This feature provides for the inclusion of the LSP primary path admin-group constraints in the computation of a Fast ReRoute (FRR) facility bypass backup LSP to protect the primary LSP path by all nodes in the LSP path.

This feature is supported with the following LSP types and in both intra-area and inter-area TE where applicable:

- Primary path of a RSVP P2P LSP.
- S2L path of an RSVP P2MP LSP instance
- LSP template for an S2L path of an RSVP P2MP LSP instance.
- LSP template for auto-created RSVP P2P LSP in intra-area TE.

2.4.2.1 Procedures at Head-End Node

The user enables the signaling of the primary LSP path admin-group constraints in the FRR object at the ingress LER with the following CLI command:

config>router>mpls>lsp>fast-reroute>propagate-admin-group

When this command is enabled at the ingress LER, the admin-group constraints configured in the context of the P2P LSP primary path, or the ones configured in the context of the LSP and inherited by the primary path, are copied into the FAST_REROUTE object. The admin-group constraints are copied into the *include-any* or *exclude-any* fields.

The ingress LER propagates these constraints to the downstream nodes during the signaling of the LSP to allow them to include the admin-group constraints in the selection of the FRR backup LSP for protecting the LSP primary path.

The ingress LER will insert the FAST_REROUTE object by default in a primary LSP path message. If the user disables the object using the following command, the admin-group constraints will not be propagated: **config>router>mpls>no frr-object**.

The same admin-group constraints can be copied into the Session Attribute object. They are intended for the use of an LSR, typically an ABR, to expand the ERO of an inter-area LSP path. They are also used by any LSR node in the path of a CSPF or non-CSPF LSP to check the admin-group constraints against the ERO regardless if the hop is strict or loose. These are governed strictly by the command:

config>router>mpls>lsp>propagate-admin-group

In other words, the user may decide to copy the primary path admin-group constraints into the FAST_REROUTE object only, or into the Session Attribute object only, or into both.

The PLR rules for processing the admin-group constraints can make use of either of the two object admin-group constraints.

2.4.2.2 Procedures at PLR Node

The user enables the use of the admin-group constraints in the association of a manual or dynamic bypass LSP with the primary LSP path at a Point-of-Local Repair (PLR) node using the following global command:

config>router>mpls>admin-group-frr

When this command is enabled, each PLR node reads the admin-group constraints in the FAST_REROUTE object in the Path message of the LSP primary path. If the FAST_REROUTE object is not included in the Path message, then the PLR will read the admin-group constraints from the Session Attribute object in the Path message.

If the PLR is also the ingress LER for the LSP primary path, then it just uses the admin-group constraint from the LSP and/or path level configurations.

Whether the PLR node is also the ingress LER or just an LSR for the protected LSP primary path, the outcome of the ingress LER configuration dictates the behavior of the PLR node and is summarized in Table 10.

Table 10 **Bypass LSP Admin-Group Constraint Behavior**

	Ingress LER Configuration	Session Attribute	FRR Object	Bypass LSP at PLR (LER/LSF) follows admin-group constraints
1	frr-object lsp>no propagate-admin group lsp>frr>propagate-admin-group	Admin color constraints not sent	Admin color constraints sent	Yes
2	frr-object lsp>propagate-admin-group lsp>frr>propagate-admin group	Admin color constraints sent	Admin color constraints sent	Yes
3	frr-object lsp>propagate-admin group lsp>frr>no propagate-admin-group	Admin color constraints sent	Admin color constraints not sent	No
4	No frr-object lsp>propagate-admin group lsp>frr>propagate-admin-group	Admin color constraints sent	Not present	Yes
5	No frr-object lsp>no propagate-admin group lsp>frr>propagate-admin-group	Admin color constraints not sent	Not present	No
6	No frr-object lsp>propagate-admin group lsp>frr>no propagate-admin-group	Admin color constraints sent	Not present	Yes

The PLR node then uses the admin-group constraints along with other constraints, such as hop-limit and SRLG, to select a manual or dynamic bypass among those that are already in use.

If none of the manual or dynamic bypass LSP satisfies the admin-group constraints, and/or the other constraints, the PLR node will request CSPF for a path that merges the closest to the protected link or node and that includes or excludes the specified admin-group IDs.

If the user changes the configuration of the above command, it will not have any effect on existing bypass associations. The change will only apply to new attempts to find a valid bypass.

2.4.3 Manual and Timer Resignal of RSVP-TE Bypass LSP

The **config**>**router**>**mpls**>**bypass-resignal-timer** command triggers the periodic global re-optimization of all dynamic bypass LSP paths associated with RSVP P2P LSP. The operation is performed at each expiry of the user-configurable bypass LSP resignal timer.

When this command is enabled, MPLS requests to CSPF for the best path for each dynamic bypass LSP originated on this node. The constraints, hop limit, SRLG and admin-group constraints, of the first associated LSP primary path that originally triggered the signaling of the bypass LSP must be satisfied. To do this, MPLS saves the original Path State Block (PSB) of that LSP primary path, even if the latter is torn down.

If CSPF returns no path or returns a new path with a cost that is lower than the current path, MPLS does not signal the new bypass path. If CSPF returns a new path with a cost that is lower than the current one, MPLS signals it. Also, if the new bypass path is SRLG strict disjoint with the primary path of the original PSB while the current path is SRLG loose disjoint, the manual bypass path is resignaled regardless of cost comparison.

After the new path is successfully signaled, MPLS evaluates each PSB of each PLR (that is, each unique avoid-node or avoid-link constraint) associated with the current bypass LSP path to check if the corresponding LSP primary path constraints are still satisfied by the new bypass LSP path. If so, the PSB association is moved to the new bypass LSP.

Each PSB for which the constraints are not satisfied remains associated with the PLR on the current bypass LSP and is checked at the next background PSB re-evaluation, or at the next timer or manual bypass re-optimization. Additionally, if SRLG FRR loose disjointness is configured using the **configure router mpls srlg-frr** command and the current bypass LSP is SRLG disjoint with a primary path while the new bypass LSP is not SRLG disjoint, the PSB association is not moved.

If a specific PLR associated with a bypass LSP is active, the corresponding PSBs remain associated with the current PLR until the Global Revertive Make-Before-Break (MBB) tears down all corresponding primary paths, which also causes the current PLR to be removed.

Note: While it is in the preceding state, the older PLR does not get any new PSB association until the PLR is removed. When the last PLR is removed, the older bypass LSP is torn down.

Additionally, PSBs that have not been moved by the dynamic or manual re-optimization of a bypass LSP, as a result of the PSB constraints not being met by the new signaled bypass LSP path, are re-evaluated by the FRR background task, which handles cases where the PSB has requested node protection but its current PLR is a link-protect.

This feature is not supported with inter-area dynamic bypass LSP and bypass LSP protecting S2L paths of a P2MP LSP.

The **tools**>**perform**>**router**>**mpls**>**resignal-bypass** command performs a manual re-optimization of a specific dynamic or manual bypass LSP, or of all dynamic bypass LSPs.

The name of a manual bypass LSP is configured by the user. The name of a dynamic bypass LSP is displayed in the output of **show>router>mpls>bypass-tunnel dynamic detail**.

The **delay** option triggers the global re-optimization of all dynamic bypass LSPs at the expiry of the specified delay. Effectively, this option forces the global bypass resignal timer to expire after an amount of time equal to the value of the **delay** parameter. This option has no effect on a manual bypass LSP.

However, when the bypass LSP name is specified, the named dynamic or manual bypass LSP is signaled and the associations moved only if the new bypass LSP path has a lower cost than the current one. This behavior is different from that of the similar command for the primary or secondary active path of an LSP, which signals and switches to the new path regardless of the cost comparison. This handling is required because a bypass LSP can have a large number of PSB associations and the associated processing churn is much higher.

In the specific case where the name corresponds to a manual bypass LSP, the LSP is torn down and resignaled using the new path provided by CSPF if no PSB associations exist. If one or more PSB associations exist but no PLR is active, the command fails and the user is prompted to explicitly enter the **force** option. In this case, the manual bypass LSP is torn down and resignaled, temporarily leaving the associated LSP primary paths unprotected. If one or more PLRs associated with the manual bypass LSP is active, the command fails.

Finally, and as with the timer based resignal, the PSB associations are checked for the SRLG and admin group constraints using the updated information provided by CSPF for the current path and new path of the bypass LSP. More details are provided in sections RSVP-TE Bypass LSP Path SRLG Information Update in Manual and Timer Resignal MBB and RSVP-TE Bypass LSP Path Administrative Group Information Update in Manual and Timer Resignal MBB.

2.4.3.1 RSVP-TE Bypass LSP Path SRLG Information Update in Manual and Timer Resignal MBB

This feature enhances procedures of the timer and manual resignal (both **delay** and **lsp** options) of the RSVP-TE bypass LSP path by updating the SRLG information of the links of the current path and checking for SRLG disjointness constraint. The following sequence describes the timer and manual resignal enhancements.

1. CSPF updates the SRLG membership of the current bypass LSP path and checks if the path violates the SRLG constraint of the first primary path that was associated with a PLR of this bypass LSP. This is referred to as the initial Path State Block (initial PSB).
2. CSPF attempts a new path computation for the bypass LSP using the initial PSB constraints.
3. MPLS uses the information returned by CSPF and determines if the new bypass path is more optimal.
    a. If SRLG FRR strict disjointness is configured (**configure**>**router**>**mpls**>**srlg-frr strict**) and CSPF indicates the updated SRLG information of current path violated the SRLG constraint of the PLR of the initial PSB, the new path is more optimal.
    b. Otherwise, MPLS performs additional checks using the PLR of the initial PSB to determine if the new path is more optimal. Table 11 summarizes the possible cases of bypass path optimality determination.

Table 11 **Determination of Bypass LSP Path Optimality**

PLR SRLG Constraint Check [a]		SRLG FRR Configuration (Strict/Loose)	Path Cumulative Cost Comparison [a]	Path Cumulative SRLG Weight Comparison [a]	More Optimal Path
Current Path	**New Path**				
Disjoint	Disjoint	—	New Cost < Current Cost	—	New

Table 11 **Determination of Bypass LSP Path Optimality (Continued)**

PLR SRLG Constraint Check [a]		SRLG FRR Configuration (Strict/Loose)	Path Cumulative Cost Comparison [a]	Path Cumulative SRLG Weight Comparison [a]	More Optimal Path
Current Path	New Path				
Disjoint	Disjoint	—	New Cost ≥ Current Cost	—	Current
Disjoint	Not Disjoint	—	—	—	Current
Not Disjoint	Not Disjoint	—	—	—	New
Not Disjoint	Not Disjoint	Strict	—	—	Current
Not Disjoint	Not Disjoint	Loose	New Cost < Current Cost	—	New
Not Disjoint	Not Disjoint	Loose	New Cost > Current Cost	—	Current
Not Disjoint	Not Disjoint	Loose	New Cost = Current Cost	New SRLG Weight < Current SRLG Weight	New
Not Disjoint	Not Disjoint	Loose	New Cost = Current Cost	New SRLG Weight ≥ Current SRLG Weight	Current

Note:

a. This check of the current path makes use of the updated SRLG and cost information provided by CSPF.

4. If the path returned by CSPF is found to be a more optimal bypass path with respect to the PLR of the initial PSB, the following sequence of actions is taken:
   i. MPLS signals and programs the new path.
   ii. MPLS moves to the new bypass path the PSB associations of all PLRs which evaluation against Table 11 results in the new bypass path being more optimal.

iii. Among the remaining PLRs, if the updated SRLG information of the current bypass path changed and SRLG FRR loose disjointness is configured (**configure**>**router**>**mpls**>**srlg-frr**), MPLS keeps this PLR PSB association with the current bypass path.

iv. Among the remaining PLRs, if the updated SRLG information of the current bypass path changed and SRLG strict disjointness is configured (**configure**>**router**>**mpls**>**srlg-frr strict**), MPLS evaluates the SRLG constraint of each PLR and performs the following actions.

i. MPLS keeps with the current bypass path the PSB associations of all PLRs where the SRLG constraint is not violated by the updated SRLG information of the current bypass path.

These PSBs are re-evaluated at the next timer or manual resignal MBB following the same procedure, as described in RSVP-TE Bypass LSP Path SRLG Information Update in Manual and Timer Resignal MBB.

ii. MPLS detaches from current bypass path the PSB associations of all PLRs where the SRLG constraint is violated by the updated SRLG information of the current bypass path.

These orphaned PSBs are re-evaluated by the FRR background task, which checks unprotected PSBs on a regular basis and following the same procedure, as described in RSVP-TE Bypass LSP Path SRLG Information Update in Manual and Timer Resignal MBB.

5. If the path returned by CSPF is found to be less optimal then the current bypass path or if CSPF did not return a new path, the following actions are performed.

i. If the updated SRLG information of the current bypass path did not change, MPLS keeps the current bypass path and the PSB associations of all PLRs.

ii. If the updated SRLG information of the current bypass path changed and SRLG FRR loose disjointness is configured (**configure**>**router**>**mpls**>**srlg-frr**), MPLS keeps the current bypass path and the PSB associations of all PLRs.

iii. If the updated SRLG information of the current bypass path changed and SRLG strict disjointness is configured (**configure**>**router**>**mpls**>**srlg-frr strict**), MPLS evaluates the SRLG constraint of each PLR and performs the following actions.

i. MPLS keeps with the current bypass path the PSB associations of all PLRs where the SRLG constraint is not violated by the updated SRLG information of the current bypass path.

These PSBs are re-evaluated at the next timer or manual resignal MBB following the same procedure, as described in RSVP-TE Bypass LSP Path SRLG Information Update in Manual and Timer Resignal MBB.

ii. MPLS detaches from current bypass path the PSB associations of all PLRs where the SRLG constraint is violated by the updated SRLG information of the current bypass path.

These orphaned PSBs are re-evaluated by the FRR background task, which checks unprotected PSBs on a regular basis and following the same procedure, as described in RSVP-TE Bypass LSP Path SRLG Information Update in Manual and Timer Resignal MBB.

2.4.3.2 RSVP-TE Bypass LSP Path Administrative Group Information Update in Manual and Timer Resignal MBB

This feature enhances procedures of the timer and manual resignal (both **delay** and **lsp** options) of a RSVP-TE bypass LSP path by updating the administrative group information of the current path links and checking for administrative group constraints. The following sequence describes the timer and manual resignal enhancements.

1. CSPF updates the administrative group membership of the current bypass LSP path and checks if the path violates the administrative group constraints of the first primary path which was associated with this bypass LSP. This is referred to as the initial PSB.
2. CSPF attempts a new path computation for the bypass LSP using the PLR constraints of the initial PSB.
3. MPLS uses the information returned by CSPF and determines if the new bypass path is more optimal.
    a. If CSPF indicated the updated administrative group information of current path violated the administrative group constraint of the initial PSB, then the new path is more optimal.
    b. Otherwise, the new path is more optimal only if its metric is lower than the updated metric of the current bypass path.
4. If the path returned by CSPF is found to be a more optimal bypass path then the following sequence of actions is performed.
    i. MPLS signals and programs the new path.
    ii. Since the administrative group constraint is not part of the PLR definition, MPLS evaluates the PSBs of all PLRs associated with the current bypass, and takes the following actions.
        i. MPLS moves to the new bypass path the PSB associations in which the administrative group constraints are not violated by the new bypass path.
        ii. Among the remaining PSBs, MPLS keeps with the current bypass path the PSB associations in which the administrative group constraints are not violated by the updated administrative group information of the current bypass path.

These PSBs are re-evaluated at the next timer or manual resignal MBB following the same procedure, as described in RSVP-TE Bypass LSP Path Administrative Group Information Update in Manual and Timer Resignal MBB.

iii. Among the remaining PSBs, MPLS detaches from current bypass path the PSB associations in which the administrative group constraints are violated by the updated administrative group information of the current bypass path.

These orphaned PSBs are re-evaluated by the FRR background task, which checks unprotected PSBs on a regular basis and following the same procedure, as described in RSVP-TE Bypass LSP Path Administrative Group Information Update in Manual and Timer Resignal MBB.

5. If the path returned by CSPF is found to be less optimal than the current bypass path or if CSPF did not return a new path, the following actions are performed.

   i. If the updated administrative group information of the current bypass path did not change, MPLS keeps the current bypass path and all PSB associations.

   ii. If the updated administrative group information of the current bypass path has changed, MPLS evaluates the PSBs of all PLRs associated with the current bypass, and performs the following actions.

      i. MPLS keeps with the current bypass path the PSB associations in which the administrative group constraints are not violated by the updated administrative group information of the current bypass path.

      ii. MPLS detaches from current bypass path the PSB associations in which the administrative group constraints are violated by the updated administrative group information of the current bypass path.

      These orphaned PSBs are re-evaluated by the FRR background task, which checks unprotected PSBs on a regular basis and following the same procedure, as described in RSVP-TE Bypass LSP Path Administrative Group Information Update in Manual and Timer Resignal MBB.

2.4.4 RSVP-TE LSP Active Path Administrative Group Information Update in Timer Resignal MBB

This feature enhances the procedures of the timer resignal and of the **delay** option of the manual resignal of the active path of a RSVP-TE LSP. The feature updates the administrative group information of the links of the current path and checks for administrative group constraint. MPLS performs the following sequence of actions.

1. CSPF checks the validity and updates the administrative group membership of the current active path. The validity of the path means that each TE link used by the path is still in the TE-DB, which ensures the continuous path form ingress to egress.
2. CSPF attempts a new path computation for the active path.
3. If CSPF returns a new path, MPLS performs the following actions.
   a. If CSPF finds the current path is invalid, MPLS signals and programs the new path.
   b. If the updated administrative group membership of the current path violates the path administrative group constraint, MPLS signals and programs the new path.
   c. If the updated administrative group membership of current path does not violate the path administrative group constraint, MPLS signals the new path only if its cumulative metric is different from the updated cumulative metric of the current path.
4. If CSPF returns no path, MPLS keeps the current path regardless of whether the updated administrative group membership of the current path violates the path administrative group constraint.

   This behavior of SR OS prevents unnecessary blackholing of traffic as a result of potential TE database churn, in which case a compliant path for the administrative group constraint is found at the next resignal timer expiry.

2.4.5 Diff-Serv Traffic Engineering

Diff-Serv traffic engineering (TE) provides the ability to manage bandwidth on a per-TE class basis as per RFC 4124. In the base traffic engineering, LER computes LSP paths based on available BW of links on the path. Diff-Serv TE adds ability to perform this on a per-TE class basis.

A TE class is a combination of Class Type and LSP priority. A Class Type is mapped to one or more system Forwarding Classes using a configuration profile. The operator sets different limits for admission control of LSPs in each TE class over each TE link. Eight TE classes are supported. Admission control of LSP paths bandwidth reservation is performed using the Maximum Allocation Bandwidth Constraint Model as per RFC 4125.

2.4.5.1 Mapping of Traffic to a Diff-Serv LSP

An LER allows the operator to map traffic to a Diff-Serv LSP using one of the following methods:

1. Explicit RSVP SDP configuration of a VLL, VPLS, or VPRN service
2. Class-based forwarding in an RSVP SDP. The operator can enable the checking by RSVP that a Forwarding Class (FC) mapping to an LSP under the SDP configuration is compatible with the Diff-Serv Class Type (CT) configuration for this LSP.
3. The **auto-bind-tunnel** RSVP-TE option in a VPRN service
4. Static routes with indirect next-hop being an RSVP LSP name

2.4.5.2 Admission Control of Classes

There are a couple of admission control decisions made when an LSP with a specified bandwidth is to be signaled. The first is in the head-end node. CSPF will only consider network links that have sufficient bandwidth. Link bandwidth information is provided by IGP TE advertisement by all nodes in that network.

Another decision made is local CAC and is performed when the RESV message for the LSP path is received in the reverse direction by a SR OS in that path. The bandwidth value selected by the egress LER is checked against link bandwidth, otherwise the reservation is rejected. If accepted, the new value for the remaining link bandwidth is advertised by IGP at the next advertisement event.

Both of these admission decisions are enhanced to be performed at the TE class level when Diff-Serv TE is enabled. In other words, CSPF in the head-end node will need to check the LSP bandwidth against the ‘unreserved bandwidth’ advertised for all links in the path of the LSP for that TE class which consists of a combination of a CT and a priority. Same for the admission control at SR OS receiving the Resv message.

2.4.5.2.1 Maximum Allocation Model

The admission control rules for this model are described in RFC 4125. Each CT shares a percentage of the Maximum Reservable Link Bandwidth through the user-configured BC for this CT. The Maximum Reservable Link Bandwidth is the link bandwidth multiplied by the RSVP interface subscription factor.

The sum of all BC values across all CTs will not exceed the Maximum Reservable Link Bandwidth. In other words, the following rule is enforced:

SUM (BCc) =< Max-Reservable-Bandwidth, $0 \leq c \leq 7$

An LSP of class-type CTc, setup priority p, holding priority h (h=<p), and bandwidth B is admitted into a link if the following condition is satisfied:

B ≤ Unreserved Bandwidth for TE-Class[i]

where TE-Class [i] maps to < CTc, p > in the definition of the TE classes on the node. The bandwidth reservation is effected at the holding priority; that is, in TE-class [j] = <CTc, h>. As such, the reserved bandwidth for CTc and the unreserved bandwidth for the TE classes using CTc are updated as follows:

Reserved(CTc) = Reserved(CTc) + B

Unreserved TE-Class [j] = BCc - SUM (Reserved(CTc,q)) for $0 \leq q \leq h$

Unreserved TE-Class [i] = BCc - SUM (Reserved(CTc,q)) for $0 \leq q \leq p$

The same is done to update the unreserved bandwidth for any other TE class making use of the same CTc. These new values are advertised to the rest of the network at the next IGP-TE flooding.

When Diff-Serv is disabled on the node, this model degenerates into a single default CT internally with eight preemption priorities and a non-configurable BC equal to the Maximum Reservable Link Bandwidth. This would behave exactly like CT0 with eight preemption priorities and BC= Maximum Reservable Link Bandwidth if Diff-Serv was enabled.

2.4.5.2.2 Russian Doll Model

The RDM model is defined using the following equations:

SUM (Reserved (CTc)) ≤ **BCb**,

where the SUM is across all values of **c** in the range **b** ≤ **c** ≤ **(MaxCT - 1)**, and **BCb** is the bandwidth constraint of **CTb**.

BC0= Max-Reservable-Bandwidth, so that:

SUM (Reserved(CTc)) ≤ **Max-Reservable-Bandwidth**,

where the **SUM** is across all values of **c** in the range **0** ≤ **c** ≤ **(MaxCT - 1)**

An LSP of class-type **CTc**, setup priority **p**, holding priority **h (h=<p)**, and bandwidth **B** is admitted into a link if the following condition is satisfied:

B $\leq$ **Unreserved Bandwidth for TE-Class[i]**,

where **TE-Class [i]** maps to < **CTc, p** > in the definition of the TE classes on the node. The bandwidth reservation is effected at the holding priority, that is, in **TE-class [j] = <CTc, h>**. As such, the reserved bandwidth for CTc and the unreserved bandwidth for the TE classes using CTc are updated as follows:

Reserved(CTc) = Reserved(CTc) + B

Unreserved TE-Class [j] = Unreserved (CTc, h) = Min [

BCc - SUM (Reserved (CTb, q) for 0≤q ≤ h, c ≤ b ≤ 7,

BC(c-1) – SUM (Reserved (CTb, q) for 0≤q ≤ h, (c-1) ≤ b ≤ 7,

…….

BC0 - SUM (Reserved (CTb, q) for 0≤q ≤ h, 0 ≤ b ≤ 7]

Unreserved TE-Class [i] = Unreserved (CTc, p) = Min [

BCc - SUM (Reserved (CTb, q) for 0≤q ≤ p, c ≤ b ≤ 7,

BC(c-1) – SUM (Reserved (CTb, q) for 0≤q ≤ p, (c-1) ≤ b ≤ 7,

…….

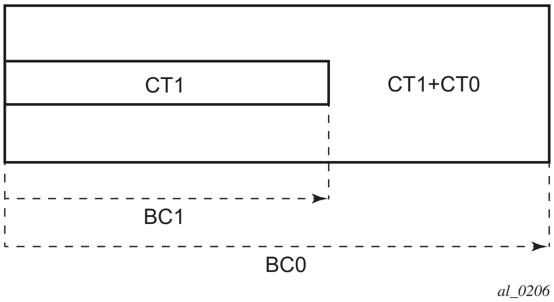
BC0 - SUM (Reserved (CTb, q) for 0≤q ≤ p, 0 ≤ b ≤7]

The same is done to update the unreserved bandwidth for any other TE class making use of the same CTc. These new values are advertised to the rest of the network at the next IGP-TE flooding.

Example CT Bandwidth Sharing with RDM

Below is a simple example with two CT values (CT0, CT1) and one priority 0 as shown in Figure 24.

Figure 24 **RDM with Two Class Types**



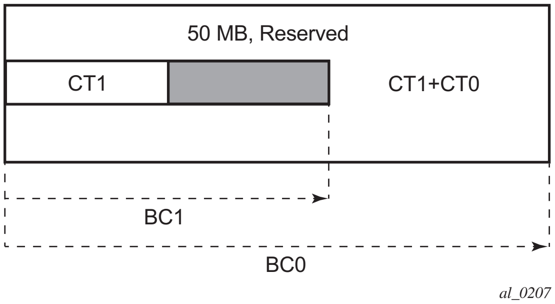
Suppose CT1 bandwidth, or the CT1 percentage of Maximum Reservable Bandwidth to be more accurate is 100 Mb/s and CT2 bandwidth is 100 Mb/s and link bandwidth is 200 Mb/s. BC constraints can be calculated as follows.

BC1 = CT1 Bandwidth = 100 Mb/s.

BC0 = {CT1 Bandwidth} + {CT0 Bandwidth} = 200 Mb/s.

Suppose an LSP comes with CT1, setup and holding priorities of 0 and a bandwidth of 50 Mb/s.

Figure 25 **First LSP Reservation**



According to the RDM admission control policy:

Reserved (CT1, 0) = 50 $\leq$ 100 Mb/s

Reserved (CT0, 0) + Reserved (CT1, 0) = 50 $\leq$ 200 Mb/s

This results in the following unreserved bandwidth calculation.

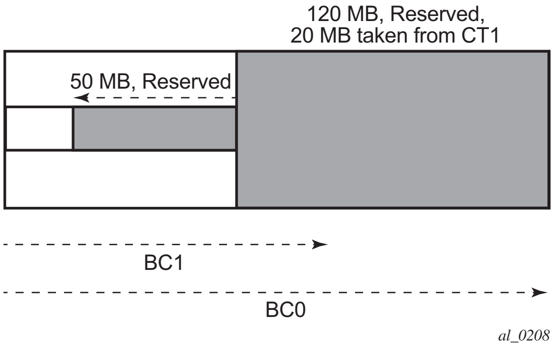
Unreserved (CT1, 0) = BC1 – Reserved (CT1, 0) = 100 – 50 = 50 Mb/s

Unreserved (CT0, 0) = BC0 – Reserved (CT0, 0) – Reserved (CT1, 0) = 200 – 0 – 50= 150 Mb/s.

The bandwidth reserved by a doll is not available to itself or any of the outer dolls.

Suppose now another LSP comes with CT0, setup and holding priorities of 0 and a bandwidth 120 Mb/s.

Figure 26 **Second LSP Reservation**



Reserved (CT0, 0) = 120 $\leq$ 150 Mb/s

Reserved (CT0, 0) + Reserved (CT1, 0) = 120 + 50 = 170 $\leq$ 200 Mb/s

Unreserved (CT0, 0) = 150 -120 = 30 Mb/s

If we simply checked BC1, the formula would yield the wrong results:

Unreserved (CT1, 0) = BC1 – Reserved (CT1, 0) = 100 -50 = 50 Mb/s

Because of the encroaching of CT0 into CT1, we would need to deduct the overlapping reservation. This would then yield:

Unreserved (CT1, 0) = BC0 – Reserved (CT0, 0) – Reserved (CT1, 0) = 200 – 120 - 50 = 30 Mb/s, which is the correct figure.

Extending the formula with both equations:

Unreserved (CT1, 0) = Min [BC1 – Reserved (CT1, 0), BC0 – Reserved (CT0, 0) – Reserved (CT1, 0)] = Min [100 – 50, 200 – 120 – 50] = 30 Mb/s

An outer doll can encroach into an inner doll, reducing the bandwidth available for inner dolls.

2.4.5.3 RSVP Control Plane Extensions

RSVP will use the Class Type object to carry LSP class-type information during path setup. Eight values are supported for class-types 0 through 7 as per RFC 4124. Class type 0 is the default class which is supported today on the router.

One or more forwarding classes will map to a Diff-Serv class type trough a system level configuration.

2.4.5.4 IGP Extensions

IGP extensions are defined in RFC 4124. Diff-Serv TE advertises link available bandwidth, referred to as unreserved bandwidth, by OSPF TE or IS-IS TE on a per TE class basis. A TE class is a combination of a class type and an LSP priority. In order to reduce the amount of per TE class flooding required in the network, the number of TE classes is set to eight. This means that eight class types can be supported with a single priority or four class types with two priorities, and so on. In that case, the operator configures the desired class type on the LSP such that RSVP-TE can signal it in the class-type object in the path message.

IGP will continue to advertise the existing Maximum Reservable Link Bandwidth TE parameter to mean the maximum bandwidth that can be booked on a given interface by all classes. The value advertised is adjusted with the link subscription factor.

2.4.5.5 Diff-Serv TE Configuration and Operation

2.4.5.5.1 RSVP Protocol Level

The following are the configuration steps at the RSVP protocol level:

1. The operator enables Diff-Serv TE by executing the **diffserv-te** command in the **config>router>rsvp** context. When this command is enabled, IS-IS and OSPF will start advertising available bandwidth for each TE class configured under the **diffserv-te** node. The operator can disable Diff-Serv TE globally by using the no form of the command.

2. The enabling or disabling of Diff-Serv on the system requires that the RSVP and MPLS protocol be shutdown. The operator must execute the **no shutdown** command in each context once all parameters under both protocols are defined. When saved in the configuration file, the **no shutdown** command is automatically inserted under both protocols to make sure they come up after a node reboot.
3. IGP will advertise the available bandwidth in each TE class in the unreserved bandwidth TE parameter for that class for each RSVP interface in the system.
4. In addition, IGP will continue to advertise the existing Maximum Reservable Link Bandwidth TE parameter so the maximum bandwidth that can be booked on a given interface by all classes. The value advertised is adjusted with the link subscription factor configured in the **config>router>rsvp>if>subscription** *percentage* context.
5. The operator can overbook (underbook) the maximum reservable bandwidth of a given CT by overbooking (underbooking) the interface maximum reservable bandwidth by configuring the appropriate value for the **subscription** *percentage* parameter.
6. The **diffserv-te** command will only have effect if the operator has already enabled TE at the IS-IS and/or OSPF routing protocol levels:

   config>router>isis>traffic-engineering

   and/or:

   config>router>ospf>traffic-engineering
7. The following Diff-Serv TE parameters are configured globally under the **diffserv-te** node. They apply to all RSVP interfaces on the system. Once configured, these parameters can only be changed after shutting down the MPLS and RSVP protocols:

   a. Definition of TE classes, TE Class = {Class Type (CT), LSP priority}. Eight TE classes can be supported. There is no default TE class once Diff-Serv is enabled. The operator must explicitly define each TE class. However, when Diff-Serv is disabled there is an internal use of the default CT (CT0) and eight preemption priorities as shown in Table 12.

Table 12 **Internal TE Class Definition when Diff-Serv TE is Disabled**

Class Type (CT internal)	LSP Priority
0	7
0	6
0	5
0	4

Table 12 **Internal TE Class Definition when Diff-Serv TE is Disabled**

Class Type (CT internal)	LSP Priority
0	3
0	2
0	1
0	0

b. A mapping of the system forwarding class to CT. The default settings are shown in Table 13.

Table 13 **Default Mapping of Forwarding Class to TE Class**

FC ID	FC Name	FC Designation	Class Type (CT)
7	Network Control	NC	7
6	High-1	H1	6
5	Expedited	EF	5
4	High-2	H2	4
3	Low-1	L1	3
2	Assured	AF	2
1	Low-2	L2	1
0	Best Effort	BE	0

c. Configuration of the percentage of RSVP interface bandwidth each CT shares, for example, the Bandwidth Constraint (BC), using the **class-type-bw** command. The absolute value of the CT share of the interface bandwidth is derived as the percentage of the bandwidth advertised by IGP in the maximum reservable link bandwidth TE parameter, for example, the link bandwidth multiplied by the RSVP interface **subscription** *percentage* parameter. Note that this configuration also exists at the RSVP interface level and the interface specific configured value overrides the global configured value. The BC value can be changed at any time. The operator can specify the BC for a CT which is not used in any of the TE class definition but that does not get used by any LSP originating or transiting this node.

d. Configuration of the Admission Control Policy to be used: only the Maximum Allocation Model (MAM) is supported. The MAM value represents the bandwidth constraint models for the admission control of an LSP reservation to a link.

2.4.5.5.2 RSVP Interface Level

The following are the configuration steps at the RSVP interface level.

1. The operator configures the percentage of RSVP interface bandwidth each CT shares, for example, the BC, using the **class-type-bw** command. The value entered at the interface level overrides the global value configured under the **diffserv-te** node.
2. The operator can overbook (underbook) the maximum reservable bandwidth of a given CT by overbooking (underbooking) the interface maximum reservable bandwidth via configuring the appropriate value for the **subscription** *percentage* parameter in the **config>router>rsvp>interface** context.
3. .Both the BC value and the subscription parameter can be changed at any time.

2.4.5.5.3 LSP and LSP Path Levels

The following are the configuration steps at the LSP and LSP path levels.

1. The operator configures the CT in which the LSP belongs by configuring the **class-type** *ct-number* command at the LSP level and/or the path level. The path level value overrides the LSP level value. By default, an LSP belongs to CT0.
2. Only one CT per LSP path is allowed per RFC 4124, *Protocol Extensions for Support of Diffserv-aware MPLS Traffic Engineering*. A multi-class LSP path is achieved through mapping multiple system Forwarding Classes to a CT.
3. The signaled CT of a dynamic bypass must always be CT0 regardless of the CT of the primary LSP path. The setup and hold priorities must be set to default values, for example, 7 and 0 respectively. This assumes that the operator configured a couple of TE classes, one which combines CT0 and a priority of 7 and the other which combines CT0 and a priority of 0. If not, the bypass LSP will not be signaled and will go into the down state.
4. The operator cannot configure the CT, setup priority, and holding priority of a manual bypass. They are always signaled with CT0 and the default setup and holding priorities.
5. The signaled CT, setup priority and holding priority of a detour LSP matches those of the primary LSP path it is associated with.

6. The operator can also configure the setup and holding priorities for each LSP path.
7. An LSP which does not have the CT explicitly configured will behave like a CT0 LSP when Diff-Serv is enabled.

If the operator configured a combination of a CT and a setup priority and/or a combination of a CT and a holding priority for an LSP path that are not supported by the user-defined TE classes, the LSP path is kept in a down state and error code is shown within the show command output for the LSP path.

2.4.6 Diff-Serv TE LSP Class Type Change under Failure

An option to configure a main Class Type (CT) and a backup CT for the primary path of a Diff-Serv TE LSP is provided. The main CT is used under normal operating conditions, for example, when the LSP is established the first time and when it gets re-optimized due to timer based or manual resignal. The backup CT is used when the LSP retries under failure.

The use of backup Class Type (CT) by an LSP is enabled by executing the **config>router>mpls>lsp>primary>backup-class-type** *ct-number* command at the LSP primary path level.

When this option is enabled, the LSP will use the CT configured using the following commands (whichever is inherited at the primary path level) as the main CT:

- config>router>mpls>lsp>class-type *ct-number*
- config>router>mpls>lsp>primary>class-type *ct-number*

The main CT is used at initial establishment and during a manual or a timer based resignal Make-Before-Break (MBB) of the LSP primary path. The backup CT is used temporarily to signal the LSP primary path when it fails and goes into retry.

Note that any valid values may be entered for the backup CT and main CT, but they cannot be the same. No check is performed to make sure that the backup CT is a lower CT in Diff-Serv Russian-Doll Model (RDM) admission control context.

The secondary paths of the same LSP are always signaled using the main CT as in existing implementation.

2.4.6.1 LSP Primary Path Retry Procedures

This feature behaves according to the following procedures.

- When an LSP primary path retries due a failure, for example, it fails after being in the up state, or undergoes any type of MBB, MPLS will retry a new path for the LSP using the main CT. If the first attempt failed, the head-end node performs subsequent retries using the backup CT. This procedure must be followed regardless if the currently used CT by this path is the main or backup CT. This applies to both CSPF and non-CSPF LSPs.
- The triggers for using the backup CT after the first retry attempt are:
  - A local interface failure or a control plane failure (hello timeout, and so on).
  - Receipt of a PathErr message with a notification of a FRR protection becoming active downstream and/or receipt of a Resv message with a ‘Local-Protection-In-Use’ flag set. This invokes the FRR Global Revertive MBB.
  - Receipt of a PathErr message with error code=25 (Notify) and sub-code=7 (Local link maintenance required) or a sub-code=8 (Local node maintenance required). This invokes the TE Graceful Shutdown MBB. Note that in this case, only a single attempt is performed by MBB as in current implementation; only the main CT is retried.
  - Receipt of a Resv refresh message with the ‘Preemption pending’ flag set or a PathErr message with error code=34 (Reroute) and a value=1 (Reroute request soft preemption). This invokes the soft preemption MBB.
  - Receipt of a ResvTear message.
  - A configuration change MBB.
- When an unmapped LSP primary path goes into retry, it uses the main CT until the number of retries reaches the value of the new main-ct-retry-limit parameter. If the path did not come up, it must start using the backup CT at that point in time. By default, this parameter is set to infinite value. The new main-ct-retry-limit parameter has no effect on an LSP primary path, which retries due to a failure event. This parameter is configured using the **main-ct-retry-limit** command in the **config>router>mpls>lsp** context. If the user entered a value of the **main-ct-retry-limit** parameter that is greater than the LSP retry-limit, the number of retries will still stop when the LSP primary path reaches the value of the LSP retry-limit. In other words, the meaning of the LSP retry-limit parameter is not changed and always represents the upper bound on the number of retries. The unmapped LSP primary path behavior applies to both CSPF and non-CSPF LSPs.
- An unmapped LSP primary path is a path that never received a Resv in response to the first path message sent. This can occur when performing a “shut/no-shut” on the LSP or LSP primary path or when the node reboots. An unmapped LSP primary path goes into retry if the retry timer expired or the head-end node received a PathErr message before the retry timer expired.
- When the **clear>router>mpls>lsp** command is executed, the retry behavior for this LSP is the same as in the case of an unmapped LSP.

- If the value of the parameter main-ct-retry-limit is changed, the new value will only be used at the next time the LSP path is put into a “no-shut” state.
- The following is the behavior when the user changes the main or backup CT:
  - If the user changes the LSP level CT, all paths of the LSP are torn down and resignaled in a break-before-make fashion. Specifically, the LSP primary path is torn down and resignaled even if it is currently using the backup CT.
  - If the user changes the main CT of the LSP primary path, the path is torn down and resignaled even if it is currently using the backup CT.
  - If the user changes the backup CT of an LSP primary path when the backup CT is in use, the path is torn down and is resignaled.
  - If the user changes the backup CT of an LSP primary path when the backup CT is not in use, no action is taken. If however, the path was in global Revertive, gshut, or soft preemption MBB, the MBB is restarted. This actually means the first attempt is with the main CT and subsequent ones, if any, with the new value of the backup CT.
  - Consider the following priority of the various MBB types from highest to lowest: Delayed Retry, Preemption, Global Revertive, Configuration Change, and TE Graceful Shutdown. If an MBB request occurs while a higher priority MBB is in progress, the latter MBB is restarted. This actually means the first attempt is with the main CT and subsequent ones, if any, with the new value of the backup CT.
- If the least-fill option is enabled at the LSP level, then CSPF must use least-fill equal cost path selection when the main or backup CT is used on the primary path.
- When the resignal timer expires, CSPF will try to find a path with the main CT. The head-end node must resignal the LSP even if the new path found by CSPF is identical to the existing one since the idea is to restore the main CT for the primary path. If a path with main CT is not found, the LSP remains on its current primary path using the backup CT. This means that the LSP primary path with the backup CT may no longer be the most optimal one. Furthermore, if the least-fill option was enabled at the LSP level, CSPF will not check if there is a more optimal path, with the backup CT, according to the least-fill criterion and, so, will not raise a trap to indicate the LSP path is eligible for least-fill re-optimization.
- When the user performs a manual resignal of the primary path, CSPF will try to find a path with the main CT. The head-end node must resignal the LSP as in current implementation.
- If a CPM switchover occurs while an the LSP primary path was in retry using the main or backup CT, for example, was still in operationally down state, the path retry is restarted with the main CT until it comes up. This is because the LSP path retry count is not synchronized between the active and standby CPMs until the path becomes up.

- When the user configured secondary standby and non-standby paths on the same LSP, the switchover behavior between primary and secondary is the same as in existing implementation.

This feature is not supported on a P2MP LSP.

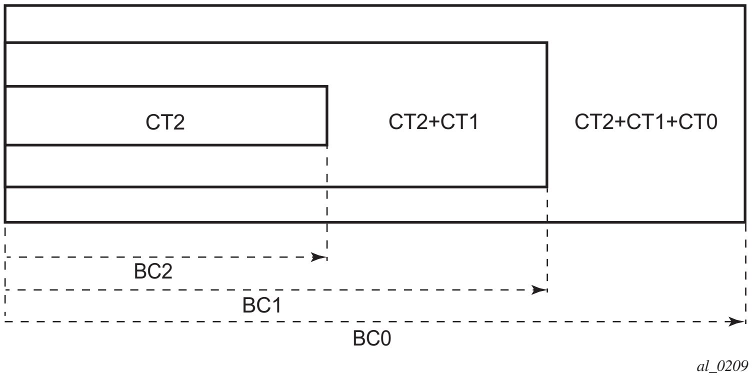
2.4.6.2 Bandwidth Sharing Across Class Types

In order to allow different levels of booking of network links under normal operating conditions and under failure conditions, it is necessary to allow sharing of bandwidth across class types.

This feature introduces the Russian-Doll Model (RDM) Diff-Serv TE admission control policy described in RFC 4127, *Russian Dolls Bandwidth Constraints Model for Diffserv-aware MPLS Traffic Engineering*. This mode is enabled using the following command: **config>router>rsvp>diffserv-te rdm**.

The Russian Doll Model (RDM) LSP admission control policy allows bandwidth sharing across Class Types (CTs). It provides a hierarchical model by which the reserved bandwidth of a CT is the sum of the reserved bandwidths of the numerically equal and higher CTs. Figure 27 shows an example.

Figure 27 **RDM Admission Control Policy Example**



CT2 has a bandwidth constraint BC2 which represents a percentage of the maximum reservable link bandwidth. Both CT2 and CT1 can share BC1 which is the sum of the percentage of the maximum reservable bandwidth values configured for CT2 and CT1 respectively. Finally, CT2, CT1, and CT0 together can share BC0 which is the sum of the percentage of the maximum reservable bandwidth values configured for CT2, CT1, and CT0 respectively. The maximum value for BC0 is of course the maximum reservable link bandwidth.

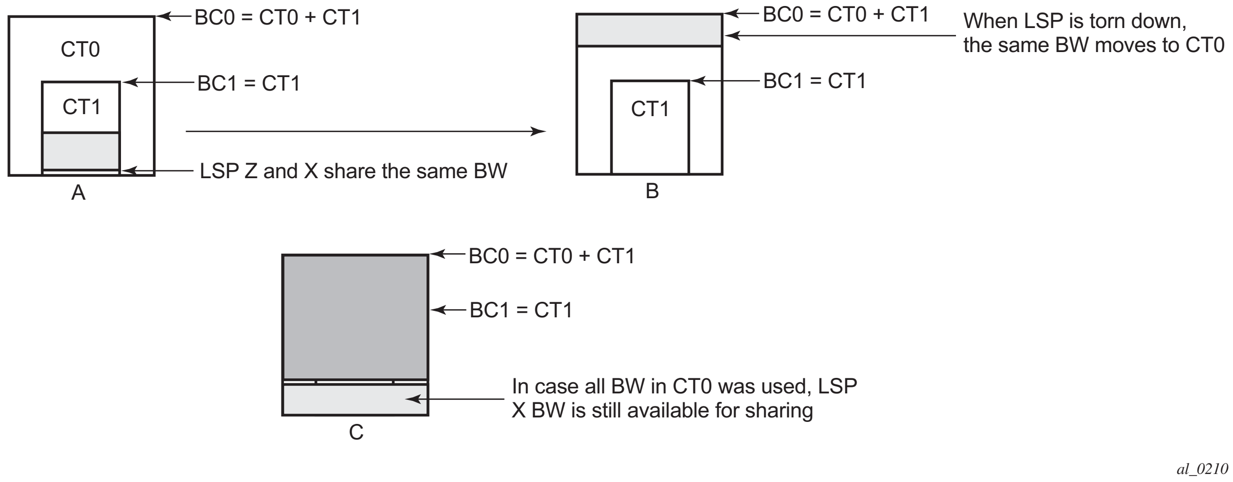
What this means in practice is that CT0 LSPs can use up to BC0 in the absence of LSPs in CT1 and CT2. When this occurs and a CT2 LSP with a reservation less than or equal to BC2 requests admission, it is only admitted by preempting one or more CT0 LSPs of lower holding priority than this LSP setup priority. Otherwise, the reservation request for the CT2 LSP is rejected.

It is required that multiple paths of the same LSP share common link bandwidth since they are signaled using the Shared Explicit (SE) style. Specifically, two instances of a primary path, one with the main CT and the other with the backup CT, must temporarily share bandwidth while MBB is in progress. Also, a primary path and one or many secondary paths of the same LSP must share bandwidth whether they are configured with the same or different CTs.

2.4.6.3 Downgrading the CT of Bandwidth Sharing LSP Paths

Consider a link configured with two class types CT0 and CT1 and making use of the RDM admission control model as shown in Figure 28.

Figure 28 **Sharing bandwidth when an LSP primary path is downgraded to backup CT**



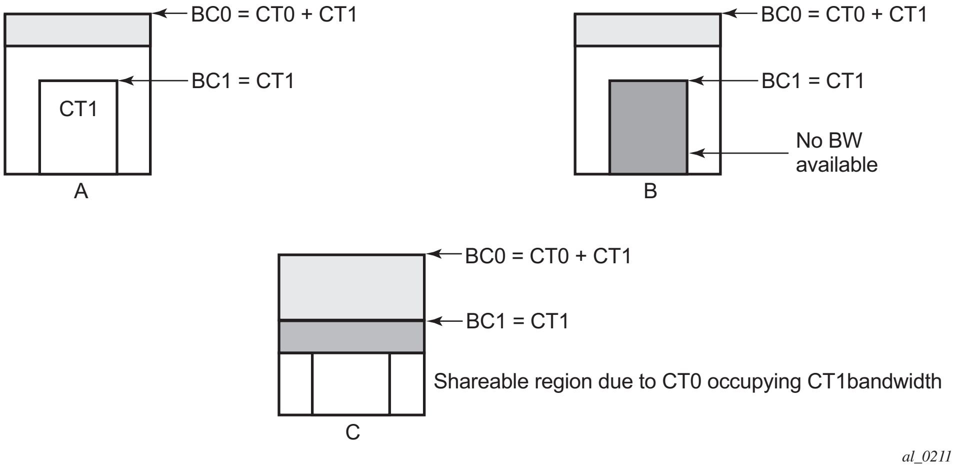
Consider an LSP path Z occupying bandwidth B at CT1. BC0 being the sum of all CTs below it, the bandwidth occupied in CT1 is guaranteed to be available in CT0. When new path X of the same LSP for CT0 is setup, it will use the same bandwidth B as used by path Z as shown in Figure 28 (a). When path Z is torn down the same bandwidth now occupies CT0 as shown in Figure 28 (b). Even if there were no new BW available in CT0 as can be seen in Figure 28 (c), path X can always share the bandwidth with path Z.

CSPF at the head-end node and CAC at the transit LSR node will share bandwidth of an existing path when its CT is downgraded in the new path of the same LSP.

2.4.6.4 Upgrading the CT of Bandwidth Sharing LSP Paths

When upgrading the CT the following issue can be apparent. Assume an LSP path X exists with CT0. An attempt is made to upgrade this path to a new path Z with CT1 using an MBB.

Figure 29 **Sharing Bandwidth When an LSP Primary Path is Upgraded to Main CT**



In Figure 29 (a), if the path X occupies the bandwidth as shown it can not share the bandwidth with the new path Z being setup. If a condition exists, as shown in Figure 29, (b) the path Z can never be setup on this particular link.

Consider Figure 29 (c). The CT0 has a region that overlaps with CT1 as CT0 has incursion into CT1. This overlap can be shared. However, in order to find whether such an incursion has occurred and how large the region is, it is required to know the reserved bandwidths in each class. Currently, IGP-TE advertises only the unreserved bandwidths. Hence, it is not possible to compute these overlap regions at the head end during CSPF. Moreover, the head end needs to then try and mimic each of the traversed links exactly which increases the complexity.

CSPF at the head-end node will only attempt to signal the LSP path with an upgraded CT if the advertised bandwidth for that CT can accommodate the bandwidth. In other words, it will assume that in the worst case this path will not share bandwidth with another path of the same LSP using a lower CT.

2.5 IPv6 Traffic Engineering

This feature extends the traffic engineering capability with the support of IPv6 TE links and nodes.

This feature enhances IS-IS, BGP-LS and the TE database with the additional IPv6 link TLVs and TE link TLVs and provides the following three modes of operation of the IPv4 and IPv6 traffic engineering in a network.

- Legacy Mode — This mode enables the existing traffic engineering behavior for IPv4 RSVP-TE and IPv4 SR-TE. Only the RSVP-TE attributes are advertised in the legacy TE TLVs that are used by both RSVP-TE and SR-TE LSP path computation in the TE domain routers. In addition, IPv6 SR-TE LSP path computation can now use these common attributes.
- Legacy Mode with Application Indication — This mode is intended for cases where link TE attributes are common to RSVP-TE and SR-TE applications and have the same value, but the user wants to indicate on a per-link basis which application is enabled.

  Routers in the TE domain use these attributes to compute path for IPv4 RSVP-TE LSP and IPv4/IPv6 SR-TE LSP.
- Application Specific Mode — This mode of operation is intended for future use cases where TE attributes may have different values in RSVP-TE and SR-TE applications or are specific to one application (for example, RSVP-TE ‘Unreserved Bandwidth' and `Max Reservable Bandwidth' attributes).

  SR OS does not support configuring TE attributes that are specific to the SR-TE application. As a result, enabling this mode advertises the common TE attributes once using a new Application Specific Link Attributes TLV. Routers in the TE domain use these attributes to compute paths for IPv4 RSVP-TE LSP and IPv4/IPv6 SR-TE LSP.

See IS-IS IPv4/IPv6 SR-TE and IPv4 RSVP-TE Feature Behavior for more details on the IPv4 and IPv6 Traffic Engineering modes of operation.

The feature also adds support of IPv6 destinations to the SR-TE LSP configuration. In addition, this feature also extends the MPLS path configuration with hop indices that include IPv6 addresses.

IPv6 SR-TE LSP is supported with the hop-to-label and the local CSPF path computation methods. It requires the enabling of the IPv6 traffic engineering feature in IS-IS.

2.5.1 Global Configuration

In order to enable IPv6 TE on the router, a new parameter referred to as IPv6 TE Router ID must have a valid IPv6 address. The following CLI command is used to configure the parameter:

configure>**router**>**ipv6-te-router-id interface** *interface-name*

The IPv6 TE Router ID is a mandatory parameter and allows the router to be uniquely identified by other routers in an IGP TE domain as being IPv6 TE capable. IS-IS advertises this information using the IPv6 TE Router ID TLV as explained in TE Attributes Supported in IGP and BGP-LS.

When the command is not configured, or the **no** form of the command is configured, the value of the IPv6 TE Router ID parameter reverts to the preferred primary global unicast address of the system interface. The user can also explicitly enter the name of the system interface to achieve the same outcome.

In addition, the user can specify a different interface and the preferred primary global unicast address of that interface is used instead. Only the system or a loopback interface is allowed since the TE router ID must use the address of a stable interface.

This address must be reachable from other routers in a TE domain and the associated interface must be added to IGP for reachability. Otherwise, IS-IS withdraws the advertisement of the IPv6 TE Router ID TLV.

When configuring a new interface name for the IPv6 TE Router ID, or when the same interface begins using a new preferred primary global unicast address, IS-IS immediately floods the new value.

If the referenced system is shut down or the referenced loopback interface is deleted or is shut down, or the last IPv6 address on the interface is removed, IS-IS withdraws the advertisement of the IPv6 TE Router ID TLV.

2.5.2 IS-IS Configuration

In order to enable the advertisement of additional link IPv6 and TE parameters, a new **traffic-engineering-options** CLI construct is used.

```
configure
     router
          ipv6-te-router-id interface interface-name
          no ipv6-te-router-id
          [no] isis [instance]
               traffic-engineering
```

```
no traffic-engineering
traffic-engineering-options
no traffic-engineering-options
     ipv6
     no ipv6
     application-link-attributes
     no application-link-attributes
          legacy
          no legacy
```

The existing **traffic-engineering** command continues its role as the main command for enabling TE in an IS-IS instance. This command enables the advertisement of the IPv4 and TE link parameters using the legacy TE encoding as per RFC 5305. These parameters are used in IPv4 RSVP-TE and IPv4 SR-TE.

When the **ipv6** command under the **traffic-engineering-options** context is also enabled, then the traffic engineering behavior with IPv6 TE links is enabled. This IS-IS instance automatically advertises the new RFC 6119 IPv6 and TE TLVs and sub-TLVs as described in TE Attributes Supported in IGP and BGP-LS.

The **application-link-attributes** context allows the advertisement of the TE attributes of each link on a per-application basis. Two applications are supported in SR OS: RSVP-TE and SR-TE. The legacy mode of advertising TE attributes that is used in RSVP-TE is still supported but can be disabled by using the **no legacy** command that enables the per-application TE attribute advertisement for RSVP-TE as well.

Additional details of the feature behavior and the interaction of the previously mentioned CLI commands are described in IS-IS IPv4/IPv6 SR-TE and IPv4 RSVP-TE Feature Behavior.

2.5.3 MPLS Configuration

The SR-TE LSP configuration can accept an IPv6 address into the **to** and **from** parameters.

In addition, the MPLS path configuration can accept a hop index with an IPv6 address. The IPv6 address used in the **from** and **to** commands in the IPv6 SR-TE LSP, as well as the address used in the **hop** command of the path used with the IPv6 SR-TE LSP must correspond to the preferred primary global unicast IPv6 address of a network interface or a loopback interface of the corresponding LER or LSR router. The IPv6 address can also be set to the system interface IPv6 address. Failure to follow the preceding IPv6 address guidelines for the **from**, **to** and **hop**, commands causes path computation to fail with failure code “noCspfRouteToDestination.

Link-local IPv6 address of a network interface is also not allowed in the **hop** command of the path used with the IPv6 SR-TE LSP.

All other MPLS level, LSP level, and primary or secondary path level configuration parameters available for a IPv4 SR-TE LSP are supported unless indicated otherwise.

2.5.4 IS-IS, BGP-LS and TE Database Extensions

IS-IS control plane extensions add support for the following RFC 6119 TLVs in IS-IS advertisements and in TE-DB.

- IPv6 interface Address TLV (ISIS_TLV_IPv6_IFACE_ADDR 0xe8)
- IPv6 Neighbor Address sub-TLV (ISIS_SUB_TLV_NBR_IPADDR6 0x0d)
- IPv6 Global Interface Address TLV (only used by ISIS in IIH PDU)
- IPv6 TE Router ID TLV
- IPv6 SRLG TLV

IS-IS also supports advertising which protocol is enabled on a given TE-link (SR-TE, RSVP-TE, or both) by using the Application Specific Link Attributes (ASLA) sub-TLV as per *draft-ietf-isis-te-app*. This causes the advertising router to send potentially different Link TE attributes for RSVP-TE and SR-TE applications and allows the router receiving the link TE attributes to know which application is enabled on the advertising router. For backward compatibility, the router continues to support the legacy mode of advertising link TE attributes, as recommended in RFC 5305, but the user can disable it.

Note: SR OS does not support configuring and advertising different link TE attribute values for RSVP-TE and SR-TE applications. The router advertises the same link TE attributes for both RSVP-TE and SR-TE applications.

See IS-IS IPv4/IPv6 SR-TE and IPv4 RSVP-TE Feature Behavior for more details of the behavior of the per-application TE capability.

The new TLVs and sub-TLVs are advertised in IS-IS and added into the local TE-DB when received from IS-IS neighbors. In addition, if the **database-export** command is enabled in this ISIS instance, then this information is also added in the Enhanced TE-DB.

This feature adds the following enhancements to support advertising of the TE parameters in BGP-LS routes over a IPv4 or IPv6 transport:

- Importing IPv6 TE link TLVs from a local Enhanced TE-DB into the local BGP process for exporting to other BGP peers using the BGP-LS route family that is enabled on an IPv4 or an IPv6 transport BGP session
  - RFC 6119 IPv6 and TE TLVs and sub-TLVs are carried in BGP-LS link NLRI as per RFC 7752
  - When the link TE attributes are advertised by IS-IS on a per-application basis using the ASLA TLV (ISIS TLV Type 16), then they are carried in the new BGP-LS ASLA TLV (TLV Type TBD) as per *draft-ietf-idr-bgp-ls-app-specific-attr*.
  - When a TE attribute of a given link is advertised for both RSVP-TE and SR-TE applications, there are three methods IS-IS can use. Each method results in a specific way the BGP-LS originator carries this information. These methods are summarized here but more details are provided in IS-IS IPv4/IPv6 SR-TE and IPv4 RSVP-TE Feature Behavior.
    - In legacy mode of operation, all TE attributes are carried in the legacy IS-IS TE TLVs and the corresponding BGP-LS link attributes TLVs as listed in Table 14.
    - In legacy with application indication mode of operation, IGP and BGP-LS advertises the legacy TE attribute TLVs and also advertises the ASLA TLV with the legacy (L) flag set and the RSVP-TE and SR-TE application flags set. No TE sub-sub TLVs are advertised within the ASLA TLV.

      The legacy with application indication mode is intended for cases where link TE attributes are common to RSVP-TE and SR-TE applications and have the same value, but the user wants to indicate on a per-link basis which application is enabled.
    - In application specific mode of operation, the TE attribute TLVs are sent as sub-sub-TLVs within the ASLA TLV. Common attributes to RSVP-TE and SR-TE applications have the main TLV Legacy (L) flag cleared and the RSVP-TE and SR-TE application flags set. Any attribute that is specific to an application (RSVP-TE or SR-TE) is advertised in a separate ASLA TLV with the main TLV Legacy (L) flag cleared and the specific application (RSVP-TE or SR-TE) flags set.

      The application specific mode of operation is intended for future cases where TE attributes may have different values in RSVP-TE and SR-TE applications or are specific to one application (for example, the RSVP-TE ’Unreserved Bandwidth' and `Max Reservable Bandwidth' attributes).
- Exporting from the local BGP process to the local Enhanced TE-DB of IPv6 and TE link TLVs received from a BGP peer via BGP-LS route family enabled on a IPv4 or IPv6 transport BGP session

- Support of exporting of IPv6 and TE link TLVs from local Enhanced TE-DB to NSP via the CPROTO channel on the VSR-NRC

2.5.4.1 BGP-LS Originator Node Handling of TE Attributes

The specification of the BGP-LS originator node in support of the ASLA TLV is written with the following main objectives in mind:

1. Accommodate IGP node advertising the TE attribute in both legacy or application specific modes of operation.
2. Allow BGP-LS consumers (for example, PCE) that support the ASLA TLV to receive per-application attributes, even if the attribute values are duplicate, and easily store them per-application in the TE-DB. Also, if the BGP-LS consumers receive the legacy attributes, then they can make a determination without ambiguity that these attributes are only for RSVP-TE LSP application.
3. Continue supporting older BGP-LS consumers that rely only on the legacy attributes. This support is taken care by the backward compatibility mode described below but is not supported in SR OS.

The following are the changes needed on the BGP-LS originator node to support objectives (1) and (2). Excerpts are directly from *draft-ietf-idr-bgp-ls-app-specific-attr*:

1. *Application specific link attributes received from an IGP node using existing RSVP-TE/GMPLS encodings only (i.e. without any ASLA sub-TLV) MUST be encoded using the respective BGP-LS top-level TLVs listed in Table 1 (i.e. not within ASLA TLV). When the IGP node is also SR enabled then another copy of application specific link attributes SHOULD be also encoded as ASLA sub-TLVs with the SR application bit for them. Further rules do not apply for such IGP nodes that do not use ASLA sub-TLVs in their advertisements*.
2. *In case of IS-IS, when application specific link attributes are received from a node with the L bit set in the ASLA sub-TLV then the application specific link attributes are picked up from the legacy ISIS TLVs/sub-TLVs and MUST be encoded within the BGP-LS ASLA TLV as sub-TLVs with the application bitmask set as per the IGP ASLA sub-TLV. When the ASLA sub-TLV with the L bit set also has the RSVP-TE application bit set then the link attributes from such an ASLA sub-TLV MUST be also encoded using the respective BGP-LS top-level TLVs listed in Table 1 (i.e. not within ASLA TLV).*

3. *In case of OSPFv2/v3, when application specific link attributes are received from a node via TE LSAs then the application specific link attributes from those LSAs MUST be encoded using the respective BGP-LS TLVs listed in Table 1 (i.e. not within ASLA TLV).*
4. *Application specific link attributes received from an IGP node within its ASLA sub-TLV MUST be encoded in the BGP-LS ASLA TLV as sub-TLVs with the application bitmask set as per the IGP advertisement.*

The following are the changes needed on the BGP-LS originator node to support of objective (3) and which is referred to as the backward compatibility mode. Excerpts are directly from *draft-ietf-idr-bgp-ls-app-specific-attr*:

Note: The backward compatibility mode is not supported in SR OS.

1. *Application specific link attribute received in IGP ASLA sub-TLVs, corresponding to RSVP-TE or SR applications, MUST be also encoded in their existing top level TLVs (as listed in Table 1) outside of the ASLA TLV in addition to them being also advertised within the ASLA TLV*
2. *When the same application specific attribute, received in IGP ASLA sub-TLVs, has different values for RSVP-TE and SR applications then the value for RSVP-TE application SHOULD be preferred over the value for SR application for advertisement as the top level TLV (as listed in Table 1). An implementation MAY provide a knob to reverse this preference.*

2.5.4.2 TE Attributes Supported in IGP and BGP-LS

Table 14 lists the TE attributes that are advertised using the legacy link TE TLVs defined in RFC 5305 for IS-IS and in RFC 3630 for OSPF. These TE attributes are carried in BGP-LS as recommended in RFC 7752. These legacy TLVs are already supported in SR OS and in IS-IS, OSPF and BGP-LS.

To support IPv6 Traffic Engineering, the IS-IS IPv6 TE attributes (IPv6 TE Router ID and IPv6 SRLG TLV) are advertised in BGP-LS as recommended in RFC 7752.

All the above attributes can now be advertised within the ASLA TLV in IS-IS as recommended in *draft-ietf-isis-te-app* and in BGP-LS as recommended in *draft-ietf-idr-bgp-ls-app-specific-attr*. In the latter case, BGP-LS uses the same TLV type as in RFC 7752 but is included as a sub-TLV of the new BGP-LS ASLA TLV. Table 14 lists the code points for IS-IS and BGP-LS TLVs.

Table 14 **Legacy Link TE TLV Support in TE-DB and BGP-LS**

Link TE TLV Description	IS-IS TLV Type (RFC 5305)	OSPF TLV Type (RFC 3630)	BGP-LS Link NLRI Link-Attribute TLV Type (RFC 7752)
Administrative group (color)	3	9	1088
Maximum link bandwidth	9	6	1089
Maximum reservable link bandwidth	10	7	1090
Unreserved bandwidth	11	8	1091
TE Default Metric	18	5	1092
SRLG	138 (RFC 4205)	16 (RFC 4203)	1096
IPv6 SRLG TLV	139 (RFC 6119)	—	1096
IPv6 TE Router ID	140 (RFC 6119)	—	1029
Application Specific Link Attributes	16 (draft-ietf-isis-te-app)	—	1122 (provisional-as per draft-ietf-idr-bgp-ls-app-specific-attr)
Application Specific SRLG TLV	238 (draft-ietf-isis-te-app)	—	1122 (provisional-as per draft-ietf-idr-bgp-ls-app-specific-attr

Table 15 lists the TE attributes that are received from a third-party router implementation in legacy TE TLVs, or in the ASLA TLV for the RSVP-TE or SR-TE applications that are added into the local SR OSTE-DB; these are also distributed by the BGP-LS originator. However, these TLVs are not originated by a SR OS router IGP implementation.

Table 15 **Additional Link TE TLV Support in TE-DB and BGP-LS**

Link TE TLV Description	IS-IS TLV Type (RFC 7810)	OSPF TLV Type (RFC 7471)	BGP-LS Link NLRI Link-Attribute TLV Type (draft-ietf-idr-te-pm-bgp)
Unidirectional Link Delay	33	27	1114
Min/Max Unidirectional Link Delay	34	28	1115
Unidirectional Delay Variation	35	29	1116

Table 15 **Additional Link TE TLV Support in TE-DB and BGP-LS (Continued)**

Link TE TLV Description	IS-IS TLV Type (RFC 7810)	OSPF TLV Type (RFC 7471)	BGP-LS Link NLRI Link-Attribute TLV Type (draft-ietf-idr-te-pm-bgp)
Unidirectional Link Loss	36	30	1117
Unidirectional Residual Bandwidth	37	31	1118
Unidirectional Available Bandwidth	38	32	1119
Unidirectional Utilized Bandwidth	39	33	1120

Any other TE attribute received in a legacy TE TLV or in an Application Specific Link Attributes TLV is not added to the local router TE-DB and therefore, not distributed by the BGP-LS originator.

2.5.5 IS-IS IPv4/IPv6 SR-TE and IPv4 RSVP-TE Feature Behavior

The TE feature in IS-IS allows the advertising router to indicate to other routers in the TE domain which applications the advertising router has enabled: RSVP-TE, SR-TE, or both. As a result, a receiving router can safely prune links that are not enabled in one of the applications from the topology when computing a CSPF path in that application.

TE behavior consists of the following steps.

1. A valid IPv6 address value must exist for the system or loopback interface assigned to the **ipv6-te-router-id** command. The IPv6 address value can be either the preferred primary global unicast address of the system interface (default value) or that of a loopback interface (user configured).

   The IPv6 TE router ID is mandatory for enabling IPv6 TE and enabling the router to be uniquely identified by other routers in an IGP TE domain as being IPv6 TE capable. If a valid value does not exist, then the IPv6 and TE TLVs described in IS-IS, BGP-LS and TE Database Extensions are not advertised.
2. The **traffic-engineering** command enables the existing traffic engineering behavior for IPv4 RSVP-TE and IPv4 SR-TE. Enable the **rsvp** context on the router and enable **rsvp** on the interfaces in order to have IS-IS begin advertising TE attributes in the legacy TLVs. By default, the **rsvp** context is enabled as soon as the **mpls** context is enabled on the interface. If **ipv6** knob is also enabled,

then the RFC 6119 IPv6 and TE link TLVs described above are advertised such that a router receiving these advertisements can compute paths for IPv6 SR-TE LSP in addition to paths for IPv4 RSVP-TE LSP and IPv4 SR-TE LSP. The receiving node cannot determine if truly IPv4 RSVP-TE, IPv4 SR-TE, or IPv6 SR-TE applications are enabled on the other routers. Legacy TE routers must assume that RSVP-TE is enabled on those remote TE links it received advertisements for.

3. When the **ipv6** command is enabled, IS-IS automatically begins advertising the RFC 6119 TLVs and sub-TLVs: IPv6 TE router ID TLV, IPv6 interface Address sub-TLV and Neighbor Address sub-TLV, or Link-Local Interface Identifiers sub-TLV if the interface has no global unicast IPv6 address. The TLVs and sub-TLVs are advertised regardless of whether TE attributes are added to the interface in the **mpls** context. The advertisement of these TLVs is only performed when the **ipv6** knob is enabled and **ipv6-routing** is enabled in this IS-IS instance and **ipv6-te-router-id** has a valid IPv6 address.

   A network IP interface is advertised with the Link-Local Interface identifiers sub-TLV if the network IP interface meets the following conditions:

   - network IP interface has link-local IPv6 address and no global unicast IPv6 address on the interface **ipv6** context
   - network IP interface has no IPv4 address and may or may not have the **unnumbered** option enabled on the interface **ipv4** context

4. The **application-link-attributes** command enables the ability to send the link TE attributes on a per-application basis and explicitly conveys that RSVP-TE or SR-TE is enabled on that link on the advertising router.

   Three modes of operation that are allowed by the **application-link-attributes** command.

   a. Legacy Mode: {**no application-link-attributes**}

      The **application-link-attributes** command is disabled by default and the **no** form matches the behavior described in list item 2. It enables the existing traffic engineering behavior for IPv4 RSVP-TE and IPv4 SR-TE. Only the RSVP-TE attributes are advertised in the legacy TE TLVs that are used by both RSVP-TE and SR-TE LSP CSPF in the TE domain routers. No separate SR-TE attributes are advertised.

      If the **ipv6** command is also enabled, then the RFC 6119 IPv6 and TE link TLVs are advertised in the legacy TLVs. A router in the TE domain receiving these advertisements can compute paths for IPv6 SR-TE LSP.

      If the user shuts down the **rsvp** context on the router or on a specific interface, the legacy TE attributes of all the MPLS interfaces or of that specific MPLS interface are not advertised. Routers can still compute SR-TE LSPs using those links but LSP path TE constraints are not enforced since the links appear in the TE Database as if they did not have TE parameters.

Table 14 shows the encoding of the legacy TE TLVs in both IS-IS and BGP-LS.

b. Legacy Mode with Application Indication: {**application-link-attributes** + **legacy**}

The legacy with application indication mode is intended for cases where link TE attributes are common to RSVP-TE and SR-TE applications and have the same value, but the user wants to indicate on a per-link basis which application is enabled.

IS-IS continues to advertise the legacy TE attributes for both RSVP-TE and SR-TE application and includes the new Application Specific Link Attributes TLV with the application flag set to RSVP-TE and/or SR-TE but without the sub-sub-TLVs. IS-IS also advertises the Application Specific SRLG TLV with the application flag set to RSVP-TE and/or SR-TE but without the actual values of the SRLGs.

Routers in the TE domain use these attributes to compute CSPF for IPv4 RSVP-TE LSP and IPv4 SR-TE LSP.

If the **ipv6** command is also enabled, then the RFC 6119 IPv6 and TE TLVs are advertised. A router in the TE domain that receives these advertisements can compute paths for IPv6 SR-TE LSP.

Note: The **segment-routing** command must be enabled in the IS-IS instance or the flag for the SR-TE application will not be set in the Application Specific Link Attributes TLV or in the Application Specific SRLG TLV.

To disable advertising of RSVP-TE attributes, shut down the **rsvp** context on the router. Note, however, doing so reverts to advertising the link SR-TE attributes using the Application Specific Link Attributes TLV and the TE sub-sub-TLVs as shown in Table 16. If legacy attributes were used, legacy routers wrongly interpret that this router enabled RSVP and may signal RSVP-TE LSP paths using its links.

Table 14 lists the code points for IS-IS and BGP-LS legacy TLVs.

The following excerpt from the Link State Database (LSDB) shows the advertisement of TE parameters for a link with both RSVP-TE and SR-TE applications enabled.

```
TE IS Nbrs   :
   Nbr   : Dut-A.00
   Default Metric  : 10
   Sub TLV Len     : 124
   IF Addr   : 10.10.2.3
   IPv6 Addr : 3ffe::10:10:2:3
   Nbr IP    : 10.10.2.1
   Nbr IPv6  : 3ffe::10:10:2:1
   MaxLink BW: 100000 kbps
   Resvble BW: 500000 kbps
```

```
Unresvd BW:
    BW[0] : 500000 kbps
    BW[1] : 500000 kbps
    BW[2] : 500000 kbps
    BW[3] : 500000 kbps
    BW[4] : 500000 kbps
    BW[5] : 500000 kbps
    BW[6] : 500000 kbps
    BW[7] : 500000 kbps
Admin Grp : 0x1
TE Metric : 123
TE APP LINK ATTR    :
     SABML-flags:Legacy SABM-flags:RSVP-TE SR-TE
Adj-SID: Flags:v4VL Weight:0 Label:524287
Adj-SID: Flags:v6BVL Weight:0 Label:524284
TE SRLGs     :
   SRLGs : Dut-A.00
   Lcl Addr  : 10.10.2.3
   Rem Addr  : 10.10.2.1
   Num SRLGs       : 1
          1003
TE APP SRLGs     :
    Nbr : Dut-A.00
    SABML-flags:Legacy SABM-flags: SR-TE
    IF Addr   : 10.10.2.3
    Nbr IP    : 10.10.2.1
```

c. Application Specific Mode: {**application-link-attributes**} or {**application-link-attributes** + **no legacy**}

   The application specific mode of operation is intended for future use cases where TE attributes may have different values in RSVP-TE and SR-TE applications (this capability is not supported in SR OS) or are specific to one application (for example, RSVP-TE 'Unreserved Bandwidth' and `Max Reservable Bandwidth' attributes).

   IS-IS advertises the TE attributes that are common to RSVP-TE and SR-TE applications in the sub-sub-TLVs of the new ASLA sub-TLV. IS-IS also advertises the link SRLG values in the Application Specific SRLG TLV. In both cases, the application flags for RSVP-TE and SR-TE are also set in the sub-TLV.

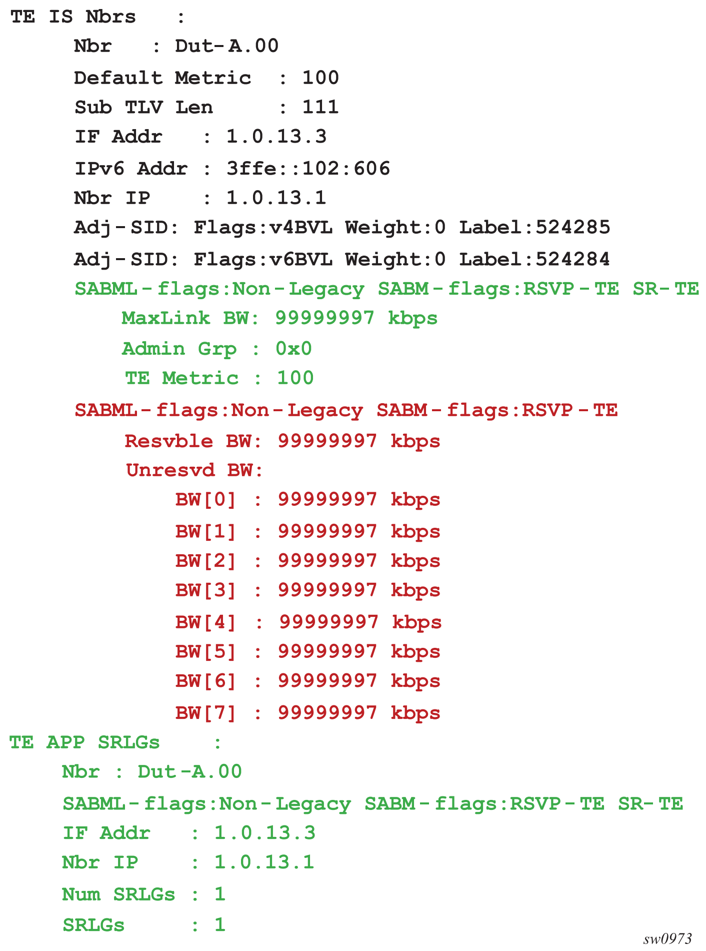
   IS-IS begins to advertise the TE attributes that are specific to the RSVP-TE application separately in the sub-sub-TLVs of the new application attribute sub-TLV. The application flag for RSVP-TE is also set in the sub-TLV.

   SR OS does not support configuring and advertising TE attributes that are specific to the SR-TE application.

   Common value RSVP-TE and SR-TE TE attributes are combined in the same application attribute sub-TLV with both application flags set, while the non-common value TE attributes are sent in their own application attribute sub-TLV with the corresponding application flag set.

Figure 30 shows an excerpt from the Link State Database (LSDB). Attributes in green font are common to both RSVP-TE and SR-TE applications and are combined, while the attribute in red font is specific to RSVP-TE application and is sent separately.

Figure 30 **Attribute Mapping per Application**



```
TE IS Nbrs   :
     Nbr   : Dut-A.00
     Default Metric  : 100
     Sub TLV Len     : 111
     IF Addr   : 1.0.13.3
     IPv6 Addr : 3ffe::102:606
     Nbr IP    : 1.0.13.1
     Adj-SID: Flags:v4BVL Weight:0 Label:524285
     Adj-SID: Flags:v6BVL Weight:0 Label:524284
     SABML-flags:Non-Legacy SABM-flags:RSVP-TE SR-TE
         MaxLink BW: 99999997 kbps
         Admin Grp : 0x0
         TE Metric : 100
     SABML-flags:Non-Legacy SABM-flags:RSVP-TE
         Resvble BW: 99999997 kbps
         Unresvd BW:
             BW[0] : 99999997 kbps
             BW[1] : 99999997 kbps
             BW[2] : 99999997 kbps
             BW[3] : 99999997 kbps
             BW[4] : 99999997 kbps
             BW[5] : 99999997 kbps
             BW[6] : 99999997 kbps
             BW[7] : 99999997 kbps
TE APP SRLGs    :
    Nbr : Dut-A.00
    SABML-flags:Non-Legacy SABM-flags:RSVP-TE SR-TE
    IF Addr   : 1.0.13.3
    Nbr IP    : 1.0.13.1
    Num SRLGs : 1
    SRLGs     : 1
```

Routers in the TE domain use these attributes to compute CSPF for IPv4 SR-TE LSP and IPv4 SR-TE LSPs. If the **ipv6** command is also enabled, then the RFC 6119 IPv6 TLVs are advertised. A router in the TE domain receiving these advertisements can compute paths for IPv6 SR-TE LSP.

Note: The **segment-routing** command must be enabled in the IS-IS instance or the common TE attribute will not be advertised for the SR-TE application.

In order to disable advertising of RSVP-TE attributes, shut down the **rsvp** context on the router.

Table 16 summarizes the IS-IS link TE parameter advertisement details for the three modes of operation of the IS-IS advertisement.

Table 16 **Details of Link TE Advertisement Methods**

IGP Traffic Engineering Options		**Link TE Advertisement Details**		
		RSVP-TE (rsvp enabled on interface)	**SR-TE (segment-routing enabled in IGP instance)**	**RSVP-TE and SR-TE (rsvp enabled on interface and segment-routing enabled in IGP instance)**
Legacy Mode: **no application-link-attributes**		Legacy TE TLVs	—	Legacy TE TLVs
Legacy Mode with Application Indication: {**application-link-attributes** + **legacy**}	**rsvp** disabled on router (**rsvp** operationally down on all interfaces)	—	Legacy TE TLVs ASLA TLV - Flags: {Legacy=0, SR-TE=1}; TE sub-sub-TLVs	Legacy TE TLVs ASLA TLV -Flags: {Legacy=1, RSVP-TE=0, SR-TE=1}
	rsvp enabled on router	Legacy TE TLVs ASLA TLV - Flags: {Legacy=1, RSVP-TE=1}	Legacy TE TLVs ASLA TLV - Flags: {Legacy=1, SR-TE=1}	Legacy TE TLVs ASLA TLV -Flags: {Legacy=1, RSVP-TE=1, SR-TE=1}

Table 16 **Details of Link TE Advertisement Methods (Continued)**

IGP Traffic Engineering Options	Link TE Advertisement Details		
	RSVP-TE (rsvp enabled on interface)	**SR-TE (segment-routing enabled in IGP instance)**	**RSVP-TE and SR-TE (rsvp enabled on interface and segment-routing enabled in IGP instance)**
Application Specific Mode: {**application-link-attributes**} or {**application-link-attributes** + **no legacy**}	ASLA TLV - Flags: {Legacy=0, RSVP-TE=1}; TE sub-sub-TLVs	ASLA TLV - Flags: {Legacy=0, SR-TE=1}; TE sub-sub-TLVs	ASLA TLV -Flags: {Legacy=0, RSVP-TE=1; SR-TE=1}; TE sub-sub-TLVs (common attributes) ASLA TLV -Flags: {Legacy=0, RSVP-TE=1}; TE sub-sub-TLVs (RSVP-TE specific attributes; e.g., Unreserved BW and Resvble BW) ASLA TLV -Flags: {Legacy=0, SR-TE=1}; TE sub-sub-TLVs (SR-TE specific attributes; not supported in SR OS 19.10.R1)

2.5.6 IPv6 SR-TE LSP Support in MPLS

This feature is supported with the hop-to-label, the local CSPF, and the PCE (PCC-initiated an PCE-initiated) path computation methods.

All capabilities of a IPv4 provisioned SR-TE LSP are supported with a IPv6 SR-TE LSP unless indicated otherwise. There are, however, some important differences with an IPv4 SR-TE LSP which are explained below.

The IPv6 address used in the **from** and **to** commands in the IPv6 SR-TE LSP, as well as the address used in the **hop** command of the path used with the IPv6 SR-TE LSP must correspond to the preferred primary global unicast IPv6 address of a network interface or a loopback interface of the corresponding LER or LSR router. The IPv6 address can also be set to the system interface IPv6 address. Failure to follow the preceding IPv6 address guidelines for the **from**, **to** and **hop**, commands causes path computation to fail with failure code “noCspfRouteToDestination. A Link-Local IPv6 address of a network interface is also not allowed in the **hop** command of the path used with the IPv6 SR-TE LSP. The configuration fails.

A TE link with no global unicast IPv6 address and only a link local IPv6 address can however be used in the path computation by the local CSPF. The address shown in the ‘Computed Hops’ and in the ‘Actual Hops’ fields of the output of the path **show** command uses the neighbor’s IPv6 TE router ID and the Link-Local Interface Identifier. The exceptions are if the interface is of type broadcast or is of type point-to-point but also has a local IPv4 address. Only the neighbor’s IPv6 TE router ID is shown as the Link-Local Interface Identifiers sub-TLV is not advertised in these situations.

The global MPLS IPv4 state UP value requires that the system interface be in the admin UP state and to have a valid IPv4 address.

The global MPLS IPv6 state UP value requires that the interface used for the IPv6 TE router ID be in admin UP state and to have a valid preferred primary IPv6 global unicast address.

The TE interface MPLS IPv4 state UP value requires the interface be in the admin UP state in the **router** context and the global MPLS IPv4 state be in UP state.

The TE interface MPLS IPv6 state UP value requires the interface be in the admin UP state in the **router** context and the global MPLS IPv6 state be in UP state.

2.5.6.1 IPv6 SR-TE auto-LSP

This feature provides for the auto-creation of an IPv6 SR-TE mesh LSP and for a IPv6 SR-TE one-hop LSP.

The SR-TE mesh LSP feature specifically binds an LSP template of type **mesh-p2p-srte** with one or more IPv6 prefix lists. When the Traffic Engineering database discovers a router, which has an IPv6 TE router ID matching an entry in the prefix list, it triggers MPLS to instantiate an SR-TE LSP to that router using the LSP parameters in the LSP template.

The SR-TE one-hop LSP feature specifically activates a LSP template of type **one-hop-p2p-srte**. In this case, the TE database keeps track of each TE link which comes up to a directly connected IGP TE neighbor. It then instructs MPLS to instantiate a SR-TE LSP with the following parameters:

- the source IPv6 address of the local router
- an outgoing interface matching the interface index of the TE-link
- a destination address matching the IPv6 TE router-id of the neighbor on the TE link

A new **family** CLI leaf is added to the LSP template configuration and must be set to the **ipv6** value. By default, this command is set to the **ipv4** value for backward compatibility. When establishing both IPv4 and IPv6 SR-TE mesh auto-LSPs with the same parameters and constraints, a separate LSP template of type **mesh-p2p-srte** must be configured for each address family with the **family** CLI leaf set to the IPv4 or IPv6 value. SR-TE one-hop auto-LSPs can only be established for either IPv4 or IPv6 family, but not both. The **family** leaf in the LSP template of type **one-hop-p2p-srte** should be set to the desired IP family value.

➡ **Note:** A IPv6 SR-TE auto-LSP can be reported to a PCE but cannot be delegated or have its paths computed by PCE.

All capabilities of an IPv4 SR-TE auto-LSP are supported with a IPv6 SR-TE auto-LSP unless indicated otherwise.

2.6 Advanced MPLS/RSVP Features

This section describes advanced MPLS/RSVP features.

2.6.1 Extending RSVP LSP to use Loopback Interfaces Other than Router-id

It is possible to configure the address of a loopback interface, other than the router-id, as the destination of an RSVP LSP, or a P2MP S2L sub-LSP. In the case of a CSPF LSP, CSPF searches for the best path that matches the constraints across all areas and levels of the IGP where this address is reachable. If the address is the router-id of the destination node, then CSPF selects the best path across all areas and levels of the IGP for that router-id; regardless of which area and level the router-id is reachable as an interface.

In addition, the user can now configure the address of a loopback interface, other than the router-id, as a hop in the LSP path hop definition. If the hop is strict and corresponds to the router-id of the node, the CSPF path can use any TE enabled link to the downstream node, based on best cost. If the hop is strict and does not correspond to the router-id of the node, then CSPF will fail.

2.6.2 LSP Path Change

The **tools perform router mpls update-path** {**lsp** *lsp-name* **path** *current-path-name* **new-path** *new-path-name*} command instructs MPLS to replace the path of the primary or secondary LSP.

The primary or secondary LSP path is indirectly identified via the current-path-name value. In existing implementation, the same path name cannot be used more than once in a given LSP name.

This command is also supported on an SNMP interface.

This command applies to both CSPF LSP and to a non-CSPF LSP. However, it will only be honored when the specified current-path-name has the adaptive option enabled. The adaptive option can be enabled the LSP level or at the path level.

The new path must be first configured in CLI or provided via SNMP. The **configure >router>mpls>path** *path-name* command is used to enter the path.

The command fails if any of the following conditions are satisfied:

- The specified current-path-name of this LSP does not have the adaptive option enabled.
- The specified new-path-name value does not correspond to a previously defined path.
- The specified new-path-name value exists but is being used by any path of the same LSP, including this one.

When the command is executed, MPLS performs the following procedures:

- MPLS performs a single MBB attempt to move the LSP path to the new path.
- If the MBB is successful, MPLS updates the new path.
  - MPLS writes the corresponding NHLFE in the data path if this path is the current backup path for the primary.
  - If the current path is the active LSP path, it will update the path, write the new NHLFE in the data path, which will cause traffic to switch to the new path.
- If the MBB is not successful, the path retains its current value.
- The update-path MBB has the same priority as the manual resignal MBB.

2.6.3 Manual LSP Path Switch

This feature provides a new command to move the path of an LSP from a standby secondary to another standby secondary.

The base version of the command allows the path of the LSP to move from a standby (or an active secondary) to another standby of the same priority. If a new standby path with a higher priority or a primary path comes up after the **tools perform** command is executed, the path re-evaluation command runs and the path is moved to the path specified by the outcome of the re-evaluation.

The CLI command for the base version is:

tools>perform>router>mpls>switch-path>lsp *lsp-name* **path** *path-name*

The sticky version of the command can be used to move from a standby path to any other standby path regardless of priority. The LSP remains in the specified path until this path goes down or the user performs the no form of the **tools perform** command.

The CLI commands for the sticky version are:

tools>perform>router>mpls>force-switch-path>lsp *lsp-name* **path** *path-name*

tools>perform>router>mpls>no force-switch-path lsp *lsp-name*

2.6.4 Make-Before-Break (MBB) Procedures for LSP/Path Parameter Configuration Change

When an LSP is switched from an existing working path to a new path, it is desirable to perform this in a hitless fashion. The Make-Before-Break (MBB) procedure consist of first signaling the new path when it is up, and having the ingress LER move the traffic to the new path. Only then the ingress LER tears down the original path.

MBB procedure is invoked during the following operations:

1. Timer based and manual resignal of an LSP path.
2. Fast-ReRoute (FRR) global revertive procedures.
3. Soft Pre-emption of an LSP path.
4. Traffic-Engineering (TE) graceful shutdown procedures.
5. Update of secondary path due to an update to primary path SRLG.
6. LSP primary or secondary path name change.
7. LSP or path configuration parameter change.

In a prior implementation, item 7 covers the following parameters:

1. Changing the primary or secondary path **bandwidth** parameter on the fly.
2. Enabling the **frr** option for an LSP.

This feature extends the coverage of the MBB procedure to most of the other LSP level and Path level parameters as follows:

1. Changes to include/exclude of admin groups at LSP and path levels.
   Enabling/disabling LSP level path-computation local-cspf option.
2. Enabling/disabling LSP level metric-type parameter.
3. Enabling/disabling LSP level propagate-admin-group option.
4. Enabling/disabling LSP level hop-limit option in the fast-reroute context.
5. Enabling the LSP level least-fill option.
6. Enabling/disabling LSP level adspec option.
7. Changing between node-protect and “no node-protect” (link-protect) values in the LSP level fast-reroute option.

8. Changing LSP primary or secondary path priority values (setup-priority and hold-priority).
9. Changing LSP primary or secondary path class-type value and primary path backup-class-type value.
10. Changing LSP level and path level hop-limit parameter value.
11. Enabling/disabling primary or secondary path record and record-label options.

This feature is not supported on a manual bypass LSP.

P2MP Tree Level Make-before-break operation is supported if changes are made to the following parameters on LSP-Template:

- Changing Bandwidth on P2MP LSP-Template.
- Enabling Fast Re-Route on P2MP LSP-Template.

2.6.5 Automatic Creation of RSVP-TE LSP Mesh

This feature enables the automatic creation of an RSVP point-to-point LSP to a destination node whose router-id matches a prefix in the specified peer prefix policy. This LSP type is referred to as auto-LSP of type mesh.

The user can associate multiple templates with the same or different peer prefix policies. Each application of an LSP template with a given prefix in the prefix list will result in the instantiation of a single CSPF computed LSP primary path using the LSP template parameters as long as the prefix corresponds to a router-id for a node in the TE database. Each instantiated LSP will have a unique LSP-id and a unique tunnel-ID.

Up to five (5) peer prefix policies can be associated with a given LSP template at all times. Each time the user executes the above command with the same or different prefix policy associations, or the user changes a prefix policy associated with an LSP template, the system re-evaluates the prefix policy. The outcome of the re-evaluation will tell MPLS if an existing LSP needs to be torn down or if a new LSP needs to be signaled to a destination address that is already in the TE database.

If a /32 prefix is added to (removed from) or if a prefix range is expanded (shrunk) in a prefix list associated with an LSP template, the same prefix policy re-evaluation described above is performed.

The trigger to signal the LSP is when the router with a router-id the matching a prefix in the prefix list appears in the TE database. The signaled LSP is installed in the Tunnel Table Manager (TTM) and is available to applications such as LDP-over-RSVP, resolution of BGP label routes, resolution of BGP, IGP, and static routes. It is, however, not available to be used as a provisioned SDP for explicit binding or auto-binding by services.

If the **one-hop** option is specified instead of a prefix policy, this command enables the automatic signaling of one-hop point-to-point LSPs using the specified template to all directly connected neighbors. This LSP type is referred to as auto-LSP of type one-hop. Although the provisioning model and CLI syntax differ from that of a mesh LSP only by the absence of a prefix list, the actual behavior is quite different. When the above command is executed, the TE database will keep track of each TE link that comes up to a directly connected IGP neighbor whose router-id is discovered. It then instructs MPLS to signal an 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 **auto-lsp** command with the **one-hop** option will result in one or more LSPs signaled to the neighboring router.

An auto-created mesh or one-hop LSP can have egress statistics collected at the ingress LER by adding the **egress-statistics** node configuration into the LSP template. The user can also have ingress statistics collected at the egress LER using the same **ingress-statistics** node in CLI used with a provisioned LSP. The user must specify the full LSP name as signaled by the ingress LER in the RSVP session name field of the Session Attribute object in the received Path message.

2.6.5.1 Automatic Creation of RSVP Mesh LSP: Configuration and Behavior

2.6.5.1.1 Feature Configuration

The user first creates an LSP template of type mesh P2P:

config>router>mpls>lsp-template *template-name* **mesh-p2p**

Inside the template the user configures the common LSP and path level parameters or options shared by all LSPs using this template.

Then the user references the peer prefix list which is defined inside a policy statement defined in the global policy manager.

config>router>mpls>auto-lsp lsp-template *template-name* **policy** *peer-prefix-policy*

The user can associate multiple templates with same or different peer prefix policies. Each application of an LSP template with a given prefix in the prefix list will result in the instantiation of a single CSPF computed LSP primary path using the LSP template parameters as long as the prefix corresponds to a router-id for a node in the TE database. This feature does not support the automatic signaling of a secondary path for an LSP. If the user requires the signaling of multiple LSPs to the same destination node, he/she must apply a separate LSP template to the same or different prefix list which contains the same destination node. Each instantiated LSP will have a unique LSP-id and a unique tunnel-ID. This feature also does not support the signaling of a non-CSPF LSP. The selection of the **no cspf** option in the LSP template is therefore blocked.

Up to 5 peer prefix policies can be associated with a given LSP template at all times. Each time the user executes the above command, with the same or different prefix policy associations, or the user changes a prefix policy associated with an LSP template, the system re-evaluates the prefix policy. The outcome of the re-evaluation will tell MPLS if an existing LSP needs to be torn down or a new LSP needs to be signaled to a destination address which is already in the TE database.

If a /32 prefix is added to (removed from) or if a prefix range is expanded (shrunk) in a prefix list associated with an LSP template, the same prefix policy re-evaluation described above is performed.

The user must perform a **no shutdown** command of the template before it takes effect. Once a template is in use, the user must shutdown the template before effecting any changes to the parameters except for those LSP parameters for which the change can be handled with the Make-Before-Break (MBB) procedures. These parameters are **bandwidth** and enabling **fast-reroute** with or without the **hop-limit** or **node-protect** options. For all other parameters, the user shuts down the template and once a it is added, removed or modified, the existing instances of the LSP using this template are torn down and re-signaled.

Finally the auto-created mesh LSP can be signaled over both numbered and unnumbered RSVP interfaces.

2.6.5.1.2 Feature Behavior

Whether the prefix list contains one or more specific /32 addresses or a range of addresses, an external trigger is required to indicate to MPLS to instantiate an LSP to a node which address matches an entry in the prefix list. The objective of the feature is to provide an automatic creation of a mesh of RSVP LSP to achieve automatic tunneling of LDP-over-RSVP. The external trigger is when the router with the router-id matching an address in the prefix list appears in the TE database. In the latter case, the TE database provides the trigger to MPLS which means this feature operates with CSPF LSP only.

Each instantiation of an LSP template results in RSVP signaling and installing state of a primary path for the LSP to the destination router. The auto- LSP is installed in the Tunnel Table Manager (TTM) and is available to applications such as LDP-over-RSVP, resolution of BGP label routes, resolution of BGP, IGP, and static routes. The auto-LSP can also be used for auto-binding by a VPRN service. The auto-LSP is however not available to be used in a provisioned SDP for explicit binding by services. Therefore, an auto-LSP can also not be used directly for auto-binding of a PW template with the **use-provisioned-sdp** option in BGP-AD VPLS or FEC129 VLL service. However, an auto-binding of a PW template to an LDP LSP, which is then tunneled over an RSVP auto-LSP is supported.

If the user changes the **bandwidth** parameter in the LSP template, an MBB is performed for all LSPs using the template. If however the **auto-bandwidth** option was enabled in the template, the bandwidth **parameter** change is saved but will only take effect at the next time the LSP bounces or is re-signaled.

Except for the MBB limitations to the configuration parameter change in the LSP template, MBB procedures for manual and timer based re-signaling of the LSP, for TE Graceful Shutdown and for soft pre-emption are supported.

Note that the use of the **tools perform router mpls update-path** command with a mesh LSP is not supported.

The **one-to-one** option under **fast-reroute** is also not supported.

If while the LSP is UP, with the bypass backup path activated or not, the TE database loses the router-id, it will perform an update to MPLS module which will state router-id is no longer in TE database. This will cause MPLS to tear down all mesh LSPs to this router-id. Note however that if the destination router is not a neighbor of the ingress LER and the user shuts down the IGP instance in the destination router, the router-id corresponding to the IGP instance will only be deleted from the TE database in the ingress LER after the LSA/LSP ages out. If the user brought back up the IGP instance before the LSA/LSP aged out, the ingress LER deletes and re-installs the same router-id at the receipt of the updated LSA/LSP. In other words, the RSVP LSPs destined to this router-id will get deleted and re-established. All other failure conditions will cause the LSP to activate the bypass backup LSP or to go down without being deleted.

2.6.5.1.3 Multi-Area and Multi-Instance Support

A router which does not have TE links within a given IGP area/level will not have its router-id discovered in the TE database by other routers in this area/level. In other words, an auto-LSP of type P2P mesh cannot be signaled to a router which does not participate in the area/level of the ingress LER.

A mesh LSP can however be signaled using TE links all belonging to the same IGP area even if the router-id of the ingress and egress routers are interfaces reachable in a different area. In this case, the LSP is considered to be an intra-area LSP.

If multiple instances of ISIS or OSPF are configured on a router, each with its own router-id value, the TE database in other routers are able to discover TE links advertised by each instance. In such a case, an instance of an LSP can be signaled to each router-id with a CSPF path computed using TE links within each instance.

Finally, if multiple instances of ISIS or OSPF are configured on a destination router each with the same router-id value, a single instance of LSP is signaled from other routers. If the user shuts down one IGP instance, this is **no op** as long as the other IGP instances remain up. The LSP will remain up and will forward traffic using the same TE links. The same behavior exists with a provisioned LSP.

2.6.5.1.4 Mesh LSP Name Encoding and Statistics

When the ingress LER signals the path of a mesh auto-LSP, it includes the name of the LSP and that of the path in the Session Name field of the Session Attribute object in the Path message. The encoding is as follows:

Session Name: <lsp-name::path-name>, where lsp-name component is encoded as follows:

TemplateName-DestIpv4Address-TunnelId

Where *DestIpv4Address* is the address of the destination of the auto-created LSP.

At ingress LER, the user can enable egress statistics for the auto-created mesh LSP by adding the following configuration to the LSP template:

```
config
    router
        [no] mpls
             lsp-template template-name mesh-p2p]
             no lsp-template template-name
                  [no] egress-statistics
                       accounting-policy policy-id
                       no accounting-policy
                       no] collect-stats
```

If there are no stat indices available when an LSP is instantiated, the assignment is failed and the egress-statistics field in the show command for the LSP path is in the operational DOWN state but in admin UP state.

An auto-created mesh LSP can also have ingress statistics enabled on the egress LER as long as the user specifies the full LSP name following the above syntax.

config>router>mpls>ingress-statistics>lsp *lsp-name* **sender** *ip-address*

2.6.5.2 Automatic Creation of RSVP One-Hop LSP: Configuration and Behavior

2.6.5.2.1 Feature Configuration

The user first creates an LSP template of type one-hop:

config>router>mpls>lsp-template t*emplate-name* **one-hop-p2p**

Then the user enables the automatic signaling of one-hop LSP to all direct neighbors using the following command:

config>router>mpls>auto-lsp lsp-template *template-name* **one-hop**

The LSP and path parameters and options supported in an LSP template of type **one-hop-p2p** are that same as in the LSP template of type **mesh-p2p** except for the parameter **from** which is not allowed in a template of type **one-hop-p2p**. The show command for the auto-LSP displays the actual outgoing interface address in the ‘from’ field.

Finally the auto-created one-hop LSP can be signaled over both numbered and unnumbered RSVP interfaces.

2.6.5.2.2 Feature Behavior

Although the provisioning model and CLI syntax differ from that of a mesh LSP only by the absence of a prefix list, the actual behavior is quite different. When the above command is executed, the TE database will keep track of each TE link which comes up to a directly connected IGP neighbor which router-id is discovered. It then instructs MPLS to signals an 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. Therefore, the **auto-lsp** command with the **one-hop** option will result in one or more LSPs signaled to the IGP neighbor.

Only the router-id of the first IGP instance of the neighbor which advertises a TE link will cause the LSP to be signaled. If subsequently another IGP instance with a different router-id advertises the same TE link, no action is taken and the existing LSP is kept up. If the router-id originally used disappears from the TE database, the LSP is kept up and is associated now with the other router-id.

The state of a one-hop LSP once signaled follows the following behavior:

- If the interface used by the TE link goes down or BFD times out and the RSVP interface registered with BFD, the LSP path moves to the bypass backup LSP if the primary path is associated with one.
- If while the one-hop LSP is UP, with the bypass backup path activated or not, the association of the TE-link with a router-id is removed in the TE databases, the one-hop LSP is torn down. This would be the case if the interface used by the TE link is deleted or if the interface is shutdown in the context of RSVP.
- If while the LSP is UP, with the bypass backup path activated or not, the TE database loses the router-id, it will perform two separate updates to MPLS module. The first one updates the loss of the TE link association which will cause action (B) above for the one-hop LSP. The other update will state router-id is no longer in TE database which will cause MPLS to tear down all mesh LSPs to this router-id. A shutdown at the neighbor of the IGP instance which advertised the router-id will cause the router-id to be removed from the ingress LER node immediately after the last IGP adjacency is lost and is not subject to age-out as for a non-directly connected destination router.

All other feature behavior, limitations, and statistics support are the same as for an auto-LSP of type **mesh-p2p**.

2.6.6 IGP Shortcut and Forwarding Adjacency

The RSVP-TE LSP or SR-TE LSP shortcut for IGP route resolution supports packet forwarding to IGP learned routes using an RSVP-TE LSP. This is also referred to as IGP shortcut. This feature instructs IGP to include RSVP-TE LSPs and SR-TE LSPs that originate on this node and terminate on the router ID of a remote node as direct links with a metric equal to the metric provided by MPLS. During the IP reach to determine the reachability of nodes and prefixes, LSPs are overlaid and the LSP metric is used to determine the subset of paths that are equal to the lowest cost to reach a node or a prefix. When computing the cost of a prefix that is resolved to the LSP, if the user enables the relative-metric option for this LSP, the IGP applies the shortest IGP cost between the endpoints of the LSP, plus the value of the offset, instead of using the LSP operational metric.

Note: Dijkstra will always use the IGP link metric to build the SPF tree and the LSP metric value does not update the SPF tree calculation.

When a prefix is resolved to a tunnel next hop, the packet is sent labeled with the label stack corresponding to the NHLFE of the RSVP LSP and the explicit-null IPv6 label at the bottom of the stack in the case of an IPv6 prefix. Any network event causing an RSVP LSP to go down will trigger a full SPF computation which may result in installing a new route over another RSVP LSP shortcut as tunnel next hop or over a regular IP next hop.

When **igp-shortcut** is enabled at the IGP instance level, all RSVP-TE and SR-TE LSPs originating on this node are eligible by default as long as the destination address of the LSP, as configured in **config>router>mpls>lsp>to**, corresponds to a router-id of a remote node. LSPs with a destination corresponding to an interface address or any other loopback interface address of a remote node are automatically not considered by IS-IS or OSPF. The user can, however, exclude a specific RSVP-TE LSP or a SR-TE LSP from being used as a shortcut for resolving IGP routes as explained in IGP Shortcut Feature Configuration.

It is specifically recommended to disable the **igp-shortcut** option on RSVP LSP which has the cspf option disabled unless the full explicit path of the LSP is provided in the path definition. MPLS tracks in RTM the destination or the first loose-hop in the path of a non CSPF LSP and as such this can cause bouncing when used within IGP shortcuts.

The SPF in OSPF or IS-IS only uses RSVP LSPs as forwarding adjacencies, IGP shortcuts, or as endpoints for LDP-over-RSVP. These applications of RSVP LSPs are mutually exclusive at the IGP instance level. If two or more options are enabled in the same IGP instance, forwarding adjacency takes precedence over the shortcut application which takes precedence over the LDP-over-RSVP application. The SPF in IGP uses SR-TE LSPs as IGP shortcuts only.

Table 17 summarizes the outcome in terms of RSVP LSP role of mixing these configuration options.

Table 17 **RSVP LSP Role As Outcome of LSP level and IGP level configuration options**

	IGP Instance level configurations					
LSP level configuration	advertise-tunnel-link enabled / igp-shortcut enabled / ldp-over-rsvp enabled	advertise-tunnel-link enabled / igp-shortcut enabled / ldp-over-rsvp disabled	advertise-tunnel-link enabled / igp-shortcut disabled / ldp-over-rsvp disabled	advertise-tunnel-link disabled / igp-shortcut disabled / ldp-over-rsvp disabled	advertise-tunnel-link disabled / igp-shortcut enabled / ldp-over-rsvp enabled	advertise-tunnel-link disabled / igp-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

The **igp-shortcut shutdown** command disables the resolution of IGP routes using IGP shortcuts.

2.6.6.1 IGP Shortcut Feature Configuration

The following CLI objects enable the resolution over IGP IPv4 shortcuts of IPv4 and IPv6 prefixes within an ISIS instance, of IPv6 prefixes within an OSPFv3 instance, and of IPv4 prefixes within an OSPFv2 instance.

```
A:Reno 194# configure router isis
           igp-shortcut
                 [no] shutdown
                 tunnel-next-hop
                      family {ipv4, ipv6}
                             resolution {any|disabled|filter|match-family-ip}
                             resolution-filter
                                   [no] rsvp
                                   exit
                              exit
                      exit
                 exit

A:Reno 194# configure router ospf
           igp-shortcut
                 [no] shutdown
                 tunnel-next-hop
                      family {ipv4}
                             resolution {any|disabled|filter|match-family-ip}
                             resolution-filter
                                   [no] rsvp
                                   exit
                              exit
                      exit
                 exit

A:Reno 194# configure router ospf3#
           igp-shortcut
                 [no] shutdown
                 tunnel-next-hop
                      family {ipv6}
                             resolution {any|disabled|filter}
                             resolution-filter
                                   [no] rsvp
                                   exit
                              exit
                      exit
                 exit
```

The new resolution node **igp-shortcut** is introduced to provide flexibility in the selection of the IP next hops or the tunnel types for each of the IPv4 and IPv6 prefix families.

When the IPv4 **family** option is enabled, the IS-IS or OSPF SPF includes the IPv4 IGP shortcuts in the IP reach calculation of IPv4 nodes and prefixes. RSVP-TE LSPs terminating on a node identified by its router ID can be used to reach IPv4 prefixes owned by this node or for which this node is the IPv4 next hop.

When the IPv6 **family** option is enabled, the IS-IS or OSPFv3 SPF includes the IPv4 IGP shortcuts in the IP reach calculation of IPv6 nodes and prefixes. RSVP-TE LSPs terminating on a node identified by its router ID can be used to reach IPv6 prefixes owned by this node or for which this node is the IPv6 next hop. The IPv6 option is supported in both ISIS MT=0 and MT=2.

The IS-IS or OSPFv3 IPv6 routes resolved to IPv4 IGP shortcuts are used to forward packets of IS-IS or OSPFv3 prefixes matching these routes but are also used to resolve the BGP next hop of BGP IPv6 prefixes, resolve the indirect next hop of static IPv6 routes, and forward CPM-originated IPv6 packets.

In the data path, a packet for an IPv6 prefix has a label stack that consists of the IPv6 Explicit-Null label value of 2 at the bottom of the label stack followed by the label of the IPv4 RSVP-TE LSP.

The following commands provide control of the use of an RSVP-TE LSP in IGP shortcuts:

- **config**>**router**>**mpls**>**lsp#** [**no**] **igp-shortcut lfa-protect** | **lfa-only**]
- **config**>**router**>**mpls**>**lsp# igp-shortcut relative-metric** *offset*

An LSP can be excluded from being used as an IGP shortcut for forwarding IPv4 and IPv6 prefixes, or the LSP in the LFA SPF can be used to protect the primary IP next hop of an IPv4 or IPv6 prefix.

2.6.6.1.1 IGP Shortcut Binding Construct

The SR OS **tunnel-next-hop** construct binds IP prefixes to IPv4 IGP shortcuts on a per-prefix family basis.

The following details the behavior of the construct.

- The construct supports the IPv4 and IPv6 families. It allows each family to resolve independently to either an IGP shortcut next hop using the unicast RTM or to the IP next hop using the multicast RTM.
- The **advertise-tunnel-link** (forwarding adjacency) takes priority over **igp-shortcut** if both CLI options are enabled. This is overall and not per family.
- The following commands are enabled based on the following relative priorities (from highest to lowest):
  - **advertise-tunnel-link** (IPv4 family with OSPFv2, IPv4, and IPv6 families with IS-IS MT=0 and IPv6 family in MT=2, no support in OSPFv3)
  - **igp-shortcut** (IPv4 family in OSPFv2, IPv6 family in OSPFv3, IPv4 and IPv6 families in IS-IS MT=0 and IPv6 family in MT=2)

  - **ldp-over-rsvp** (IPv4 FECs only)

  More details can be found in RSVP LSP Role As Outcome of LSP level and IGP level configuration options.

- No default behavior exists for IPv4 prefixes to automatically resolve to RSVP LSPs used as IGP shortcut by enabling the **igp-shortcut** context only. The IPv4 family must be enabled and the **resolution-filter** set to the value of **rsvp** which selects the RSVP-TE tunnel type.
- A [**no**] **shutdown** command under the **igp-shortcut** context enforces that the IGP shortcut context cannot be enabled unless at least one family is configured under the tunnel-next-hop node to a value other than **resolution disabled**, which is the default value for all families and that a tunnel type is selected if the **resolution** is set to **filter**.
- To disable IGP shortcuts globally, shutdown the **igp-shortcut** context.
- When computing the backup next hop of an IPv4 or IPv6 prefix, LFA considers the IP links and tunnels of the selected tunnel type which are the result of the configuration of the **tunnel-next-hop** for the IPv4 or IPv6 prefix family.

The resolution outcome for each of the IPv4 and IPv6 prefix families is summarized in Table 18. The description and behavior of the SRv4 and SRv6 families are described in SR Shortest Path Tunnel Over RSVP-TE IGP Shortcut Feature Configuration. The description and behavior of the sr-te resolution option using SR-TE IGP shortcuts are described in IPv4 IGP Shortcuts using SR-TE LSP Feature Configuration.

Table 18 **IGP Shortcut Binding Resolution Outcome**

	igp-shortcut CLI context	**IP family (v4/v6) CLI config**	**SR family (v4/v6) CLI config**	**IPv4 ECMP NH SET Computed**	**SRv4 ECMP NH SET Computed**	**IPv6 ECMP NH SET Computed**	**SRv6 ECMP NH SET Computed**
0	shutdown	—	—	IP (unicast RTM)	IP (mcast RTM)	IP (unicast RTM)	IP (mcast RTM)
1	no shutdown	resolution disabled	resolution disabled	IP (mcast RTM)	IP (mcast RTM)	IP (mcast RTM)	IP (mcast RTM)
			resolution match-family-ip	IP (mcast RTM)	IP (mcast RTM)	IP (mcast RTM)	IP (mcast RTM)
2	no shutdown	resolution-filter {rsvp}	resolution disabled	RSVP+IP	IP (mcast RTM)	RSVP+IP	IP (mcast RTM)
			resolution match-family-ip	RSVP+IP	RSVP+IP	RSVP+IP	RSVP+IP

Table 18 **IGP Shortcut Binding Resolution Outcome (Continued)**

	igp-shortcut CLI context	IP family (v4/v6) CLI config	SR family (v4/v6) CLI config	IPv4 ECMP NH SET Computed	SRv4 ECMP NH SET Computed	IPv6 ECMP NH SET Computed	SRv6 ECMP NH SET Computed
3	no shutdown	resolution-filter {sr-te}	resolution disabled	SRTE+IP	IP (mcast RTM)	SRTE+IP	IP (mcast RTM)
			resolution match-family-ip	SRTE+IP	IP (mcast RTM)	SRTE+IP	IP (mcast RTM)
4	no shutdown	resolution {any}/ resolution-filter {rsvp,sr-te}	resolution disabled	RSVP+IP	IP (mcast RTM)	RSVP+IP	IP (mcast RTM)
				SRTE+IP	IP (mcast RTM)	SRTE+IP	IP (mcast RTM)
			resolution match-family-ip	RSVP+IP	RSVP+IP	RSVP+IP	RSVP+IP
				SRTE+IP	IP (mcast RTM)	SRTE+IP	IP (mcast RTM)

2.6.6.2 IPv4 IGP Shortcuts using SR-TE LSP Feature Configuration

The configuration value of **sr-te** is added to the **resolution-filter** context of the **igp-shortcut** construct. When enabled, this value allows IGP to resolve IPv4 prefixes, IPv6 prefixes, and LDP IPv4 prefix FECs over SR-TE LSPs used as IGP shortcuts.

In addition, the value of **any** in the **resolution-filter** context allows the user to resolve IP prefixes and LDP FECs to either RSVP-TE or SR-TE LSPs used as IGP shortcuts.

```
A:Reno 194# configure router isis
           igp-shortcut
                 [no] shutdown
                 tunnel-next-hop
                      family {ipv4, ipv6}
                             resolution {any|disabled|filter|match-family-ip}
                             resolution-filter
                                   [no] rsvp
                                   [no] sr-te
                                   exit
                             exit
                      exit
                 exit


A:Reno 194# configure router ospf
           igp-shortcut
                 [no] shutdown
                 tunnel-next-hop
```

```
                      family {ipv4}
                             resolution {any|disabled|filter|match-family-ip}
                             resolution-filter
                                   [no] rsvp
                                   [no] sr-te
                                   exit
                             exit
                      exit
                 exit
```

```
A:Reno 194# configure router ospf3
           igp-shortcut
                 [no] shutdown
                 tunnel-next-hop
                      family {ipv6}
                             resolution {any|disabled|filter}
                             resolution-filter
                                   [no] rsvp
                                   [no] sr-te
                                   exit
                             exit
                      exit
                 exit
```

See Family Prefix Resolution and Tunnel Selection Rules for an explanation of the rules for the resolution of IPv4 prefixes, IPv6 prefixes, and LDP FECs, and for the selection of the tunnel types on a per family basis.

2.6.6.2.1 Family Prefix Resolution and Tunnel Selection Rules

The IGP instance SPF routine performs the Dijkstra tree calculation on the topology with IP links only and saves the information in both the unicast routing table and in the multicast routing table. It then performs the IP reach calculation in the multicast routing table for each prefix family that disabled IGP shortcuts. Concurrently, it lays the tunnels on the tree and performs the IP reach calculation in the unicast routing table for each prefix family that enabled IGP shortcuts.

The following are the details of the resolution of prefix families in the unicast or multicast routing tables.

a. OSPF supports IPv4 prefixes by enabling **family**=**ipv4**. IPv4 prefix resolution in the unicast routing table can mix IP and tunnel next hops with the preference given to tunnel next hops. A maximum of 64 ECMP tunnel and IP next hops can be programmed for an IPv4 prefix.
b. OSPFv3 supports IPv6 prefixes by enabling **family=ipv6**. IPv6 prefix resolution in the unicast routing table can mix IP and tunnel next hops with the preference given to tunnel next hops. A maximum of 64 ECMP tunnel and IP next hops can be programmed for an IPv6 prefix.

c. IS-IS supports IPv4 prefixes in MT=0 by enabling **family**=**ipv4** and **ipv6** prefixes in both MT=0 and MT=2 by enabling **family**=**ipv6**. IPv4 and IPv6 prefix resolution in the unicast routing table can mix IP and tunnel next hops with the preference given to tunnel next hops. A maximum of 64 ECMP tunnel and IP next hops can be programmed for an IPv4 or IPv6 prefix.

d. **family**=**ipv4** also enables the resolution in the unicast routing table of LDP IPv4 prefix FEC in OSPF or IS-IS. When prefer-tunnel-in-tunnel is enabled (disabled) in LDP, an LDP FEC selects tunnel next hops (IP next hops) only and does not mix these next hop types when both are eligible in the unicast routing table.

   A maximum of 32 ECMP tunnels next hops can be programmed for a LDP FEC.

   LDP IPv6 prefix FECs are not supported over IPv4 IGP shortcuts when enabling **family**=**ipv6**. A consequence of this is that if the corresponding IPv6 prefix resolves to tunnel next hops only, the LDP IPv6 prefix FEC will remain unresolved.

e. In all cases, the IP reach calculation in the unicast routing table will first follow the ECMP tunnel and IP next hop selection rules, described in ECMP Considerations, when resolving a prefix over IGP shortcuts. After the set of ECMP tunnel and IP next hops have been selected, the preference of tunnel type is then applied based on the user setting of the resolution of the family of the prefix. If the user enabled resolution of the prefix family to both RSVP-TE and SR-TE tunnel types, the TTM tunnel preference value is used to select one type for the prefix. In other words, RSVP-TE LSP type is preferred to a SR-TE LSP type on a per-prefix basis.

f. One or more SR-TE LSPs can be selected in the unicast routing table if **resolution**=**filter** and the **resolution-filter**=**sr-te**.

g. One or more SR-TE LSPs can also be selected in the unicast routing table if **resolution**=**any** and one or more SR-TE LSPs are available but no RSVP-TE LSPs are available for resolving the prefix by IGP.

h. An intra-area IP prefix of **family**=**ipv4**, or **family**=**ipv6**, or an LDP IPv4 prefix FEC always resolves to a single type of tunnel **rsvp-te** or **sr-te**. **rsvp-te** type is preferred if both types are allowed by the prefix family resolution and both types exist in the set of tunnel next hops of the prefix. The feature does not support mixing tunnel types per prefix.

i. An inter-area IP prefix of **family**=**ipv4**, or **family**=**ipv6**, or an LDP IPv4 prefix FECs always resolves to a single tunnel type and selects the tunnel next hops to the advertising ABR node from the most preferred tunnel type if the prefix family resolution allowed both types. If the prefix resolves to multiple ABR next hops, ABR nodes with the preferred tunnel type are selected. In other words, if RSVP-TE LSPs exist to at least one ABR node, ABR nodes that are the tail-end of only SR-TE LSPs will not be used in the set of ECMP tunnel next hops for the inter-area prefix.

j. The feature does not support configuring a different tunnel type per prefix family in **resolution-filter**. The **no shutdown** command within the **igp-shortcut** context fails if the user configured **family**=**ipv4** to resolve to **sr-te** and **family**=**ipv6** to **rsvp-te** or vice-versa. This is true for both inter-area and intra-area prefixes.

   The feature does, however, support selecting the best tunnel-type per prefix within each family as explained in (e). For instance, **family**=**ipv4** and **family**=**ipv6** can both configure **resolution**=**any**. On a per prefix-basis, the best tunnel type is selected, thus allowing both tunnel types to be deployed in the network.

k. The user can set **resolution**=**disabled** for each family independently, which disables IGP shortcuts for the corresponding prefix family in this IGP instance. IP Prefixes and LDP FECs of this family will resolve over IP links in the multicast routing table.

2.6.6.2.2 Application Support

SR-TE IGP shortcuts can be used in the following applications.

a. **family**=**ipv4** resolves IPv4 prefixes in RTM for the following:
   - IGP routes
   - indirect next hop of static routes
   - BGP next hop of BGP routes
   - LDP IPv4 prefix FEC

b. **family**=**ipv6** resolves IPv6 prefixes in RTM for the following:
   - IGP routes
   - indirect next hop of static routes
   - BGP next hop of BGP routes

c. When an LDP IPv4 FEC prefix is resolved to one or more SR-TE LSPs, then the following applications can resolve to LDP in TTM:
   - L2 service FECs
   - BGP next hop of VPN IPv4/IPv6 prefixes
   - BGP next hop of EVPN routes
   - BGP next hop of IPv4 prefixes
   - BGP next hop of IPv6 prefixes (6PE)
   - IGP IPv4 routes (ldp-shortcut feature)
   - indirect next hop of IPv4 static routes

d. When an LDP IPv4 FEC prefix is resolved to one or more SR-TE LSPs, then the following applications cannot resolve to LDP in TTM:
   - next hop of BGP LU routes

Note: Next hops of BGP LU routes cannot resolve to LDP in TTM because SR OS supports three levels of hierarchy in the data path and, because SR-TE LSP is a hierarchical LSP already, this makes the BGP-over-LDP-over-SRTE a 4-level hierarchy. BGP will keep these BGP-LU routes unresolved.

2.6.6.2.3 Loop-free Alternate (LFA) Protection Support

The following are the details of the Loop-free Alternate (LFA) Protection Support.

a. Prefixes that use one or more SR-TE LSPs as their primary next hops are automatically protected by one of the LFA features, base LFA, remote LFA, or TI-LFA, when enabled on any of the SR-TE LSPs.
b. If the user specifies the **lfa-only** option for a specified SR-TE LSP, then if the application prefix has a single IP primary next hop (no ECMP next hops). It is protected by an LFA backup, which can use an SR-TE LSP.

Note: The LFA SPF calculation cannot check that the outgoing interface of the protecting SR-TE LSP is different from the primary next hop of the prefix. The prefix will still be protected by either the ECMP next hops or the LFA backup next hop of the first segment of the protecting SR-TE LSP. This is a difference in behavior with that of an RSVP-TE LSP used with the **lfa-only** option. In that case, such an LSP is excluded from being used as a LFA backup next hop.

c. Application prefixes that resolve in TTM to an LDP IPv4 prefix FEC, which itself is resolved to one or more SR-TE LSPs, are equally protected either by the SR-TE LSP FRR (a) or the LDP LFA backup using an SR-TE LSP (b).
d. Assume **resolution**=**disabled** for one prefix family (for example, IPv6) while it is enabled to sr-te for the other (for example, IPv4). Also, assume a node is resolving an IPv6 prefix and an IPv4 prefix, both of which share the same downstream parent node in the Dijkstra tree. If the IPv4 prefix is protected by the LFA of one or more SR-TE LSP primary next hops (a), the feature supports computing a LFA IP backup next hop for the IPv6 prefix which is resolved to a IP primary next hop. This behavior aligns with the behavior over RSVP-TE LSP used as IGP shortcut for IPv6 and IPv4 prefixes.

e. Assume **resolution**=**disabled** for one prefix family (for example, IPv6) while it is enabled to sr-te for the other (for example, IPv4). Also, assume a node is resolving an IPv6 prefix and an IPv4 prefix, both of which share the same downstream parent node in the Dijkstra tree. If the IPv4 prefix resolves to a single primary IP next hop but is protected by the LFA backup next hop that uses an SR-TE LSP (b), the feature does not support computing an LFA IP backup next hop for IPv6 prefix, which then remains unprotected. This is a limitation of the feature that also exists with RSVP-TE LSP used as IGP shortcut for IPv6 and IPv4 prefixes.

This behavior also applies if the configuration of the resolution command for IPv4 and IPv6 families are reversed.

If the user enabled the remote LFA or the TI-LFA feature and enabled the use of SR IPv6 or SR IPv6 tunnels as an LFA backup next hop by the LDP IPv6 or IPv4 FEC prefix (LDP **fast-reroute backup-sr-tunnel** option), the LDP FEC is protected if such a backup SR tunnel is found.

2.6.6.3 SR Shortest Path Tunnel Over RSVP-TE IGP Shortcut Feature Configuration

Two prefix family values of **srv4** and **srv6** are added to the **igp-shortcut** construct.

When enabled, the **srv4** value allows IGP to resolve SR-ISIS IPv4 tunnels in MT=0 or SR-OSPF IPv4 tunnels over RSVP-TE LSPs used as IGP shortcuts.

When enabled, the **srv6** value allows IGP to resolve SR-ISIS IPv6 tunnels in MT=0 over RSVP-TE LSPs used as IGP shortcuts.

```
A:Reno 194# configure router isis
           igp-shortcut
                 [no] shutdown
                 tunnel-next-hop
                      family {srv4, srv6}
                             resolution {disabled | match-family-ip}
                             exit
                      exit
                 exit
```

```
A:Reno 194# configure router ospf
           igp-shortcut
                 [no] shutdown
                 tunnel-next-hop
                      family {srv4}
                             resolution {disabled | match-family-ip}
                             exit
                      exit
```

```
                exit
```

See Family Prefix Resolution and Tunnel Selection Rules for an explanation of the rules for the resolution of SR-ISIS IPv4 tunnels, SR-ISIS IPv6 tunnels, and SR-OSPF IPV4 tunnels, and the selection of the tunnel types on a per-family basis.

2.6.6.3.1 Family Prefix Resolution and Tunnel Selection Rules

The following are the details of the resolution of prefix families in the unicast or multicast routing tables.

a. **family**=**srv4** enables the resolution of SR-OSPF IPv4 tunnels and SR-ISIS IPv4 tunnels in MT=0 over RSVP-TE IPv4 IGP shortcuts. A maximum of 32 ECMP tunnel next hops can be programmed for an SR-OSPF or an SR-ISIS IPv4 tunnel.

b. **family**=**srv6** enables the resolution of SR-ISIS IPv6 tunnels in MT=0 over RSVP-TE IPv4 IGP shortcuts. A maximum of 32 ECMP tunnel next hops can be programmed for an SR-ISIS IPv6 tunnel.

Note: Segment routing is not supported in IS-IS MT=2.

c. One or more RSVP-TE LSPs can be selected if **resolution**=**match-family-ip** and the corresponding IPv4 or IPv6 prefix is resolved to RSVP-TE LSPs.

d. An SR tunnel cannot resolve to SR-TE IGP shortcuts. If **resolution**=**match-family-ip** and the corresponding IPv4 or IPv6 prefix is resolved to SR-TE LSPs, the SR tunnel is resolved to IP next hops in the multicast routing table.

e. For an SR tunnel corresponding to an inter-area prefix with best routes via multiple ABRs, setting **resolution**=**match-family-ip** means the SR tunnel can resolve to RSVP-TE LSPs to one or more ABR nodes. If, however, only SR-TE LSPs exist to any of the ABR nodes, IGP will not include this ABR in the selection of ECMP next hops for the tunnel. If there exists no RSVP-TE LSPs to all ABR nodes, the inter-area prefix is resolved to IP next hops in the multicast routing table.

Note: While this feature is intended to tunnel SR-ISIS IPv4 and IPv6 tunnels and SR-OSPF IPv4 tunnels over RSVP-TE IPv4 IGP shortcuts, an SR-TE LSP that has its first segment (ingress LER role) or its next segment (LSR role) correspond to one of these SR-ISIS or SR-OSPF tunnels will also be tunneled over RSVP-TE LSP.

f. **resolution**=**disabled** is the default value for the **srv4** and **srv6** families and means that SR-ISIS and SR-OSPF tunnels are resolved to IP links in the multicast routing table.

2.6.6.3.2 Application Support

The following describes how SR-ISIS IPv4 or IPv6 or a SR-OSPF IPv4 tunnels are resolved.

a. When an SR-ISIS IPv4 or an SR-OSPF IPv4 tunnel is resolved to one or more RSVP-TE LSPs, then the following applications can resolve to the SR-ISIS or SR-OSPF tunnel in TTM:
   - L2 service FECs
   - BGP next hop of VPN IPv4/IPv6 prefixes
   - BGP next hop of EVPN routes
   - BGP next hop of IPv4 prefixes
   - BGP next hop of IPv6 prefixes (6PE)
   - next hop of a BGP LU IPv4 route
   - indirect next hop of IPv4 static routes

b. When an SR-ISIS IPv6 tunnel is resolved to one or more RSVP-TE LSPs, then the following applications can resolve to the SR-ISIS tunnel in TTM:
   - L2 service FECs
   - next hop of VPN-IPv4 and VPN-IPv6 over a spoke SDP interface using the SR tunnel
   - indirect next hop of IPv6 static routes

c. When an SR-ISIS IPv4 or an SR-OSPF IPv4 tunnel is resolved to one or more RSVP-TE LSPs, then the following applications cannot resolve in TTM to a SR-TE LSP that is using an SR-ISIS or SR-OSPF segment:
   - next hop of a BGP LU route

Note: Next hops of BGP LU routes cannot resolve to LDP in TTM to a SR-TE LSP that is using an SR-ISIS or SR-OSPF segment because SR OS supports three levels of hierarchy in the data path and, because SR-TE LSP is a hierarchical LSP already, this makes the BGP-over-SRTE-over-RSVPTE a 4-level hierarchy. BGP will keep these BGP-LU routes unresolved.

2.6.6.3.3 Loop-free Alternate (LFA) Protection Support

The following are the details of the Loop-free Alternate (LFA) Protection Support.

a. Prefixes that resolve to one or more RSVP-TE LSPs as their primary next hops are automatically protected by RSVP-TE LSP FRR if enabled.

b. If the user specifies the l**fa-only** option for a specified RSVP-TE LSP, then if the SR-ISIS or SR-OSPF has a single IP primary next hop (no ECMP next hops), it is protected by a FRR backup that can use a RSVP-TE LSP.

c. Applications that resolve in TTM to an SR-ISIS or SR-OSPF, which itself is resolved to one or more RSVP-TE LSPs, will equally be protected either by the RSVP-TE LSP FRR (a) or the SR LFA using a RSVP-TE LSP (b).

d. Assume **family**=**ipv4** resolves to RSVP-TE in the unicast routing table while **family**=**srv4** resolves to IP links in the multicast routing table. If the IP prefix of an SR tunnel is resolved to a RSVP-TE LSP primary next hop, and is protected by RSVP-TE LSP FRR (a), this feature supports computing an LFA next hop for the SRv4 tunnel of the same prefix using IP next hops.

e. Assume **family**=**ipv4** or **family**=**ipv6** resolves to RSVP-TE in the unicast routing table while **family**=**srv4** or **family**=**srv6** resolves to IP links in the multicast routing table. If the IP prefix of an SRv4 or SRv6 tunnel is resolved to a single IP primary next hop and is protected by an SR LFA backup using an RSVP-TE LSP FRR (b), the feature does not support computing a LFA next hop for the SRv4 or SRv6 tunnel and remains unprotected.

   If, however, the user enabled the remote LFA or the TI-LFA feature, then an SR backup next hop may be found for the SR IPv4 or SR IPv6 tunnel, which then becomes protected.

2.6.6.4 Using LSP Relative Metric with IGP Shortcut

By default, the absolute metric of the LSP is used to compute the contribution of a IGP shortcut to the total cost of a prefix or a node after the SPF is complete. The absolute metric is the operational metric of the LSP populated by MPLS in the Tunnel Table Manager (TTM). This corresponds to the cumulative IGP-metric of the LSP path returned by CSPF or the static admin metric value of the LSP if the user configured one using the **config>router>mpls>lsp>metric** command. Note that MPLS populates the TTM with the maximum metric value of 16777215 in the case of a CSPF LSP using the TE-metric and a non-CSPF LSP with a loose or strict hop in the path. A non-CSPF LSP with an empty hop in the path definition returns the IGP cost for the destination of the LSP.

The user enables the use of the relative metric for an IGP shortcut with the following CLI command:

config>router>mpls>lsp>igp-shortcut relative-metric [*offset*]

IGP will apply the shortest IGP cost between the endpoints of the LSP plus the value of the offset, instead of the LSP operational metric, when computing the cost of a prefix which is resolved to the LSP.

The offset value is optional and it defaults to zero. An offset value of zero is used when the **relative-metric** option is enabled without specifying the offset parameter value.

The minimum net cost for a prefix is capped to the value of one (1) after applying the offset:

Prefix cost = max(1, IGP cost + relative metric offset)

Note that the TTM continues the show the LSP operational metric as provided by MPLS. In other words, applications such as LDP-over-RSVP (when IGP shortcut is disabled) and BGP and static route shortcuts will continue to use the LSP operational metric.

The **relative-metric** option is mutually exclusive with the **lfa-protect** or the **lfa-only** options. In other words, an LSP with the **relative-metric** option enabled cannot be included in the LFA SPF and vice-versa when the **igp-shortcut** option is enabled in the IGP.

Finally, it should be noted that the **relative-metric** option is ignored when forwarding adjacency is enabled in IS-IS or OSPF by configuring the **advertise-tunnel-link** option. In this case, IGP advertises the LSP as a point-to-point unnumbered link along with the LSP operational metric capped to the maximum link metric allowed in that IGP.

2.6.6.5 ECMP Considerations

When you enable ECMP on the system and multiple equal-cost paths exist for a prefix, the following selection criteria are used to pick up the set of next hops to program in the data path:

- for a destination = tunnel-endpoint (including external prefixes with tunnel-endpoint as the next hop):
  - select tunnel with lowest tunnel-index (ip next hop is never used in this case)
- for a destination != tunnel-endpoint:
  - exclude LSPs with metric higher than underlying IGP cost between the endpoint of the LSP

  - prefer tunnel next hop over ip next hop
  - within tunnel next hops:
    1. select lowest endpoint to destination cost
    2. if same endpoint to destination cost, select lowest endpoint node router-id
    3. if same router-id, select lowest tunnel-index
  - within ip next hops:
    - select lowest downstream router-id
    - if same downstream router-id, select lowest interface-index
- Although no ECMP is performed across both the IP and tunnel next hops, the tunnel endpoint lies in one of the shortest IGP paths for that prefix. As a result, the tunnel next hop is always selected as long as the prefix cost using the tunnel is equal or lower than the IGP cost.

The ingress IOM will spray the packets for a prefix over the set of tunnel next hops and IP next hops based on the hashing routine currently supported for IPv4 packets.

2.6.6.6 Handling of Control Packets

All control plane packets that require an RTM lookup and whose destination is reachable over the RSVP shortcut are forwarded over the shortcut. This is because RTM keeps a single route entry for each prefix unless there is ECMP over different outgoing interfaces.

Interface bound control packets are not impacted by the RSVP shortcut since RSVP LSPs with a destination address different than the router-id are not included by IGP in its SPF calculation.

2.6.6.7 Forwarding Adjacency

The forwarding adjacency feature can be enabled independently from the IGP shortcut feature in CLI. To enable forwarding adjacency, the user enters the following command in IS-IS or OSPF:

- **config>router>isis>advertise-tunnel-link**
- **config>router>ospf>advertise-tunnel-link**

If both **igp-shortcut** and **advertise-tunnel-link** options are enabled for a given IGP instance, then the **advertise-tunnel-link** will win. With this feature, ISIS or OSPF advertises an RSVP LSP as a link so that other routers in the network can include it in their SPF computations. An SR-TE LSP is not supported with forwarding adjacency. The RSVP LSP is advertised as an unnumbered point-to-point link and the link LSP/LSA has no TE opaque sub-TLVs as per RFC 3906 *Calculating Interior Gateway Protocol (IGP) Routes Over Traffic Engineering Tunnels*.

When the forwarding adjacency feature is enabled, each node advertises a p2p unnumbered link for each best metric tunnel to the router-id of any endpoint node. The node does not include the tunnels as IGP shortcuts in SPF computation directly. Instead, when the LSA/LSP advertising the corresponding P2P unnumbered link is installed in the local routing database, then the node performs an SPF using it like any other link LSA/LSP. The link bi-directional check requires that a link, regular link or tunnel link, exists in the reverse direction for the tunnel to be used in SPF.

The forwarding adjacency feature supports forwarding of both IPv4 and IPv6 prefixes. Specifically, it supports family IPv4 in OSPFv2, family IPv6 in OSPFv3, families IPv4 and IPv6 in ISIS MT=0, and family IPv6 in ISIS MT=2. Note that the **igp-shortcut** option under the LSP name governs the use of the LSP with both the **igp-shortcut** and the **advertise-tunnel-link** options in IGP. Table 19 describes the interactions of the actions of the forwarding adjacency feature.

Table 19 **Impact of LSP Level Configuration on IGP Shortcut and Forwarding Adjacency Features**

LSP Level Configuration	Actions with IGP Shortcut Feature	Actions with Forwarding Adjacency Feature
igp-shortcut	Tunnel is used in main SPF, but is not used in LFA SPF	Tunnel is advertised as a P2P link if it has the best LSP metric, is used in the main SPF if advertised, but is not used in LFA SPF
igp-shortcut lfa-protect	Tunnel is used in main SPF, and is used in LFA SPF	Tunnel is advertised as a P2P link if it has the best LSP metric, is used in the main SPF if advertised, and is used in LFA SPF regardless of whether it is advertised or not
igp-shortcut lfa-only	Tunnel is not used in main SPF, but is used in LFA SPF	Tunnel is not advertised as a P2P link, if not used in main SPF, but is used in LFA SPF

2.6.6.8 SR Shortest Path Tunnel Over RSVP-TE Forwarding Adjacency

This feature is enabled by configuring both the segment routing and forwarding adjacency features within an IS-IS instance in a multi-topology MT=0.

Both IPv4 and IPv6 SR-ISIS tunnels can be resolved and further tunneled over one or more RSVP-TE LSPs used as forwarding adjacencies.

This feature uses the following procedures.

- The forwarding adjacency feature only advertises into IS-IS RSVP-TE LSPs. SR-TE LSPs are not supported.
- An SR-ISIS tunnel (node SID) can have up to 32 next hops, some of which can resolve to a forwarding adjacency and some to a direct IP link. When the router **ecmp** value is configured lower than the number of next hops for the SR-ISIS tunnel, the subset of next hops selected prefers a forwarding adjacency over an IP link.
- In SR OS, ECMP and LFA are mutually exclusive on per-prefix basis. This is not specific to SR-ISIS but also applies to IP FRR, LDP FRR, and SR-ISIS FRR. If an SR-ISIS tunnel has one or more next hops that resolve to forwarding adjacencies, each next hop is protected by the FRR mechanism of the RSVP-TE LSP through which it is tunneled. In this case, LFA backup is not programmed by IS-IS.
- If an SR-ISIS tunnel has a single primary next hop that resolves to a direct link (not to a forwarding adjacency), base LFA may protect it if a loop-free alternate path exists. The LFA path may or may not use a forwarding adjacency.
- IS-IS does not compute a remote LFA or a TI-LFA backup for an SR-ISIS tunnel when forwarding adjacency is enabled in the IS-IS instance, even if these two types of LFAs are enabled in the configuration of that same IS-IS instance.

2.6.6.9 LDP Forwarding over IGP Shortcut

The user can enable LDP FECs over IGP shortcuts by configuring T-LDP sessions to the destination of the RSVP LSP. In this case, LDP FEC is tunneled over the RSVP LSP, effectively implementing LDP-over-RSVP without having to enable the **ldp-over-rsvp** option in OSPF or IS-IS. The **ldp-over-rsvp** and **igp-shortcut** options are mutually exclusive under OSPF or IS-IS.

2.6.6.10 LDP Forwarding over Static Route Shortcut Tunnels

Similar to LDP forwarding over IGP shortcut tunnels, the user can enable the resolution of LDP FECs over static route shortcuts by configuring T-LDP sessions and a static route that provides tunneled next hops corresponding to RSVP LSPs. In this case, indirect tunneled next hops in a static route are preferred over IP indirect next hops. For more information, refer to the *7450 ESS, 7750 SR, 7950 XRS, and VSR Router Configuration Guide*.

2.6.6.11 Handling of Multicast Packets

This feature supports multicast Reverse-Path Check (RPF) in the presence of IGP shortcuts. When the multicast source for a packet is reachable via an IGP shortcut, the RPF check fails since PIM requires a bi-directional path to the source but IGP shortcuts are unidirectional.

The implementation of the IGP shortcut feature provides IGP with the capability to populate the multicast RTM with the prefix IP next-hop when both the **igp-shortcut** option and the **multicast-import** option are enabled in IGP.

This change is made possible with the enhancement introduced by which SPF keeps track of both the direct first hop and the tunneled first hop of a node that is added to the Dijkstra tree.

Note that IGP will not pass LFA next-hop information to the mcast RTM in this case. Only ECMP next-hops are passed. As a consequence, features such as PIM Multicast-Only FRR (MoFRR) will only work with ECMP next-hops when IGP shortcuts are enabled.

Finally, note that the concurrent enabling of the **advertise-tunnel-link** option and the **multicast-import** option will result a multicast RTM that is a copy of the unicast RTM and is populated with mix of IP and tunnel NHs. RPF will succeed for a prefix resolved to a IP NH, but will fail for a prefix resolved to a tunnel NH. Table 20 summarizes the interaction of the **igp-shortcut** and **advertise-tunnel-link** options with unicast and multicast RTMs.

Table 20 **Impact of IGP Shortcut and Forwarding Adjacency on Unicast and Multicast RTM**

		Unicast RTM (Primary SPF)	**Multicast RTM (Primary SPF)**	**Unicast RTM (LFA SPF)**	**Multicast RTM (LFA SPF)**
OSPF	igp-shortcut	√	√ [1]	√	X [3]
	advertise-tunnel-link	√	√ [2]	√	√ [4]

Table 20 **Impact of IGP Shortcut and Forwarding Adjacency on Unicast and Multicast RTM**

		Unicast RTM (Primary SPF)	Multicast RTM (Primary SPF)	Unicast RTM (LFA SPF)	Multicast RTM (LFA SPF)
IS-IS	igp-shortcut	√	√ [1]	√	X [3]
	advertise-tunnel-link	√	√ [2]	√	√ [4]

Notes:

1. Multicast RTM is different from unicast RTM as it is populated with IP NHs only, including ECMP IP NHs. RPF check can be performed for all prefixes.
2. Multicast RTM is a copy of the unicast RTM and, so, is populated with mix of IP and tunnel NHs. RPF will succeed for a prefix resolved to a IP NH but will fail for a prefix resolved to a tunnel NH.
3. LFA NH is not computed for the IP primary next-hop of a prefix passed to multicast RTM even if the same IP primary next-hop ends up being installed in the unicast RTM. The LFA next-hop will, however, be computed and installed in the unicast RTM for a primary IP next-hop of a prefix.
4. Multicast RTM is a copy of the unicast RTM and, so, is populated with mix of IP and tunnel LFA NHs. RPF will succeed for a prefix resolved to a primary or LFA IP NH but will fail for a prefix resolved to a primary or LFA tunnel NH.

2.6.6.12 MPLS Entropy Label on Shortcut Tunnels

The router supports the MPLS entropy label (RFC 6790) on RSVP-TE LSPs used for IGP and BGP shortcuts. LSR nodes in a network can load-balance labeled packets in a more granular way than by hashing on the standard label stack. See MPLS Entropy Label and Hash Label for more information.

To configure insertion of the entropy label on IGP or BGP shortcuts, use the **entropy-label** command under the **configure>router** context.

2.6.7 Disabling TTL Propagation in an LSP Shortcut

This feature provides the option for disabling TTL propagation from a transit or a locally generated IP packet header into the LSP label stack when an RSVP LSP is used as a shortcut for BGP next-hop resolution, a static-route-entry next-hop resolution, or for an IGP route resolution.

A transit packet is a packet received from an IP interface and forwarded over the LSP shortcut at ingress LER.

A locally-generated IP packet is any control plane packet generated from the CPM and forwarded over the LSP shortcut at ingress LER.

TTL handling can be configured for all RSVP LSP shortcuts originating on an ingress LER using the following global commands:

config>router>mpls>[no] shortcut-transit-ttl-propagate
config>router>mpls>[no] shortcut-local-ttl-propagate

These commands apply to all RSVP LSPs which are used to resolve static routes, BGP routes, and IGP routes.

When the **no** form of the above command is enabled for local packets, TTL propagation is disabled on all locally generated IP packets, including ICMP Ping, trace route, and OAM packets that are destined to a route that is resolved to the LSP shortcut. In this case, a TTL of 255 is programmed onto the pushed label stack. This is referred to as pipe mode.

Similarly, when the **no** form is enabled for transit packets, TTL propagation is disabled on all IP packets received on any IES interface and destined to a route that is resolved to the LSP shortcut. In this case, a TTL of 255 is programmed onto the pushed label stack.

2.6.8 RSVP-TE LSP Signaling using LSP Template

An LSP template can be used for signaling RSVP-TE LSP to far-end PE node that is detected based on auto-discovery method by a client application. RSVP-TE P2MP LSP signaling based on LSP template is supported for Multicast VPN application on SR OS platform. LSP template avoids an explicit LSP or LSP S2L configuration for a node that is dynamically added as a receiver.

An LSP template has the option to configure TE parameters that apply to LSP that is set up using the template. TE options that are currently supported are:

- adaptive
- admin-group
- bandwidth
- CSPF calculation
- fast-reroute
- hop-limit
- record-label
- retry-timer

2.6.9 Shared Risk Link Groups

Shared Risk Link Groups (SRLGs) is a feature that allows the user to establish a backup secondary LSP path or a FRR LSP path which is disjoint from the path of the primary LSP. Links that are members of the same SRLG represent resources sharing the same risk, for example, fiber links sharing the same conduit or multiple wavelengths sharing the same fiber.

When the SRLG option is enabled on a secondary path, CSPF includes the SRLG constraint in the computation of the secondary LSP path. CSPF would return the list of SRLG groups along with the ERO during primary path CSPF computation. At a subsequent establishment of a secondary path with the SRLG constraint, the MPLS task queries again CSPF providing the list of SRLG group numbers to be avoided. If the primary path was not successfully computed, MPLS assumes an empty SRLG list for the primary. CSPF prunes all links with interfaces which belong to the same SRLGs as the interfaces included in the ERO of the primary path. If CSPF finds a path, the secondary is setup. If not, MPLS keeps retrying the requests to CSPF.

When the SRLG option is enabled on FRR, CSPF includes the SRLG constraint in the computation of a FRR detour or bypass for protecting the primary LSP path. CSPF prunes all links with interfaces which belong to the same SRLG as the interface, which is being protected, that is, the outgoing interface at the PLR the primary path is using. If one or more paths are found, the MPLS task selects one based on best cost and signals the bypass/detour. If not and the user included the strict option, the bypass/detour is not setup and the MPLS task keeps retrying the request to CSPF. Otherwise, if a path exists which meets the other TE constraints, other than the SRLG one, the bypass/detour is setup.

A bypass or a detour LSP is not intended to be SRLG disjoint from the entire primary path. This is because only the SRLGs of the outgoing interface at the PLR the primary path is using are avoided.

2.6.9.1 Enabling Disjoint Backup Paths

A typical application of the SRLG feature is to provide for an automatic placement of secondary backup LSPs or FRR bypass/detour LSPs that minimizes the probability of fate sharing with the path of the primary LSP (Figure 31).

The following details the steps necessary to create shared risk link groups:

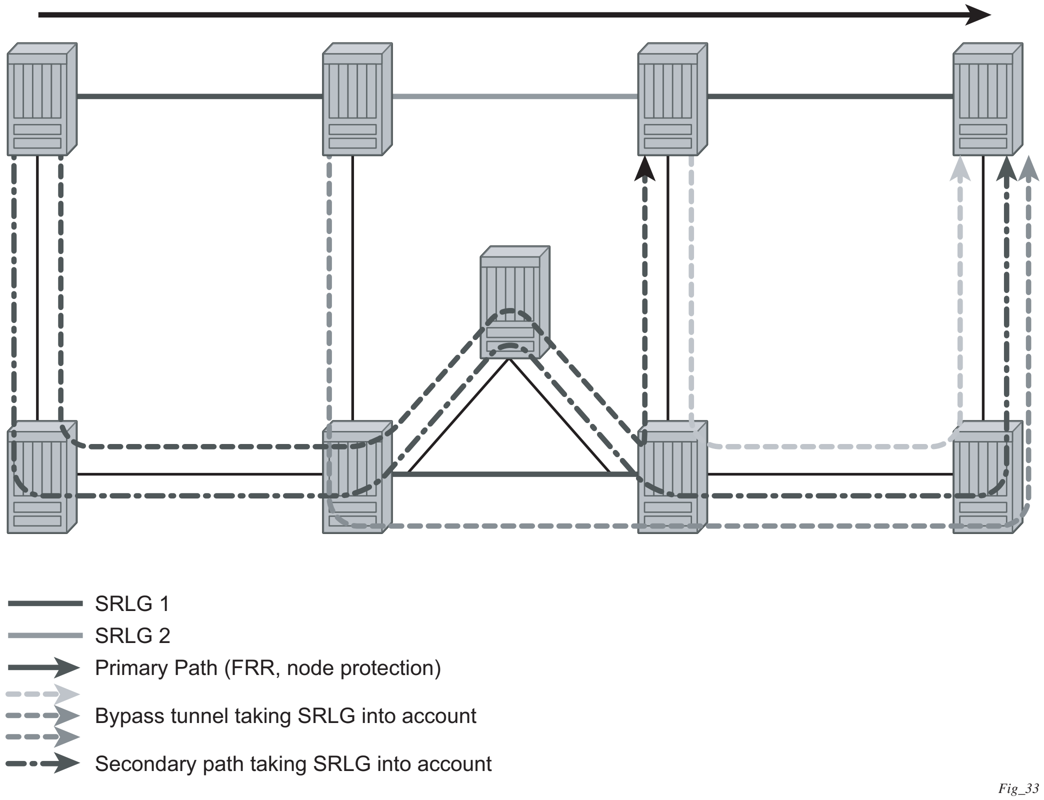
- For primary/standby SRLG disjoint configuration:
  - Create an SRLG-group, similar to admin groups.
  - Link the SRLG-group to MPLS interfaces.

    - Configure primary and secondary LSP paths and enable SRLG on the secondary LSP path. Note that the SRLG secondary LSP path(s) will *always* perform a strict CSPF query. The **srlg-frr** command is irrelevant in this case.
- For FRR detours/bypass SRLG disjoint configuration:
    - Create an SRLG group, similar to admin groups.
    - Link the SRLG group to MPLS interfaces.
    - Enable the **srlg-frr** (strict/non-strict) option, which is a system-wide parameter, and it force every LSP path CSPF calculation, to take the configured SRLG membership(s) (and propagated through the IGP opaque-te-database) into account.
    - Configure primary FRR (one-to-one/facility) LSP path(s). Consider that each PLR will create a detour/bypass that will only avoid the SRLG memberships configured on the primary LSP path egress interface. In a one-to-one case, detour-detour merging is out of the control of the PLR. As such, the latter will not ensure that its detour is prohibited to merge with a colliding one. For facility bypass, with the presence of several bypass type to bind to, the following priority rules are followed:

1. Manual bypass disjoint
2. Manual bypass non-disjoint (eligible only if srlg-frr is non-strict)
3. Dynamic disjoint
4. Dynamic non-disjoint (eligible only if srlg-frr is non-strict)

Non-CSPF manual bypass is not considered.

Figure 31 **Shared Risk Link Groups**


SRLG 1
SRLG 2
Primary Path (FRR, node protection)
Bypass tunnel taking SRLG into account
Secondary path taking SRLG into account
Fig_33

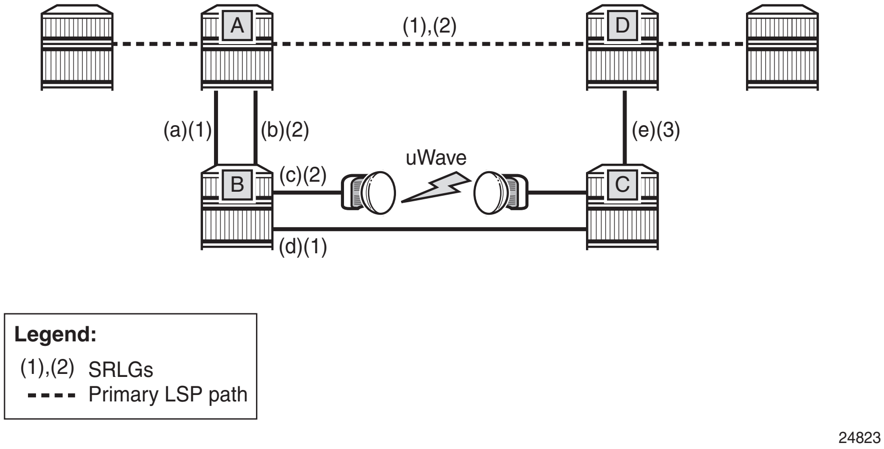
This feature is supported on OSPF and IS-IS interfaces on which RSVP is enabled.

2.6.9.2 SRLG Penalty Weights for Detour and Bypass LSPs

The likelihood of paths with links sharing SRLG values with a primary path being used by a bypass or detour LSP can be configured if a penalty weight is specified for the link. The higher the penalty weight, the less desirable it is to use the link with a given SRLG.

Figure 32 illustrates the operation of SRLG penalty weights.

Figure 32 **SRLG Penalty Weight Operation**



The primary LSP path includes a link between A and D with SRLG (1) and (2). The bypass around this link through nodes B and C includes links (a) and (d), which are members of SRLG (1), and links (b) and (c), which are members of SRLG 2. If the link metrics are equal, then this gives four ECMP paths from A to D via B and C:

- (a), (d), (e)
- (a), (c), (e)
- (b), (c), (e)
- (b), (d), (e)

Two of these paths include undesirable (from a reliability perspective) link (c). SRLG penalty weights or costs can be used to provide a tiebreaker between these paths so that the path including (c) is less likely to be chosen. For example, if the penalty associated with SRLG (1) is 5, and the penalty associated with SRLG (2) is 10, and the penalty associated with SRLG (3) is 1, then the cumulative penalty of each of the paths above is calculated by summing the penalty weights for each SRLG that a path has in common with the primary path:

- (a), (d), (e) = 10
- (a), (c), (e) = 15
- (b), (c), (e) = 20
- (b), (d), (e) = 15

Therefore path (a), (d), (e) is chosen since it has the lowest cumulative penalty.

Penalties are applied by summing the values for SRLGs in common with the protected part of the primary path.

A user can define a penalty weight value associate with an SRLG group using the **penalty-weight** parameter of the **srlg-group** command under the **configure**>**router-if-attribute** context. If an SRLG penalty weight is configured, then CSPF will include the SRLG penalty weight in the computation of an FRR detour or bypass for protecting the primary LSP path at a PLR node. Links with a higher SRLG penalty should be more likely to be pruned than links with a lower SRLG penalty.

Note that the configured penalty weight is not advertised in the IGP.

An SRLG penalty weight is applicable whenever an SRLG group is applied to an interface, including in the static SRLG database. However, penalty weights are used in bypass and detour path computation only when the srlg-frr (loose) flag is enabled.

2.6.9.3 Static Configurations of SRLG Memberships

This feature provides operations with the ability to manually enter the link members of SRLG groups for the entire network at any SR OS which will need to signal LSP paths (for example, a head-end node).

The operator may explicitly enable the use by CSPF of the SRLG database. In that case, CSPF will not query the TE database for IGP advertised interface SRLG information.

Note, however, that the SRLG secondary path computation and FRR bypass/detour path computation remains unchanged.

There are deployments where the SR OS will interoperate with routers that do not implement the SRLG membership advertisement via IGP SRLG TLV or sub-TLV.

In these situations, the user is provided with the ability to enter manually the link members of SRLG groups for the entire network at any SR OS which will need to signal LSP paths, for example, a head-end node.

The user enters the SRLG membership information for any link in the network by using the **interface** *ip-int-name* **srlg-group** *group-name* command in the **config>router>mpls> srlg-database>router-id** context. An interface can be associated with up to 5 SRLG groups for each execution of this command. The user can associate an interface with up to 64 SRLG groups by executing the command multiple times. The user must also use this command to enter the local interface SRLG membership into the user SRLG database. The user deletes a specific interface entry in this database by executing the **no** form of this command.

The *group-name* must have been previously defined in the **srlg-group** *group-name* **value** *group-value* command in the **config**>**router**>**if-attribute**. The maximum number of distinct SRLG groups the user can configure on the system is 1024.

The parameter value for *router-id* must correspond to the router ID configured under the base router instance, the base OSPF instance or the base IS-IS instance of a given node. Note however that a single user SRLG database is maintained per node regardless if the listed interfaces participate in static routing, OSPF, IS-IS, or both routing protocols. The user can temporarily disable the use by CSPF of all interface membership information of a specific router ID by executing the **shutdown** command in the **config>router>mpls> srlg-database> router-id** context. In this case, CSPF will assume these interfaces have no SRLG membership association. The operator can delete all interface entries of a specific router ID entry in this database by executing the **no router-id** *router-address* command in the **config>router>mpls> srlg-database** context.

CSPF will not use entered SRLG membership if an interface is not listed as part of a router ID in the TE database. If an interface was not entered into the user SRLG database, it is assumed that it does not have any SRLG membership. CSPF will not query the TE database for IGP advertised interface SRLG information.

The operator enables the use by CSPF of the user SRLG database by entering the user-srlg-db enable command in the **config>router>mpls** context. When the MPLS module makes a request to CSPF for the computation of an SRLG secondary path, CSPF will query the local SRLG and computes a path after pruning links which are members of the SRLG IDs of the associated primary path. Similarly, when MPLS makes a request to CSPF for a FRR bypass or detour path to associate with the primary path, CSPF queries the user SRLG database and computes a path after pruning links which are members of the SRLG IDs of the PLR outgoing interface.

The operator can disable the use of the user SRLG database by entering the user-srlg-db disable in command in the **config>router>mpls** context. CSPF will then resumes queries into the TE database for SRLG membership information. However, the user SRLG database is maintained

The operator can delete the entire SRLG database by entering the **no srlg-database** command in the **config>router>mpls** context. In this case, CSPF will assume all interfaces have no SRLG membership association if the user has not disabled the use of this database.

2.6.10 TE Graceful Shutdown

Graceful shutdown provides a method to bulk re-route transit LSPs away from the node during software upgrade of a node. A solution is described in RFC 5817, *Graceful Shutdown in MPLS and Generalized MPLS Traffic Engineering Networks*. This is achieved in this RFC by using a PathErr message with a specific error code Local Maintenance on TE link required flag. When a LER gets this message, it performs a make-before-break on the LSP path to move the LSP away from the links/ nodes which IP addresses were indicated in the PathErr message.

Graceful shutdown can flag the affected link/node resources in the TE database so other routers will signal LSPs using the affected resources only as a last resort. This is achieved by flooding an IGP TE LSA/LSP containing link TLV for the links under graceful shutdown with the TE metric set to 0xffffffff and 0 as unreserved bandwidth.

2.6.11 Soft Preemption of Diff-Serv RSVP LSP

A Diff-Serv LSP can preempt another LSP of the same or of a different CT if its setup priority is strictly higher (numerically lower) than the holding priority of that other LSP.

2.6.12 Least-Fill Bandwidth Rule in CSPF ECMP Selection

When multiples equal-cost paths satisfy the constraints of a given RSVP LSP path, CSPF in the router head-end node will select a path so that LSP bandwidth is balanced across the network links. In releases prior to R7.0, CSPF used a random number generator to select the path and returned it to MPLS. In the course of time, this method actually balances the number of LSP paths over the links in the network; it does not necessarily balance the bandwidth across those links.

The least-fill path selection algorithm identifies the single link in each of the equal cost paths which has the least available bandwidth in proportion to its maximum reserved bandwidth. It then selects the path which has the largest value of this figure. The net effect of this algorithm is that LSP paths are spread over the network links over time such that percentage link utilization is balanced. When the least-fill option is enabled on an LSP, during a manual reset CSPF will apply this method to all path calculations of the LSP, also at the time of the initial configuration.

2.6.13 Inter-Area TE LSP (ERO Expansion Method)

Inter-area contiguous LSP scheme provides end-to-end TE path. Each transit node in an area can set up a TE path LSP based on TE information available within its local area.

A PE node initiating an inter-area contiguous TE LSP does partial CSPF calculation to include its local area border router as a loose node.

Area border router on receiving a PATH message with loose hop ERO does a partial CSPF calculation to the next domain border router as loose hop or CSPF to reach the final destination.

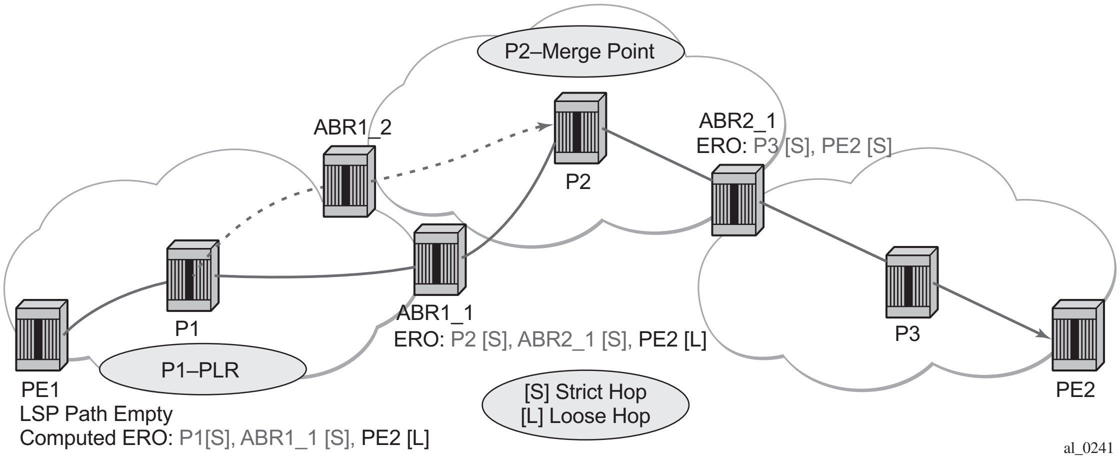
2.6.13.1 Area Border Node FRR Protection for Inter-Area LSP

This feature enhances the prior implementation of an inter-area RSVP P2P LSP by making the ABR selection automatic at the ingress LER. The user will not need to include the ABR as a loose-hop in the LSP path definition.

CSPF adds the capability to compute all segments of a multi-segment intra-area or inter-area LSP path in one operation.

Figure 33 illustrates the role of each node in the signaling of an inter-area LSP with automatic ABR node selection.

Figure 33 **Automatic ABR Node Selection for Inter-Area LSP**

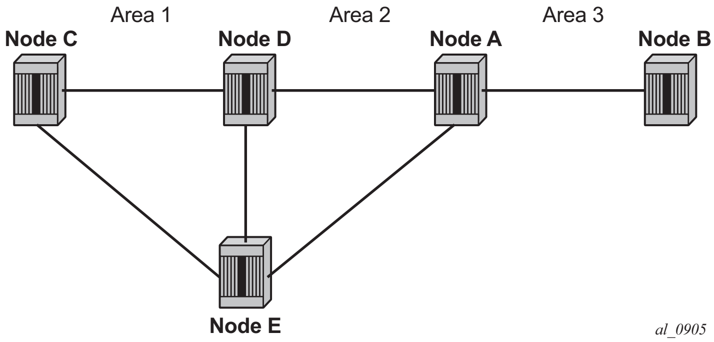


CSPF for an inter-area LSP operates as follows:

1. CSPF in the Ingress LER node determines that an LSP is inter-area by doing a route lookup with the destination address of a P2P LSP (that is the address in the to field of the LSP configuration). If there is no intra-area route to the destination address, the LSP is considered as inter-area.
2. When the path of the LSP is empty, CSPF will compute a single-segment intra-area path to an ABR node that advertised a prefix matching with the destination address of the LSP.
3. When the path of the LSP contains one or more hops, CSPF will compute a multi-segment intra-area path including the hops that are in the area of the Ingress LER node.
4. When all hops are in the area of the ingress LER node, the calculated path ends on an ABR node that advertised a prefix matching with the destination address of the LSP.
5. When there are one or more hops that are not in the area of the ingress LER node, the calculated path ends on an ABR node that advertised a prefix matching with the first hop-address that is not in the area of the ingress LER node.
6. Note the following special case of a multi-segment inter-area LSP. If CSPF hits a hop that can be reached via an intra-area path but that resides on an ABR, CSPF only calculates a path up to that ABR. This is because there is a better chance to reach the destination of the LSP by first signaling the LSP up to that ABR and continuing the path calculation from there on by having the ABR expand the remaining hops in the ERO.

   This behavior can be illustrated in the Figure 34. The TE link between ABR nodes D and E is in area 0. When node C computes the path for LSP from C to B which path specified nodes C and D as loose hops, it would fail the path computation if CSPF attempted a path all the way to the last hop in the local area, node E. Instead, CSPF stops the path at node A which will further expand the ERO by including link D-E as part of the path in area 0.

Figure 34 **CSPF for an Inter-area LSP**



7. If there is more than 1 ABR that advertised a prefix, CSPF will calculate a path for all ABRs. Only the shortest path is withheld. If more than one path has the shortest path, CSPF will pick a path randomly or based on the least-fill criterion if enabled. If more than one ABR satisfies the least-fill criterion, CSPF will also pick one path randomly.
8. The path for an intra-area LSP path will not be able to exit and re-enter the local area of the ingress LER. This behavior was possible in prior implementation when the user specified a loose hop outside of the local area or when the only available path was via TE links outside of the local area.

2.6.13.1.1 Rerouting of Inter-Area LSP

In prior implementation, an inter-area LSP path would have been re-routed if a failure or a topology change occurred in the local or a remote area while the ABR loose-hop in the path definition was still up. If the exit ABR node went down, went into IS-IS overload, or was put into node TE graceful shutdown, the LSP path will remain down at the ingress LER.

One new behavior introduced by the automatic selection of ABR is the ability of the ingress LER to reroute an inter-area LSP primary path via a different ABR in the following situations:

- When the local exit ABR node fails, There are two cases to consider:
  - The primary path is not protected at the ABR and, so, is torn down by the previous hop in the path. In this case the ingress LER will retry the LSP primary path via the ABR which currently has the best path for the destination prefix of the LSP.
  - The primary path is protected at the ABR with a manual or dynamic bypass LSP. In this case the ingress LER will receive a Path Error message with a notification of a protection becoming active downstream and a RESV with a *Local-Protection-In-Use* flag set. At the receipt of first of these two messages, the ingress LER will then perform a Global Revertive Make-Before-Break (MBB) to re-optimize the LSP primary path via the ABR which currently has the best path for the destination prefix of the LSP.
- When the local exit ABR node goes into IS-IS overload or is put into node TE Graceful Shutdown. In this case, the ingress LER will perform a MBB to re-optimize the LSP primary path via the ABR which currently has the best path for the destination prefix of the LSP. The MBB is performed at the receipt of the PathErr message for the node TE shutdown or at the next timer or manual re-optimization of the LSP path in the case of the receipt of the IS-IS overload bit.

2.6.13.1.2 Behavior of MPLS Options in Inter-Area LSP

The automatic ABR selection for an inter-area LSP does not change prior implementation inter-area LSP behavior of many of the LSP and path level options. There is, however, a number of enhancements introduced by the automatic ABR selection feature as explained in the following.

- Features such as path bandwidth reservation and admin-groups continue to operate within the scope of all areas since they rely on propagating the parameter information in the Path message across the area boundary.
- The TE graceful shutdown and soft preemption features will continue to support MBB of the LSP path to avoid the link or node that originated the PathErr message as long as the link or node is in the local area of the ingress LER. If the PathErr originated in a remote area, the ingress LER will not be able to avoid the link or node when it performs the MBB since it computes the path to the local ABR exit router only. There is, however, an exception to this for the TE graceful shutdown case only. An enhancement has been added to cause the upstream ABR nodes in the current path of the LSP to record the link or node to avoid and will use it in subsequent ERO expansions. This means that if the ingress LER computes a new MBB path which goes via the same exit ABR router as the current path and all ABR upstream nodes of the node or link which originated the PathErr message are also selected in the new MBB path when the ERO is expanded, the new path will indeed avoid this link or node. The latter is a new behavior introduced with the automatic ABR selection feature.
- The support of MBB to avoid the ABR node when the node is put into TE Graceful Shutdown is a new behavior introduced with the automatic ABR selection feature.
- The **metric-type te** option in CSPF cannot be propagated across the area boundary and will operate within the scope of the local area of the ingress LER node. This is a new behavior introduced with the automatic ABR selection feature.
- The **srlg** option on bypass LSP will continue to operate locally at each PLR within each area. The PLR node protecting the ABR will check the SRLG constraint for the path of the bypass within the local area.
- The **srlg** option on secondary path is allowed to operate within the scope of the local area of the ingress LER node with the automatic ABR selection feature.
- The **least-fill** option support with an inter-area LSP is introduced with the automatic ABR selection feature. When this option is enabled, CSPF applies the least-fill criterion to select the path segment to the exit ABR node in the local area.
- 1The PLR node must indicate to CSPF that a request to one-to-one detour LSP path must remain within the local area. If the destination for the detour, which is the same as that of the LSP, is outside of the area, CSPF must return no path.

- The **propagate-admin-group** option under the LSP will still need to be enabled on the inter-area LSP if the user wants to have admin-groups propagated across the areas.
- With the automatic ABR selection feature, timer based re-signal of the inter-area LSP path is supported and will re-signal the path if the cost of the path segment to the local exit ABR changed. The cost shown for the inter-area LSP at ingress LER is the cost of the path segments to the ABR node.

2.6.13.2 Inter-Area LSP support of OSPF Virtual Links

The OSPF virtual link extends area 0 for a router that is not connected to area 0. As a result, it makes all prefixes in area 0 reachable via an intra-area path but in reality, they are not since the path crosses the transit area through which the virtual link is set up to reach the area 0 remote nodes.

The TE database in a router learns all of the remote TE links in area 0 from the ABR connected to the transit area, but an intra-area LSP path using these TE links cannot be signaled within area 0 since none of these links is directly connected to this node.

This inter-area LSP feature can identify when the destination of an LSP is reachable via a virtual link. In that case, CSPF will automatically compute and signal an inter-area LSP via the ABR nodes that is connected to the transit area.

However, when the ingress LER for the LSP is the ABR connected to the transit area and the destination of the LSP is the address corresponding to another ABR router-id in that same transit area, CSPF will compute and signal an intra-area LSP using the transit area TE links, even when the destination router-id is only part of area 0.

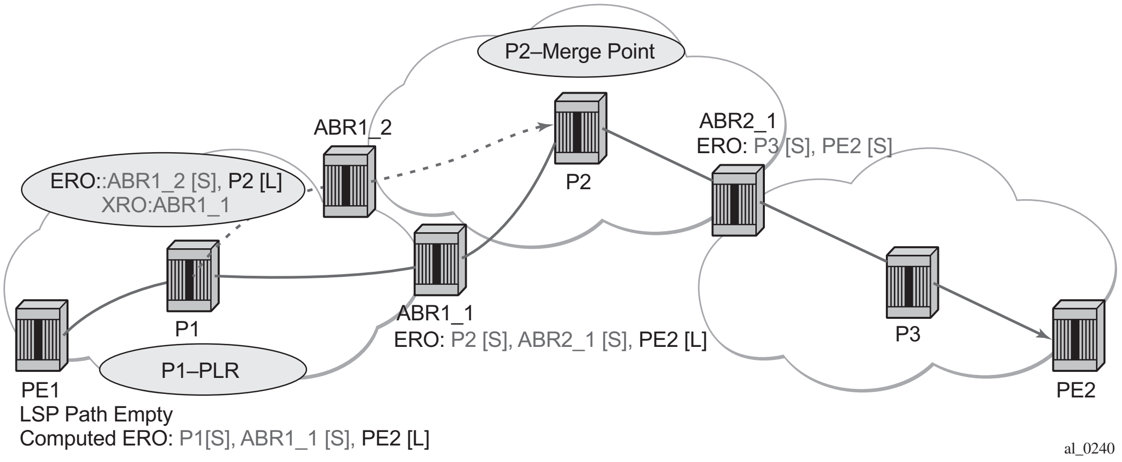
2.6.13.3 Area Border Node FRR Protection for Inter-Area LSP

For protection of the area border router, the upstream node of the area border router acts as a point-of-local-repair (PLR), and the next-hop node to the protected domain border router is the merge-point (MP). Both manual and dynamic bypass are available to protect area border node.

Manual bypass protection works only when a proper completely strict path is provisioned that avoids the area border node.

Dynamic bypass protection provides for the automatic computation, signaling, and association with the primary path of an inter-area P2P LSP to provide ABR node protection. Figure 35 illustrates the role of each node in the ABR node protection using a dynamic bypass LSP.

Figure 35 **ABR Node Protection Using Dynamic Bypass LSP**



In order for a PLR node within the local area of the ingress LER to provide ABR node protection, it must dynamically signal a bypass LSP and associate it with the primary path of the inter-area LSP using the following new procedures:

- The PLR node must inspect the node-id RRO of the LSP primary path to determine the address of the node immediately downstream of the ABR in the other area.
- The PLR signals an inter-area bypass LSP with a destination address set to the address downstream of the ABR node and with the XRO set to exclude the node-id of the protected ABR node.
- The request to CSPF is for a path to the merge-point (that is the next-next-hop in the RRO received in the RESV for the primary path) along with the constraint to exclude the protected ABR node and the include/exclude admin-groups of the primary path. If CSPF returns a path that can only go to an intermediate hop, then the PLR node signals the dynamic bypass and will automatically include the XRO with the address of the protected ABR node and propagate the admin-group constraints of the primary path into the Session Attribute object of the bypass LSP. Otherwise, the PLR signals the dynamic bypass directly to the merge-point node with no XRO object in the Path message.
- If a node-protect dynamic bypass cannot be found or signaled, the PLR node attempts a link-protect dynamic bypass LSP. As in existing implementation of dynamic bypass within the same area, the PLR attempts in the background to signal a node-protect bypass at the receipt of every third Resv refresh message for the primary path.
- Refresh reduction over dynamic bypass will only work if the node-id RRO also contains the interface address. Otherwise the neighbor will not be created once the bypass is activated by the PLR node. The Path state will then time out after three refreshes following the activation of the bypass backup LSP.

Note that a one-to-one detour backup LSP cannot be used at the PLR for the protection of the ABR node. As a result, a PLR node will not signal a one-to-one detour LSP for ABR protection. In addition, an ABR node will reject a Path message, received from a third party implementation, with a detour object and with the ERO having the next-hop loose. This is performed regardless if the **cspf-on-loose-hop** option is enabled or not on the node. In other words, the router as a transit ABR for the detour path will reject the signaling of an inter-area detour backup LSP.

2.6.14 Timer-based Reversion for RSVP-TE LSPs

The following secondary to primary path reversion is supported for RSVP-TE LSPs:

- Configurable timer-based reversion for primary LSP path
- Manual reversion from secondary to primary path

Normally, an RSVP-TE LSP automatically switches back from using a secondary path to the primary path as soon as the primary path recovers. In some deployments, it is useful to delay reversion or allow manual reversion, rather than allowing an LSP to revert to the primary path as soon as it is available. This feature provides a method to manage fail-overs in the network.

If manual reversion is used, a fall-back timer-based mechanism is required in case a human operator fails to execute the switch back to the primary path. This function is also useful to stagger reversion for large numbers of LSPs.

A reversion timer for an LSP is configured using the CLI as follows:

```
config
   router
      [no] mpls
             lsp
               [no] revert-timer <timer-value>
```

When configured, the revert timer is started as soon as a primary path recovers. The LSP does not revert from the currently used secondary path to the primary path until the timer expires. When configured, the revert-timer is used instead of the existing hold timer.

The timer value can be configured in one minute increments, up to 4320 minutes (72 hours). Once a timer has started, it can be modified using this command. If a new value is entered, then the current timer is canceled (without reverting the LSP) and then restarted using the new value. The revert timer should always be configured to a higher value than the hold timer. This prevents the router from reverting to the primary path and sending traffic before the downstream LSRs have programmed their data path.

The **no** form of the command cancels any currently outstanding revert timer and causes the LSP to revert to the primary path if it is up.

If the LSP secondary path fails while the revert timer is still running, the system cancels the revert- timer and the LSP will revert to the primary path immediately. A user can manually force an LSP to revert to the primary path while the revert-timer is still running, using the following tools command:

tools>perform>router>mpls revert lsp *lsp-name*

This command forces the early expiry of the revert timer for the LSP. The primary path must be up in order for this command to work.

2.6.15 LSP Tagging and Auto-Bind Using Tag Information

RSVP and SR-TE LSPs can be configured with an administrative tag.

The primary application of LSP tagging is to enable the system to resolve to specific transport tunnels (or groups of eligible transport tunnels) for BGP routes for applications such as BGP labeled unicast, VPRN, or EVPN. Additionally, LSP tagging specifies a finer level of granularity on the next-hop or the far-end prefix associated with a BGP labeled unicast route or unlabeled BGP route shortcut tunnels.

LSP tagging is supported using the following capabilities in SR OS.

- The ability to associate a color with an exported BGP route. This is signaled using the BGP Color Extended Community described in Section 4.3 of *draft-ietf-idr-tunnel-encaps-03*. This provides additional context associated with a route that an upstream router can use to help select a distinct transport for traffic associated with that route.
- The ability to define a set of administrative tags on a node for locally-coloring imported routes and consequent use in transport tunnel selection. Up to 256 discrete tag values are supported.
- The ability to configure a set of administrative tags on an RSVP or SR-TE LSP. This tag is used by applications to refer to the LSP (or set of LSPs with the same tag) for the purposes of transport tunnel selection. Up to four tags are supported per LSP.
- The ability to apply one or more administrative tags to include or exclude as an action to a matching route in a BGP route policy. Different admin-tag values can be applied to different VPRN routes, such that different VPRNs can ultimately share the same set of tunnels by having the same admin-tags associated to their VPN routes via matching on RT extended community values.

- The ability to match an administrative tag in a route policy for the following service types to the list of available RSVP or SR-TE tunnels (potentially filtered by the resolution filter):
    - BGP labeled unicast and BGP shortcuts
    - VPRN with auto-bind-tunnel
    - EVPN with auto-bind-tunnel

The following provides an overview of how the feature is intended to operate:

1. Configure a nodal database of admin-tags. Each tag is automatically assigned an internal color. The nodal admin tag database is configured under **config**>**router**>**admin-tags** in the CLI.
2. Optionally, configure export route policies associating routes with a color extended community. The color extended community allows for a color to be advertised along with specific routes, intended to indicate some property of a transport that a route can be associated with.
3. Configure a named **route-admin-tag-policy** containing a list of admin-tags to include or exclude. The **route-admin-tag-policy** is configured under **config**>**router**>**admin-tags** in the CLI. Up to eight include and exclude statements are supported per policy.
4. Configure a named **route-admin-tag-policy** as an action against matching routes in a route policy. An internal route color is applied to matching routes. Examples of a match are on a BGP next-hop or an extended community; for example, the color extended community specified in Section 4.3 of *draft-ietf-idr-tunnel-encaps-03*. That is, if that policy is later used as an import policy by a service, routes received from, for example, a matching BGP next hop or color-extended community in the policy will be given the associated internal color.
5. Configure admin-tags on RSVP or SR-TE LSPs so that different groups of LSPs can be treated differently by applications that intend to use them. More than one admin-tag can be configured against a specified LSP. Admin-tags are configured using the **admin-tag** command under **config**>**router**>**mpls**>**lsp** in the CLI.
6. Apply a route policy to a service or other object as an import policy. The system then matches the internal color policy of a route against corresponding LSP internal colors in the tunnel table. That set of LSPs can subsequently be limited by a resolution filter. For BGP-LU and BGP shortcut routes, the resolution filter can optionally be restricted to only those LSPs matching the pattern of admin-tags in the **route-admin-tag-policy** (otherwise the resolution fails) using the **enforce-strict-tunnel-tagging** option. If **enforce-strict-tunnel-tagging** is not

specified, then the router falls back to untagged LSPs. The tunnels that VPRN and EVPN services can auto-bind to can also be restricted using the **enforce-strict-tunnel-tagging** option in the **auto-bind-tunnel** configuration for the service. The following subsections provide more details about how the matching algorithm works.

2.6.15.1 Internal Route Color to LSP Color Matching Algorithm

This section describes how the matrix of **include** or **exclude** colors in a **route-admin-tag-policy** *policy-name*, which is assigned to a route, are matched against LSP internal colors. This is a generic algorithm. The following sections provide further details of how this applies to specific use cases.

Internal color matching occurs before any resolution filter is applied.

The following selection process assumes the system starts with a set of eligible RSVP and SR-TE LSPs to the appropriate BGP next hop.

1. Prune the following RSVP and SR-TE LSPs from the eligible set:
   - uncolored LSPs
   - LSPs where none of the internal colors match any "include" color for the route
   - LSPs where any of the internal colors match any "exclude" color for the route
2. If none of the LSPs match, then the default behavior is that the route does not resolve. Depending on the context, configure a fall-back method, as described in LSP Admin Tag use in Tunnel Selection for VPRN and E-VPN Auto-Bind.
3. If a route does not have an admin-tag policy, it is assumed that the operator does not wish to express a preference for the LSP to use. Therefore, routes with no admin-tag policy can still resolve to any tagged or untagged LSP.

This selection process results in a set of one or more ECMP LSPs, which may be further reduced by a resolution filter.

2.6.15.2 LSP Admin Tag use in Tunnel Selection for VPRN and E-VPN Auto-Bind

For VPRN, EVPN-VPLS, and EVPN-VPWS, routes may be imported via peer route import policies that contain route admin-tag policies or via VRF import for VPRN and VSI import for E-VPN VPLS used for auto-bind-tunnel.

VRF import and VSI import policies take precedence over the peer route import policy.

For policies that contain route admin-tag policies, the set of available RSVP and SR-TE LSPs in TTM are first pruned as described in Internal Route Color to LSP Color Matching Algorithm. This set may then be further reduced by a resolution filter. If **weighted-ecmp** is configured, then this is applied across the resulting set.

Routes with no admin-tag, or a tag that is not explicitly excluded by the route admin tag policy, can still resolve to any tagged or untagged LSP but matching tagged LSPs are used in preference to any other. It is possible that following the resolution filter no eligible RSVP or SR-TE LSP exists. By default, the system will fall back to regular auto-bind behavior using LDP, SR-ISIS, SR-OSPF, or any other lower priority configured tunnel type, otherwise the resolution will fail. That is, matching admin-tagged RSVP or SR-TE LSPs are used in preference to other LSP types, whether tagged or untagged. However, it is possible on a per-service basis to enforce that only specific tagged tunnels should be considered, otherwise resolution will fail, using the **enforce-strict-tunnel-tagging** command in the **auto-bind-tunnel** context.

For E-VPN VPWS, VSI import is not supported. Therefore, admin-tag policies can only be applied via a peer route import policy based on a match on the route target for the BGP peer for the VPWS.

2.6.15.3 LSP Admin Tag Use for BGP Next Hop or BGP Prefix for Labeled and Unlabeled Unicast Routes

A specific LSP can be selected as transport to a specified BGP next hop for BGP labeled unicast and unlabeled BGP routes tunneled over RSVP and SR-TE LSPs.

Routes are imported via import route policies. Named routing policies may contain route admin-tag policies. For route import policies that contain route admin-tag policies, the set of available RSVP and SR-TE LSPs in TTM are first pruned as described in Internal Route Color to LSP Color Matching Algorithm.

This set may then be further reduced by a resolution filter.

If **weighted-ecmp** is configured, then this is applied across the resulting set.

Routes with no admin-tag can still resolve to any tagged or untagged LSP. It is possible that, following the resolution filter, no eligible RSVP or SR-TE LSP exists. By default, the system falls back to using LDP, SR-ISIS, SR-OSPF, or any other lower-priority tunnel type; otherwise the resolution fails. That is, matching admin-tagged RSVP or SR-TE LSPs are preferred to other LSP types. On a per-address family basis, the **enforce-strict-tunnel-tagging** command in the **next-hop-resolution** filter for BGP labeled routes or shortcut tunnels can be used to enforce that only tagged tunnels are considered; otherwise, resolution fails.

2.6.16 LSP Self-Ping

LSP Self-ping is specified in RFC 7746, *Label Switched Path (LSP) Self-Ping*. LSP Self-ping provides a lightweight, periodic connectivity check by the head-end LER of an LSP with no session state in the tail-end LER. LSP Self-ping checks that an LSP data path has been programmed following the receipt of the RESV message for the path. LSP Self-ping defines a new OAM packet with a locally unique session ID. The IP source address of this packet is set to the address of the egress LER, and the destination address is set to that of the ingress LER, such that when the packet exits the egress LER the packet is simply forwarded back to the ingress LER. LSP Self-ping is a distinct OAM mechanism from LSP ping, despite the similar name.

SR OS supports LSP Self-ping for point-to-point RSVP-TE LSPs and point-to-point RSVP auto-LSPs, including PCC-initiated and PCC-controlled LSPs, and PCC-initiated and PCE-controlled LSPs.

An SR OS router can use LSP Self-ping to test that the data path of an LSP has been fully programmed along its length before moving traffic onto it. When enabled, LSP Self-ping packets are periodically sent on a candidate path that the router intends to switch to, for example, during primary or secondary switching (with FRR on the primary) or MBB of a path, following the receipt of the RESV message, until a reply is received from the far end. When a reply is received, the system determines that the data path of the LSP must have been programmed. LSP Self-ping is used instead of the LSP hold timer (**config**>**router**>**mpls**>**hold-timer**). This is particularly useful in multi-vendor networks where certain nodes may take unexpectedly long times to program their data path.

LSP BFD is not supported if LSP Self-ping is enabled. The router ignores the LSP Self Ping configuration if **configure**>**router**>**mpls**>**lsp**>**bfd**>**failure-action failover-or-down** is configured for an LSP.

LSP Self-ping is configured under the MPLS context using the **lsp-self-ping** command.

```
configure
```

```
router
   mpls
      [no] lsp-self-ping
        interval <seconds>
        timeout <seconds>
        timeout-action {retry | switch}
        rsvp-te {enable | disable}
```

LSP Self-ping is enabled for all RSVP-TE LSPs using the **rsvp-te enable** command. However, it is possible to enable or disable LSP Self-ping for a specific LSP or LSP template regardless of the setting at the MPLS level.

The **interval** command sets the interval, in seconds, that periodic LSP Self-ping packets are sent. The **timeout** command configures a timer that is started when the first LSP Self-ping packet for a given event is sent on an LSP path. The **timeout-action** specifies what action to take if no LSP Self-ping reply is received before the timer expires. If **timeout-action** is set to **retry**, then the router tries to signal a new path and the process repeats (see Detailed Behavior of LSP Self-Ping for more information). If **timeout-action** is set to **switch**, then the router uses the new path regardless and stops the LSP Self-ping cycle.

LSP Self-ping can also be enabled or disabled for a given LSP or LSP template:

```
configure router mpls
   lsp
     lsp-self-ping {enable | disable | inherit}

configure router mpls
   lsp-template
      lsp-self-ping {enable | disable | inherit}
```

By default, LSPs and LSP templates inherit the configuration at the MPLS level. However, LSP Self-ping may be enabled for a specific LSP or LSP template using the **lsp-self-ping enable** command. LSP Self-ping may be explicitly disabled for a given LSP or LSP template, even if enabled at the MPLS level, using the **lsp-self-ping disable** command.

2.6.16.1 Detailed Behavior of LSP Self-Ping

When LSP Self-ping is enabled, destination UDP port 8503 is opened and a unique session ID is allocated for each RSVP LSP path. When an RESV message is received following a resignaling event, LSP Self-ping packets are sent at configurable periodic intervals until a reply is received from the far end for that session ID.

LSP Self-ping applies in cases where the active path is changed, while the previous active path remains up, whether it is FRR/MBB or pre-empted. These cases are as follows:

- Primary in degraded state -> standby or secondary path
- Standby or secondary path -> primary path (reversion)
- Standby or secondary path -> another standby or secondary path (**tools**>**perform**>**router**>**mpls**>**switch-path** command or path preference change)
- Degraded standby/secondary path -> degraded primary path (degraded primary is preferred to degraded standby/secondary path)
- MBB on active path

A path can go to a degraded state either due to FRR active (only on the primary path), soft pre-emption, or LSP BFD down (when the failure action is failover).

The system does not activate a candidate path until the first LSP Self-ping reply is received, subject to the timeout. The LSP Self-ping timer is started when the RESV message is received. The system will then periodically send LSP Self-ping packets until the timer expires or the first LSP Self-ping reply is received, whichever comes first. If the timeout expires before an LSP Self-ping reply has been received and the **timeout-action** is set to **retry**, then the system tears down the candidate path (in the case of switching between paths) and go back to CSPF for a new path. The system will then start the LSP Self-ping cycle again after a new path is obtained. In the case of switching between paths, the system retries immediately and increments the retry counter. In the case of MBB, the system retries immediately, but will not increment the retry counter, which has the effect of continuously repeating the retry/LSP Self-ping cycle until a new path is successfully established.

Note: If the configured timeout value is changed for an LSP with an in-progress LSP Self-ping session, the previous timer will complete, and the new value is not used until the next lsp-self-ping session.

If no timeout is configured, then the default value is used.

2.6.16.2 Considerations for Scaled Scenarios

The router can send LSP Self-ping packets at a combined rate across all sessions of 125 packets per second. This means that it takes 10 seconds to detect that the data plane is forwarding for 1250 LSPs. If the number of currently in-progress LSP Self-ping sessions reaches 125 PPS with no response, then the system continues with these LSP Self-ping sessions until the timeout is reached and is not able to test additional LSP paths. In scaled scenarios, it is recommended that the lsp-self-ping interval and timeout values be configured so that LSP Self-ping sessions are completed (either successfully or through timing out) so that all required LSP paths are tested within an acceptable timeframe. A count of the number of LSP Self-ping and OAM resource exhaustion timeouts is shown in the output of the **show>router>mpls>lsp detail** and **show>router>mpls>lsp-self-ping** commands.

2.6.17 Accounting for Dark Bandwidth

In traffic engineered networks, IGP-TE advertisements are used to distribute bandwidth availability on each link. This bandwidth availability information only accounts for RSVP-TE LSP set-ups and tear-downs. However, network deployments often have labeled traffic (other than RSVP-TE LSP) flowing on the same links as these RSVP-TE LSPs, in particular when MPLS Segment Routing (MPLS-SR) is deployed. The bandwidth consumed by this labeled traffic is often referred to as dark bandwidth.

The bandwidth consumed by, for example, MPLS-SR traffic is not accounted for in IGP-TE advertisements. This unaccounted-for traffic may result in suboptimal constrained routing decisions or contention for the access to the bandwidth resource. SR OS enables accounting for dark bandwidth in IGP-TE advertisement and provides the means to control the behavior of this accounting.

To configure dark bandwidth accounting:

1. Enable collection of traffic statistics for dark bandwidth, using the command **configure>router>mpls>aux-stats sr**

Note: Only one keyword parameter is available (**sr**) for this command, so only MPLS-SR is considered as contributing to dark bandwidth.

2. Enable dark bandwidth accounting on each SE, using the command **configure>router>rsvp>dbw-accounting**

Note: After dark bandwidth has been enabled, auxiliary statistics collection cannot be disabled. Dark bandwidth accounting must be disabled (**no dbw-accounting**) before auxiliary statistics collection can be disabled.

3. Configure the dark bandwidth accounting parameters to control the behavior of the system.

When dark bandwidth accounting is enabled, the system samples dark bandwidth at the end of every sample interval and computes an average after *sample-multiplier* samples. The system applies a multiplier (*dbw-multiplier*) to the computed average dark bandwidth and then determines whether an IGP-TE update is required based on whether one of the thresholds (*up-threshold* or *down-threshold*) has been crossed. If an IGP-TE advertisement is required, the bandwidth information is updated, considering that dark bandwidth has the highest priority among the eight available priorities. These thresholds represent a change of Maximum Reservable Bandwidth (OSPF) or Maximum Reservable Link Bandwidth (IS-IS) compared to the previously advertised bandwidth. These parameters are generally global parameters, but it is possible to override the global value of some parameters on a per-interface basis.

The **show>router>rsvp>status** command allows the user to view, on a global or per-interface basis, key values associated with the dark bandwidth accounting process.

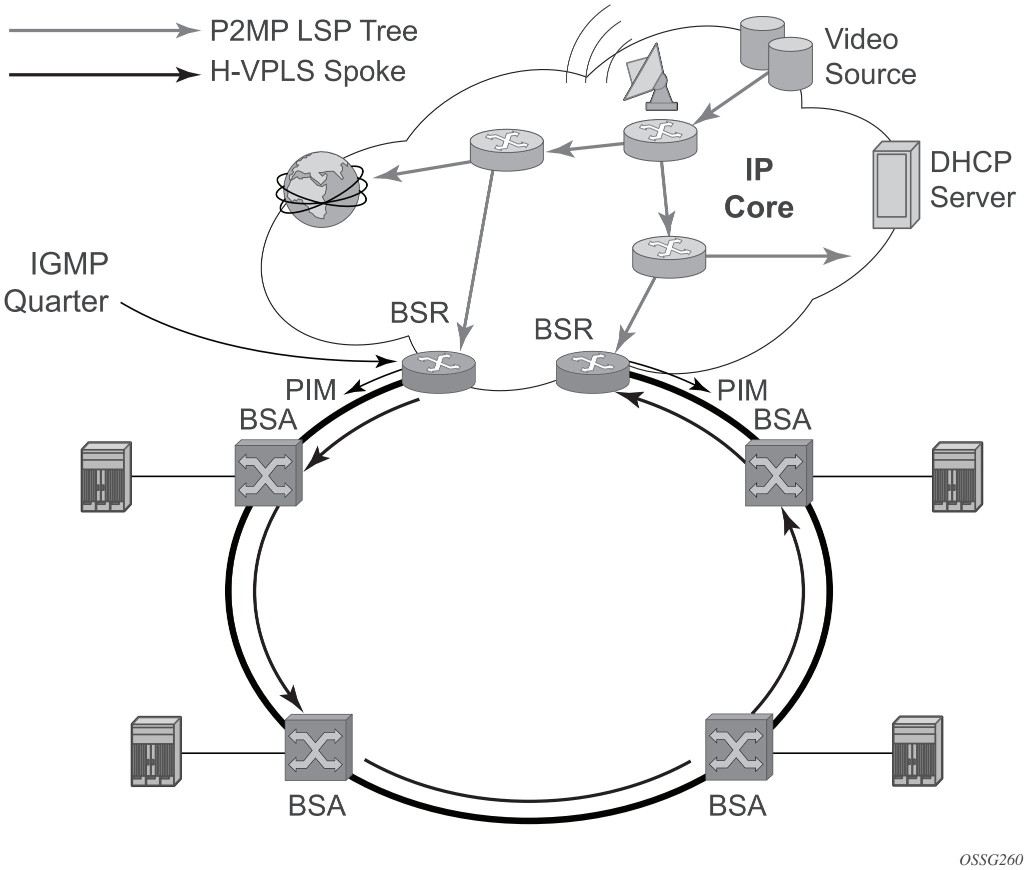
2.7 Point-to-Multipoint (P2MP) RSVP LSP

Point-to-multipoint (P2MP) RSVP LSP allows the source of multicast traffic to forward packets to one or many multicast receivers over a network without requiring a multicast protocol, such as PIM, to be configured in the network core routers. A P2MP LSP tree is established in the control plane which path consists of a head-end node, one or many branch nodes, and the leaf nodes. 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.

2.7.1 Application in Video Broadcast

Figure 36 illustrates the use of the 7450 ESS, 7750 SR, and 7950 XRS in triple play application (TPSDA). The Broadband Service Router (BSR) is a 7750 SR and the Broadband Service Aggregator (BSA) is the 7450 ESS.

Figure 36   Application of P2MP LSP in Video Broadcast



A PIM-free core network can be achieved by deploying P2MP LSPs using other core routers. The router can act as the ingress LER receiving the multicast packets from the multicast source and forwarding them over the P2MP LSP.

A router can act as a leaf for the P2MP LSP tree initiated from the head-end router co-located with the video source. The router can also act as a branch node serving other leaf nodes and supports the replication of multicast packets over P2MP LSPs.

2.7.2 P2MP LSP Data Plane

A P2MP LSP is a unidirectional label switched path (LSP) which inserts packets at the root (ingress LER) and forwards the exact same replication of the packet to one or more leaf nodes (egress LER). The packet can be replicated at the root of P2MP LSP tree and/or at a transit LSR which acts as a branch node for the P2MP LSP tree.

Note that the data link layer code-point, for example Ethertype when Ethernet is the network port, continues to use the unicast codepoint defined in RFC 3032, *MPLS Label Stack Encoding*, and which is used on P2P LSP. This change is specified in draft-ietf-mpls-multicast-encaps, *MPLS Multicast Encapsulations*.

When a router sends a packet over a P2MP LSP which egresses on an Ethernet-based network interface, the Ethernet frame uses a MAC unicast destination address when sending the packet over the primary P2MP LSP instance or over a P2P bypass LSP). Note that a MAC multicast destination address is also allowed in the *draft-ietf-mpls-multicast-encaps*. Therefore, at the ingress network interface on an Ethernet port, the router can accept both types of Ethernet destination addresses.

2.7.2.1 Procedures at Ingress LER Node

The following procedures occur at the root of the P2MP LSP (head-end or ingress LER node):

1. First, the P2MP LSP state is established via the control plane. Each leaf of the P2MP LSP will have a next-hop label forwarding entry (NHLFE) configured in the forwarding plane for each outgoing interface.
2. The user maps a specific multicast destination group address to the P2MP LSP in the base router instance by configuring a static multicast group under a tunnel interface representing the P2MP LSP.
3. An FTN entry is programmed at the ingress of the head-end node that maps the FEC of a received user IP multicast packet to a list of outgoing interfaces (OIF) and corresponding NHLFEs.
4. The head-end node replicates the received IP multicast packet to each NHLFE. Replication is performed at ingress toward the fabric and/or at egress forwarding engine depending on the location of the OIF.
5. At ingress, the head-end node performs a PUSH operation on each of the replicated packets.

2.7.2.2 Procedures at LSR Node

The following procedures occur at an LSR node that is not a branch node:

- The LSR performs a label swapping operation on a leaf of the P2MP LSP. This is a conventional operation of an LSR in a P2P LSP. An ILM entry is programmed at the ingress of the LSR to map an incoming label to a NHLFE.

The following is an exception handling procedure for control packets received on an ILM in an LSR.

- Packets that arrive with the TTL in the outer label expiring are sent to the CPM for further processing and are not forwarded to the egress NHLFE.

2.7.2.3 Procedures at Branch LSR Node

The following procedures occur at an LSR node that is a branch node:

- The LSR performs a replication and a label swapping for each leaf of the P2MP LSP. An ILM entry is programmed at the ingress of the LSR to map an incoming label to a list of OIF and corresponding NHLFEs.
- There is a limit of 127 OIF/NHLFEs per ILM entry.

The following is an exception handling procedure for control packets received on an ILM in a branch LSR:

- Packets that arrive with the TTL in the outer label expiring are sent to the CPM for further processing and not copied to the LSP branches.

2.7.2.4 Procedures at Egress LER Node

The following procedures occur at the leaf node of the P2MP LSP (egress LER):

- The egress LER performs a pop operation. An ILM entry is programmed at the ingress of the egress LER to map an incoming label to a list of next-hop/OIF.

The following is an exception handling procedure for control packets received on an ILM in an egress LER.

- The packet is sent to the CPM for further processing if there is any of the IP header exception handling conditions set after the label is popped: 127/8 destination address, router alert option set, or any other options set.

2.7.2.5 Procedures at BUD LSR Node

The following are procedures at an LSR node which is both a branch node and an egress leaf node (bud node):

- The bud LSR performs a pop operation on one or many replications of the received packet and a swap operation of the remaining replications. An ILM entry is programmed at ingress of the LSR to map the incoming label to list of NHLFE/OIF and next-hop/OIF.

  Note however, the exact same packets are replicated to an LSP leaf and to a local interface.

The following are the exception handling procedures for control packets received on an ILM in a bud LSR:

- Packets which arrive with the TTL in the outer label expiring are sent to the CPM and are not copied to the LSP branches.
- Packets whose TTL does not expire are copied to all branches of the LSP. The local copy of the packet is sent to the CPM for further processing if there is any of the IP header exception handling conditions set after the label is popped: 127/8 destination address, router alert option set, or any other options set.

2.7.3 Ingress Path Management for P2MP LSP Packets

The SR OS provides the ingress multicast path management (IMPM) capability that allows users to manage the way IP multicast streams are forwarded over the router's fabric and to maximize the use of the fabric multicast path capacity.

IMPM consists of two components, a bandwidth policy and a multicast information policy. The bandwidth policy configures the parameters of the multicast paths to the fabric. This includes the multicast queue parameters of each path. The multicast information policy configures the bandwidth and preference parameters of individual multicast flows corresponding to a channel, for example, a <*,G> or a <S,G>, or a bundle of channels.

By default, the XCM (on the 7950 XRS) and the IOM/IMM (on the 7750 SR and 7450 ESS) ingress data paths provides two multicast paths through the fabric referred to as high-priority path and low-priority path respectively. When a multicast packet is received on an ingress network or access interface or on a VPLS SAP, the packet's classification will determine its forwarding class and priority or profile as per the ingress QoS policy. This then determines which of the SAP or interface multicast queues it must be stored in. By default SAP and interface expedited forwarding class queues forward over the high-priority multicast path and the non-expedited forwarding class queues forward over the low-priority multicast path.

When IMPM on the ingress FP is enabled on the 7950 XRS, 7750 SR, or 7450 ESS, one or more multicast paths are enabled depending on the hardware in use. In addition, for all routers, multicast flows managed by IMPM are stored in a separate shared multicast queue for each multicast path. These queues are configured in the bandwidth policy.

IMPM maps a packet to one of the paths dynamically based on monitoring the bandwidth usage of each packet flow matching a <*,G> or <S,G> record. The multicast bandwidth manager also assigns multicast flows to a primary path based on the flow preference until the rate limits of each path is reached. At that point in time, a multicast flow is mapped to the secondary flow. If a path congests, the bandwidth manager will remove and black-hole lower preference flows to guarantee bandwidth to higher preference flows. The preference of a multicast flow is configured in the multicast info policy.

A packet received on a P2MP LSP ILM is managed by IMPM when IMPM is enabled on the ingress XMA or the ingress FP and the packet matches a specific multicast record. When IMPM is enabled but the packet does not match a multicast record, or when IMPM is disabled, a packet received on a P2MP LSP ILM is mapped to a multicast path.

2.7.3.1 Ingress P2MP Path Management on XCM/IOM/IMMs

On an ingress XCM or IOM/IMM, there are multiple multicast paths available to forward multicast packets, depending on the hardware being used. Each path has a set of multicast queues and associated with it. Two paths are enabled by default, a primary path and a secondary path, and represent the high-priority and low-priority paths respectively. Each VPLS SAP, access interface, and network interface will have a set of per forwarding class multicast and/or broadcast queues which are defined in the ingress QoS policy associated with them. The expedited queues are attached to the primary path while the non-expedited queues are attached to secondary path.

When IMPM is enabled and/or when a P2MP LSP ILM exists on the ingress XCM or IOM/IMM, the remaining multicast paths are also enabled. 16 multicast paths are supported by default with 28 on 7950 XRS systems and 7750 SR-12e systems, with the latter having the **tools** perform **system set-fabric-speed fabric-speed-b**. One path remains as a secondary path and the rest are primary paths.

A separate pair of shared multicast queues is created on each of the primary paths, one for IMPM managed packets and one for P2MP LSP packets not managed by IMPM. The secondary path does not forward IMPM managed packets or P2MP LSP packets. These queues have a default rate (PIR=CIR) and CBS/MBS/low-drop-tail thresholds, but these can be changed under the bandwidth policy.

A VPLS snooped packet, a PIM routed packet, or a P2MP LSP packet is managed by IMPM if it matches a <*,G> or a <S,G> multicast record in the ingress forwarding table and IMPM is enabled on the ingress XMA or the FP where the packet is received. The user enables IMPM on the ingress XMA data path or the FP data path using the **config>card>fp>ingress>mcast-path-management** command.

A packet received on an IP interface and to be forwarded to a P2MP LSP NHLFE or a packet received on a P2MP LSP ILM is not managed by IMPM when IMPM is disabled on the ingress XMA or the FP where the packet is received or when IMPM is enabled but the packet does not match any multicast record. A P2MP LSP packet duplicated at a branch LSR node is an example of a packet not managed by IMPM even when IMPM is enabled on the ingress XMA or the FP where the P2MP LSP ILM exists. A packet forwarded over a P2MP LSP at an ingress LER and which matches a <*,G> or a <S,G> is an example of a packet which is not managed by IMPM if IMPM is disabled on the ingress XMA or the FP where the packet is received.

When a P2MP LSP packet is not managed by IMPM, it is stored in the unmanaged P2MP shared queue of one of the primary multicast paths.

By default, non-managed P2MP LSP traffic is distributed across the IMPM primary paths using hash mechanisms. This can be optimized by enabling IMPM on any forwarding complex, which allows the system to redistribute this traffic on all forwarding complexes across the IMPM paths to achieve a more even capacity distribution. Be aware that enabling IMPM will cause routed and VPLS (IGMP and PIM) snooped IP multicast groups to be managed by IMPM.

The above ingress data path procedures apply to packets of a P2MP LSP at ingress LER, LSR, branch LSR, bud LSR, and egress LER. Note that in the presence of both IMPM managed traffic and unmanaged P2MP LSP traffic on the same ingress forwarding plane, the user must account for the presence of the unmanaged traffic on the same path when setting the rate limit for an IMPM path in the bandwidth policy.

2.7.4 RSVP Control Plane in a P2MP LSP

P2MP RSVP LSP is specified in RFC 4875, *Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs).*

A P2MP LSP is modeled as a set of source-to-leaf (S2L) sub-LSPs. The source or root, for example the head-end node, triggers signaling using one or multiple path messages. A path message can contain the signaling information for one or more S2L sub-LSPs. The leaf sub-LSP paths are merged at branching points.

A P2MP LSP is identified by the combination of <P2MP ID, tunnel ID, extended tunnel ID> part of the P2MP session object, and <tunnel sender address, LSP ID> fields in the P2MP sender_template object.

A specific sub-LSP is identified by the <S2L sub-LSP destination address> part of the S2L_SUB_LSP object and an ERO and secondary ERO (SERO) objects.

The following are characteristics of this feature:

- Supports the de-aggregated method for signaling the P2MP RSVP LSP. Each root to leaf is modeled as a P2P LSP in the RSVP control plane. Only data plane merges the paths of the packets.
- Each S2L sub-LSP is signaled in a separate path message. Each leaf node responds with its own resv message. A branch LSR node will forward the path message of each S2L sub-LSP to the downstream LSR without replicating it. It will also forward the resv message of each S2L sub-LSP to the upstream LSR without merging it with the resv messages of other S2L sub-LSPs of the same P2MP LSP. The same is done for subsequent refreshes of the path and resv states.
- The node will drop aggregated RSVP messages on the receive side if originated by another vendor’s implementation.
- The user configures a P2MP LSP by specifying the optional create-time parameter **p2mp-lsp** following the LSP name. Next, the user creates a primary P2MP instance using the keyword **primary-p2mp-instance**. Then a path name of each S2L sub-LSP must added to the P2MP instance using the keyword **s2l-path**. The paths can be empty paths or can specify a list of explicit hops. The path name must exist and must have been defined in the **config>router>mpls>path** context.
- The same path name can be re-used by more than one S2L of the primary P2MP instance. However the **to** keyword must have a unique argument per S2L as it corresponds to the address of the egress LER node.
- The user can configure a secondary instance of the P2MP LSP to backup the primary one. In this case, the user enters the name of the secondary P2MP LSP instance under the same LSP name. One or more secondary instances can be created. The trigger for the head-end node to switch the path of the LSP from the primary P2MP instance to the secondary P2MP instance is to be determined. This could be based on the number of leaf LSPs which went down at any given time.
- The following parameters can be used with a P2MP LSP: adaptive, cspf, exclude, fast-reroute, from, hop-limit, include, metric, retry-limit, retry-timer, resignal-timer.
- The following parameters cannot be used with a P2MP LSP: adspec, primary, secondary, to.

- The node ingress LER will not inset an adspec object in the path message of an S2L sub-LSP. If received in the resv message, it is dropped. The operational MTU of an S2L path is derived from the MTU of the outgoing interface of that S2L path.
- The **to** parameter is not available at the LSP level but at the path level of each S2L sub-LSP of the primary or secondary instance of this P2MP LSP.
- The hold-timer configured in the **config>router>mpls>hold-timer** context applies when signaling or re-signaling an individual S2L sub-LSP path. It does not apply when the entire tree is signaled or re-signaled.
- The head-end node can add and/or remove a S2L sub-LSP of a specific leaf node without impacting forwarding over the already established S2L sub-LSPs of this P2MP LSP and without re-signaling them.
- The head-end node performs a make-before break (MBB) on an individual S2L path of a primary P2MP instance whenever it applies the FRR global revertive procedures to this path. If CSPF finds a new path, RSVP signals this S2L path with the same LSP-ID as the existing path.
- All other configuration changes, such as adaptive/no-adaptive (when an MBB is in progress), metric-type te, no-frr, path-computation-method/no path-computation-method, result in the tear-down and re-try of all affected S2L paths.
- MPLS requests CSPF to re-compute the whole set of S2L paths of a given active P2MP instance each time the P2MP re-signal timer expires. The P2MP re-signal timer is configured separately from the P2P LSP. MPLS performs a global 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. This is regardless of the cost of the new S2L path.
- MPLS will request CSPF to re-compute the whole set of S2L paths of a given active P2MP instance each time the user performs a manual re-signal of the P2MP instance. MPLS then always performs a global 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. This is regardless of the cost of the new S2L path. The user executes a manual re-signal of the P2MP LSP instance using the command: **tools>perform>router>mpls>resignal p2mp-lsp** *lsp-name* **p2mp-instance** *instance-name*.
- When performing global MBB, MPLS runs a separate MBB on each S2L in the P2MP LSP instance. If an S2L MBB does not succeed the first time, MPLS will re-try the S2L using the re-try timer and re-try count values inherited from P2MP LSP configuration. However, there is a global MBB timer set to 600 seconds and which is not configurable. If the global MBB succeeds, for example, all S2L MBBs have succeeded, before the global timer expires, MPLS moves the all S2L sub-LSPs into their new path. Otherwise when this timer expires, MPLS

checks if all S2L paths have at least tried once. If so, it then aborts the global MBB. If not, it will continue until all S2Ls have re-tried once and then aborts the global MBB. Once global MBB is aborted, MPLS will move all S2L sub-LSPs into the new paths only if the set of S2Ls with a new path found is a superset of the S2Ls which have a current path which is up.

- While make-before break is being performed on individual S2L sub-LSP paths, the P2MP LSP will continue forwarding packets on S2L sub-LSP paths which are not being re-optimized and on the older S2L sub-LSP paths for which make-before-break operation was not successful. MBB will therefore result in duplication of packets until the old path is torn down.
- The MPLS data path of an LSR node, branch LSR node, and bud LSR node is able to re-merge S2L sub-LSP paths of the same P2MP LSP in case their ILM is on different incoming interfaces and their NHLFE is on the same or different outgoing interfaces. This could occur anytime there are equal cost paths through this node for the S2L sub-LSPs of this P2MP LSP.
- Link-protect FRR bypass using P2P LSPs is supported. In link protect, the PLR protecting an interface to a branch LSR will only make use of a single P2P bypass LSP to protect all S2L sub-LSPs traversing the protected interface.
- Refresh reduction on RSVP interface and on P2P bypass LSP protecting one or more S2L sub-LSPs.
- A manual bypass LSP cannot be used for protecting S2L paths of a P2MP LSP.
- The following MPLS features do operate with P2MP LSP:
  - BFD on RSVP interface.
  - MD5 on RSVP interface.
  - IGP metric and TE metric for computing the path of the P2MP LSP with CSPF.
  - SRLG constraint for computing the path of the P2MP LSP with CSPF. SRLG is supported on FRR backup path only.
  - TE graceful shutdown.
  - Admin group constraint.
- The following MPLS features are not operable with P2MP LSP:
  - Class based forwarding over P2MP RSVP LSP.
  - LDP-over-RSVP where the RSVP LSP is a P2MP LSP.
  - Diff-Serv TE.
  - Soft preemption of RSVP P2MP LSP.

2.7.5 Forwarding Multicast Packets over RSVP P2MP LSP in the Base Router

Multicast packets are forwarded over the P2MP LSP at the ingress LER based on a static join configuration of the multicast group against the tunnel interface associated with the originating P2MP LSP. At the egress LER, packets of a multicast group are received from the P2MP LSP via a static assignment of the specific <S,G> to the tunnel interface associated with a terminating LSP.

2.7.5.1 Procedures at Ingress LER Node

To forward multicast packets over a P2MP LSP, perform the following steps:

1. Create a tunnel interface associated with the P2MP LSP: **config>router>tunnel-interface rsvp-p2mp** *lsp-name*. (The config>router>pim>tunnel-interface command has been discontinued.)
2. Add static multicast group joins to the PIM interface, either as a specific <S,G> or as a <*,G>: **config>router>igmp>tunnel-if>static>group>source** *ip-address* and **config>router>igmp>tunnel-if>static>group>starg**.

The tunnel interface identifier consists of a string of characters representing the LSP name for the RSVP P2MP LSP. Note that MPLS will actually pass to PIM a more structured tunnel interface identifier. The structure will follow the one BGP uses to distribute the PMSI tunnel information in BGP multicast VPN as specified in draft-ietf-l3vpn-2547bis-mcast-bgp, *Multicast in MPLS/BGP IP VPNs*. The format is: <extended tunnel ID, reserved, tunnel ID, P2MP ID> as encoded in the RSVP-TE P2MP LSP session_attribute object in RFC 4875.

The user can create one or more tunnel interfaces in PIM and associate each to a different RSVP P2MP LSP. The user can then assign static multicast group joins to each tunnel interface. Note however that a given <*,G> or <S,G> can only be associated with a single tunnel interface.

A multicast packet which is received on an interface and which succeeds the RPF check for the source address is replicated and forwarded to all OIFs which correspond to the branches of the P2MP LSP. The packet is sent on each OIF with the label stack indicated in the NHLFE of this OIF. The packets will also be replicated and forwarded natively on all OIFs which have received IGMP or PIM joins for this <S,G>.

The multicast packet can be received over a PIM or IGMP interface which can be an IES interface, a spoke SDP-terminated IES interface, or a network interface.

In order to duplicate a packet for a multicast group over the OIF of both P2MP LSP branches and the regular PIM or IGMP interfaces, the tap mask for the P2MP LSP and that of the PIM based interfaces will need to be combined into a superset MCID.

2.7.5.2 Procedures at Egress LER Node

2.7.5.2.1 Procedures with a Primary Tunnel Interface

The user configures a tunnel interface and associates it with a terminating P2MP LSP leaf using the command: **config>router>tunnel-interface rsvp-p2mp lsp-name sender** *sender-address*. The **config>router>pim>tunnel-interface** command has been discontinued.

The tunnel interface identifier consists of a couple of string of characters representing the LSP name for the RSVP P2MP LSP followed by the system address of the ingress LER. The LSP name must correspond to a P2MP LSP name configured by the user at the ingress LER and must not contain the special character ":" Note that MPLS will actually pass to PIM a more structured tunnel interface identifier. The structure will follow the one BGP uses to distribute the PMSI tunnel information in BGP multicast VPN as specified in draft-ietf-l3vpn-2547bis-mcast-bgp.The format is: <extended tunnel ID, reserved, tunnel ID, P2MP ID> as encoded in the RSVP-TE P2MP LSP session_attribute object in RFC 4875.

The egress LER accepts multicast packets using the following methods:

1. The regular RPF check on unlabeled IP multicast packets, which is based on routing table lookup.
2. The static assignment which specifies the receiving of a multicast group <*,G> or a specific <S,G> from a primary tunnel-interface associated with an RSVP P2MP LSP.

One or more primary tunnel interfaces in the base router instance can be configured. In other words, the user is able to receive different multicast groups, <*,G> or specific <S,G>, from different P2MP LSPs. This assumes that the user configured static joins for the same multicast groups at the ingress LER to forward over a tunnel interface associated with the same P2MP LSP.

A multicast info policy CLI option allows the user to define a bundle and specify channels in the bundle that must be received from the primary tunnel interface. The user can apply the defined multicast info policy to the base router instance.

At any given time, packets of the same multicast group can be accepted from either the primary tunnel interface associated with a P2MP LSP or from a PIM interface. These are mutually exclusive options. As soon as a multicast group is configured against a primary tunnel interface in the multicast info policy, it is blocked from other PIM interfaces.

However, if the user configured a multicast group to be received from a given primary tunnel interface, there is nothing preventing packets of the same multicast group from being received and accepted from another primary tunnel interface. However, an ingress LER will not allow the same multicast group to be forwarded over two different P2MP LSPs. The only possible case is that of two ingress LERs forwarding the same multicast group over two P2MP LSPs towards the same egress LER.

A multicast packet received on a tunnel interface associated with a P2MP LSP can be forwarded over a PIM or IGMP interface which can be an IES interface, a spoke SDP-terminated IES interface, or a network interface.

Note that packets received from a primary tunnel-interface associated with a terminating P2MP LSP cannot be forwarded over a tunnel interface associated with an originating P2MP LSP.

2.8 Segment Routing With Traffic Engineering (SR-TE)

Segment routing adds to IS-IS and OSPF protocols the ability to perform shortest path routing and source routing using the concept of abstract segment. A segment can represent a local prefix of a node, a specific adjacency of the node (interface/ next-hop), a service context, or a specific explicit path over the network. For each segment, the IGP advertises an identifier referred to as Segment ID (SID).

When segment routing is used together with MPLS data plane, the SID is a standard MPLS label. A router forwarding a packet using segment routing will therefore push one or more MPLS labels.

Segment routing using MPLS labels can be used in both shortest path routing applications (refer to the *7450 ESS, 7750 SR, 7950 XRS, and VSR Unicast Routing Protocols Guide* for more information) and in traffic engineering (TE) applications, as described in this section.

The following are the objectives and applications of Segment Routing:

- ability for a node to specify a unicast shortest- or source-routed forwarding path with the same mechanism; re-use IGP to minimize the number of control plane protocols
- IGP-based MPLS tunnels without the addition of any other signaling protocol
- ability to tunnel services from ingress PE to egress PE with or without an explicit path, and without requiring forwarding plane or control plane state in intermediate nodes
- FRR: expand coverage of basic LFA to any topology with the use of source-routed backup path; pre-computation and set up of backup path without additional signaling
- support LFA policies with shared-risk constraints, admin-groups, link/node protection
- TE should include loose/strict options, distributed and centralized TE, path disjointness, ECMP-awareness, limited or no per-service state on midpoint and tail-end routers
- Fine Grained Flow Steering and Service Chaining via a centralized stateful Path Computation Element (PCE) such as the one provided by the Nokia Network Services Platform (NSP)

2.8.1 SR-TE Support

The following MPLS commands and nodes are supported:

- Global MPLS-level commands and nodes:

  interface, **lsp**, **path**, **shutdown**
- LSP-level commands and nodes:

  bfd, **bgp-shortcut**, **bgp-transport-tunnel**, **cspf, exclude**, **hop-limit**, **igp-shortcut, include**, **metric**, **metric-type, path-computation-method, primary**, **retry-limit**, **retry-timer**, **revert-timer**, **shutdown**, **to**, **from**, **vprn-auto-bind**
- Both primary and secondary paths are supported with a SR-TE LSP. The following primary path level commands and nodes are supported with SR-TE LSP:

  bandwidth, **bfd**, **exclude**, **hop-limit**, **include**, **priority**, **shutdown**

  The following secondary path level commands and nodes are supported with SR-TE LSP:

  bandwidth, **bfd**, **exclude**, **hop-limit**, **include**, **path-preference**, **priority**, **shutdown**, **srlg**, **standby**

The following MPLS commands and nodes are not supported:

- Global MPLS level commands and nodes not applicable to SR-TE LSP (configuration is ignored):

  admin-group-frr, **auto-bandwidth-multipliers**, **auto-lsp**, **bypass-resignal-timer**, **cspf-on-loose-hop**, **dynamic-bypass**, **exponential-backoff-retry**, **frr-object**, **hold-timer**, **ingress-statistics**, **least-fill-min-thd**, **least-fill-reoptim-thd**, **logger-event-bundling**, **lsp-init-retry-timeout**, **lsp-template**, **max-bypass-associations**, **mbb-prefer-current-hops**, **mpls-tp**, **p2mp-resignal-timer**, **p2mp-s2l-fast-retry**, **p2p-active-path-fast-retry**, **retry-on-igp-overload**, **secondary-fast-retry-timer**, **shortcut-local-ttl-propagate**, **shortcut-transit-ttl-propagate**, **srlg-database**, **srlg-frr**, **static-lsp**, **static-lsp-fast-retry**, **user-srlg-db**
- LSP level commands and nodes not supported with SR-TE LSP (configuration blocked):

  adaptive, **adspec**, **auto-bandwidth**, **class-type**, **dest-global-id**, **dest-tunnel-number**, **exclude-node**, **fast-reroute**, **ldp-over-rsvp**, **least-fill**, **main-ct-retry-limit**, **p2mp-id**, **primary-p2mp-instance**, **propagate-admin-group**, **protect-tp-path**, **rsvp-resv-style**, **working-tp-path**
- The following primary path level commands and nodes are not supported with SR-TE LSP:

  adaptive, **backup-class-type**, **class-type**, **record**, **record-label**

- The following secondary path level commands and nodes are not supported with SR-TE LSP:

  adaptive, **class-type**, **record**, **record-label**

The user can associate an empty path or a path with strict or loose explicit hops with the paths of the SR-TE LSP using the **hop**, **primary**, and **secondary** commands.

A hop that corresponds to an adjacency SID must be identified with its far-end host IP address (next-hop) on the subnet. If the local end host IP address is provided, this hop is ignored because this router can have multiple adjacencies (next-hops) on the same subnet.

A hop that corresponds to a node SID is identified by the prefix address.

Details of processing the user configured path hops are provided in SR-TE LSP Instantiation.

2.8.2 SR-TE LSP Instantiation

When an SR-TE LSP is configured on the router, its path can be computed by the router or by an external TE controller referred to as a Path Computation Element (PCE). This feature works with the Nokia stateful PCE which is part of the Network Services Platform (NSP).

The SR OS supports the following modes of operations which are configurable on a per SR-TE LSP basis:

- When the path of the LSP is computed by the router acting as a PCE Client (PCC), the LSP is referred to as PCC-initiated and PCC-controlled.

  A PCC-initiated and controlled SR-TE LSP has the following characteristics:
  - Can contain strict or loose hops, or a combination of both
  - Supports both a basic hop-to-label translation and a full CSPF as a path computation method.
  - The capability exists to report a SR-TE LSP to synchronize the LSP database of a stateful PCE server using the **pce-report** option, but the LSP path cannot be updated by the PCE. In other words, the control of the LSP is maintained by the PCC
- When the path of the LSP is computed by the PCE at the request of the PCC, it is referred to as PCC-initiated and PCE-computed.

A PCC-initiated and PCE-computed SR-TE LSP supports the Passive Stateful Mode, which enables the **path-computation-method pce** option for the SR-TE LSP so PCE can perform path computation at the request of the PCC only. PCC retains control.

The capability exists to report a SR-TE LSP to synchronize the LSP database of a stateful PCE server using the **pce-report** option.

- When the path of the LSP is computed and updated by the PCE following a delegation from the PCC, it is referred to as PCC-initiated and PCE-controlled.

  A PCC-initiated and PCE-controlled SR-TE LSP allows Active Stateful Mode, which enables the **pce-control** option for the SR-TE LSP so PCE can perform path computation and updates following a network event without the explicit request from the PCC. PCC delegates full control.

The user can configure the path computation requests only (PCE-computed) or both path computation requests and path updates (PCE-controlled) to PCE for a specific LSP using the **path-computation-method pce** and **pce-control** commands.

The **path-computation-method pce** option sends the path computation request to the PCE instead of the local CSPF. When this option is enabled, the PCE acts in Passive Stateful mode for this LSP. In other words, the PCE can perform path computations for the LSP only at the request of the router. This is used in cases where the operator wants to use the PCE specific path computation algorithm instead of the local router CSPF algorithm.

The default value is **no path-computation-method**.

The user can also enable the router's full CSPF path computation method. See SR-TE LSP Path Computation Using Local CSPF for more details.

The **pce-control** option allows the router to delegate full control of the LSP to the PCE (PCE-controlled). Enabling it means the PCE is acting in Active Stateful mode for this LSP and allows PCE to reroute the path following a failure or to re-optimize the path and update the router without requiring the router to request it.

Note:

- The user can delegate LSPs computed by either the local CSPF or the hop-to-label translation path computation methods.
- The user can delegate LSPs which have the **path-computation-method pce** option enabled or disabled. The LSP maintains its latest active path computed by PCE or the router at the time it was delegated. The PCE will only make an update to the path at the next network event or re-optimization. The default value is **no pce-control**.
- PCE report is supported for SR-TE LSPs with more than one path. However, PCE computation and PCE control are not supported in such cases. PCE computation and PCE control are supported for SR-TE LSPs with only one path that is either primary or secondary.

In all cases, the PCC LSP database is synchronized with the PCE LSP database using the PCEP Path Computation State Report (PCRpt) message for LSPs that have the **pce-report** command enabled.

The global MPLS level **pce-report** command can be used to enable or disable PCE reporting for all SR-TE LSPs for the purpose of LSP database synchronization. This configuration is inherited by all LSPs of a given type. The PCC reports both CSPF and non-CSPF LSP. The default value is disabled (**no pce-report**). This default value controls the introduction of PCE into an existing network and allows the operator to decide if all LSP types need to be reported.

The LSP level **pce-report** command overrides the global configuration for reporting LSP to PCE. The default value is to inherit the global MPLS level value. The **inherit** value returns the LSP to inherit the global configuration for that LSP type.

Note: If PCE reporting is disabled for the LSP, either due to inheritance or due to LSP level configuration, enabling the **pce-control** option for the LSP has no effect. To help troubleshoot this situation, operational values of both the **pce-report** and **pce-control** are added to the output of the LSP path **show** command.

For more information about configuring PCC-Initiated and PCC-Controlled LSPs, see Configuring PCC-controlled, PCE-computed, and PCE-controlled SR-TE LSPs.

2.8.2.1 PCC-Initiated and PCC-Controlled LSP

In this mode of operation, the user configures the LSP name, primary path name and optional secondary path name with the path information in the referenced path name, entering a full or partial explicit path with all or some hops to the destination of the LSP. Each hop is specified as an address of a node or an address of the next hop of a TE link. Optionally, each hop may be specified as a SID value corresponding to the MPLS label to use on a given hop. In this case, the whole path must consist of SIDs.

To configure a primary or secondary path to always use a specific link whenever it is up, the strict hop must be entered as an address corresponding to the next-hop of an adjacency SID, or the path must consist of SID values for every hop. If the strict hop corresponds to an address of a loopback address, it is translated into an adjacency SID as explained below and therefore does not guarantee that the same specific TE link is picked.

MPLS assigns a Tunnel-ID to the SR-TE LSP and a path-ID to each new instantiation of the primary path, as in an RSVP-TE LSP. These IDs are useful to represent the MBB path of the same SR-TE LSP which need to co-exist during the update of the primary path.

➡ **Note:** The concept of MBB is not exactly accurate in the context of a SR-TE LSP because there is no signaling involved and, as such, the new path information immediately overrides the older one.

The router retains full control of the path of the LSP. The LSP path label stack size is checked by MPLS against the maximum value configured for the LSP after the TE-DB returns the label stack. See Service and Shortcut Application SR-TE Label Stack Check for more information about this check.

The ingress LER performs the following steps to resolve the user-entered path before programming it in the data path:

Step 1. MPLS passes the path information to the TE-DB, which uses the hop-to-label translation or the full CSPF method to convert the list of hops into a label stack. The TE database returns the actual selected hop SIDs plus labels as well the configured path hop addresses which were used as the input for this conversion.

Step 2. The ingress LER validates the first hop of the path to determine the outgoing interface and next hop where the packet is to be forwarded and programs the data path according to the following conditions.

- If the first hop corresponds to an adjacency SID (host address of next-hop on the link's subnet), the adjacency SID label is not pushed. In other words, the ingress LER treats forwarding to a local interface as a push of an implicit-null label.
- If the first hop is a node SID of some downstream router, then the node SID label is pushed.

In both cases, the SR-TE LSP tracks and rides the SR shortest path tunnel of the SID of the first hop.

Step 3. In the case where the router is configured as a PCC and has a PCEP session to a PCE, the router sends a PCRpt message to update PCE with the state of UP and the RRO object for each LSP which has the **pce-report** option enabled. PE router does not set the delegation control flag to keep LSP control. The state of the LSP is now synchronized between the router and the PCE.

2.8.2.1.1 Guidelines for PCC-Initiated and PCC-Controlled LSPs

The router supports both a full CSPF and a basic hop-to-label translation path computation methods for a SR-TE LSP. In addition, the user can configure a path for the SR-TE LSP by explicitly entering SID label values.

The ingress LER has a few ways to detect a path is down or is not optimal and take immediate action:

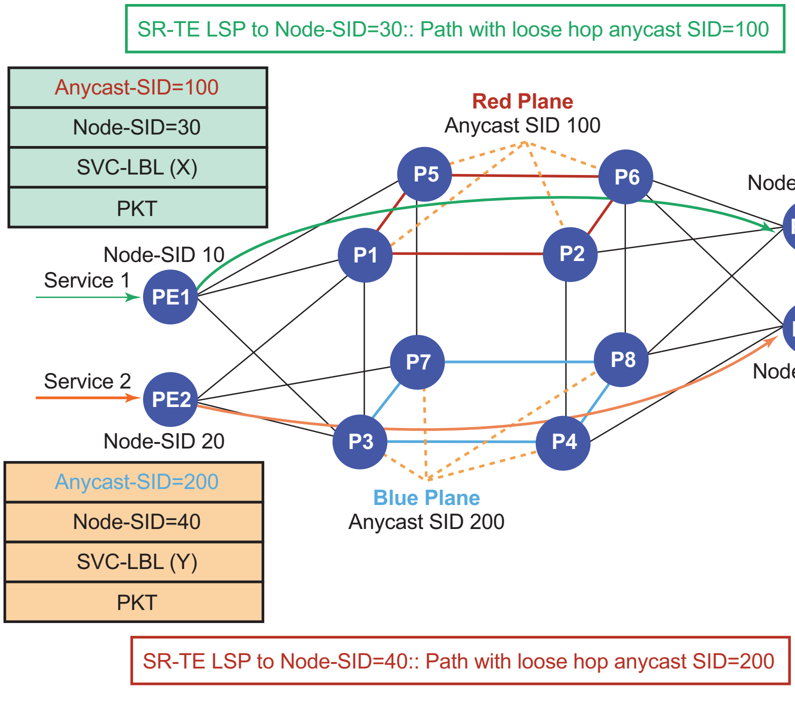
- Failure of the top SID detected via a local failure or an IGP network event. In this case, the LSP path goes down and MPLS will retry it.
- Timeout of the seamless BFD session when enabled on the LSP and the **failure-action** is set to the value of **failover-or-down**. In this case, the path goes down and MPLS will retry it.
- Receipt of an IGP link event in the TE database. In this case, MPLS performs an ad-hoc re-optimization of the paths of all SR-TE LSPs if the user enabled the MPLS level command **sr-te-resignal resignal-on-igp-event**. This capability only works when the path computation method is the local CSPF. It allows the ingress LER not only to detect a single remote failure event which causes packets to drop but also a network event which causes a node SID to reroute and thus forwarding packets on a potentially sub-optimal TE path.
- Performing a manual or timer based resignal of the SR-TE LSP. This applies only when the path computation method is the local CSPF. In this case, MPLS re-optimizes the paths of all SR-TE LSPs.

With both the hop-to-label path computation method and the user configured SID labels, the ingress LER does not monitor network events which affect the reachability of the adjacency SID or node SID used in the label stack of the LSP, except for the top SID. As a result, the label stack may not be updated to reflect changes in the path except when seamless BFD is used to detect failure of the path. It is therefore recommended to use this type of SR-TE LSP in the following configurations only:

- empty path
- path with a single node-SID loose-hop
- path of an LSP to a directly-connected router (single-hop LSP) with an adjacency-SID or a node-SID loose/strict hop
- strict path with hops consisting of adjacencies explicitly configured in the path as IP addresses or SID labels.

The user can also configure a SR-TE LSP with a single loose-hop using the anycast SID concept to provide LSR node protection within a given plane of the network TE topology. This is illustrated in Figure 37. The user configures all LSRs in a given plane with the same loopback interface address, which must be different from that of the system interface and the router-id of the router, and assigns them the same node-SID index value. All routers must use the same SRGB.

Figure 37 **Multi-plane TE with Node Protection**



Then user configures in a LER a SR-TE LSP to some destination and adds to its path either a loose-hop matching the anycast loopback address or the explicit label value of the anycast SID. The SR-TE LSP to any destination will hop over the closest of the LSRs owning the anycast SID because the resolution of the node-SID for that anycast loopback address uses the closest router. When that router fails, the resolution is updated to the next closest router owning the anycast SID without changing the label stack of the SR-TE LSP.

2.8.2.2 PCC-Initiated and PCE-Computed or Controlled LSP

In this mode of operation, the ingress LER uses Path Computation Element Communication Protocol (PCEP) to communicate with a PCE-based external TE controller (also referred to as the PCE). The router instantiates a PCEP session to the PCE. The router is referred to as the PCE Client (PCC).

The following PCE control modes are supported:

- passive control mode

  In this mode, the user enables the **path-computation-method pce** command for one or more SR-TE LSPs and a PCE performs path computations at the request of the PCC.
- active control mode

  In this mode, the user enables the **pce-control** command for an LSP, which allows the PCE to perform both path computation and periodic reoptimization of the LSP path without an explicit request from the PCC.

For the PCC to communicate with a PCE about the management of the path of a SR-TE LSP, the router implements the extensions to PCEP in support of segment routing (see the PCEP section for more information).This feature works with the Nokia stateful PCE, which is part of the Network Services Platform (NSP).

The following procedure describes configuring and programming a PCC-initiated SR-TE LSP when passive or active control is given to the PCE.

1. The SR-TE LSP configuration is created on the PE router via CLI or via OSS/SAM.

   The configuration dictates which PCE control mode is desired: active (**pce-control** option enabled) or passive (**path-computation-method pce** enabled and **pce-control** disabled).

2. The PCC assigns a unique PLSP-ID to the LSP. The PLSP-ID uniquely identifies the LSP on a PCEP session and must remain constant during its lifetime. PCC on the router tracks the association of {PLSP-ID, SRP-ID} to {Tunnel-ID, Path-ID} and uses the latter to communicate with MPLS about a specific path of the LSP.
3. The PE router does not validate the entered path. While the PCC can include the IRO objects for any loose or strict hop in the configured LSP path in the Path Computation Request (PCReq) message to PCE, the PCE ignores them and computes the path with the other constraints, excepting the IRO.
4. The PE router sends a PCReq message to the PCE to request a path for the LSP and includes the LSP parameters in the METRIC object, the LSPA object, and the Bandwidth object. It also includes the LSP object with the assigned PLSP-ID. At this point, the PCC does not delegate control of the LSP to the PCE.
5. PCE computes a new path, reserves the bandwidth, and returns the path in a Path Computation Reply (PCRep) message with the computed ERO in the ERO object. It also includes the LSP object with the unique PLSP-ID, the METRIC object with the computed metric value if any, and the Bandwidth object.

Note: For the PCE to use the SRLG path diversity and admin-group constraints in the path computation, the user must configure the SRLG and admin-group membership against the MPLS interface and verify that the traffic-engineering option is enabled in IGP. This causes IGP to flood the link SRLG and admin-group membership in its participating area and for the PCE to learn it in its TE database.

6. The PE router updates the CPM and the data path with the new path.

   Up to this step, the PCC and PCE are using passive stateful PCE procedures. The next steps synchronize the LSP database of the PCC and PCE for both PCE-computed and PCE-controlled LSPs. They also initiate the active PCE stateful procedures for the PCE-controlled LSP only.
7. PE router sends a PCRpt message to update PCE with the state of UP and the RRO as confirmation, including the LSP object with the unique PLSP-ID. For a PCE-controlled LSP, the PE router also sets a delegation control flag to delegate control to the PCE. The state of the LSP is now synchronized between the router and the PCE.
8. Following a network event or re-optimization, PCE computes a new path for a PCE-Controlled LSP and returns it in a Path Computation Update (PCUpd) message with the new ERO. It includes the LSP object with the same unique PLSP-ID assigned by the PCC and the Stateful Request Parameter (SRP) object with a unique SRP-ID-number to track error and state messages specific to this new path.

Note: If the **no pce-control** command is performed while a PCUpdate MBB is in progress on the LSP, the router aborts and removes the information and state related to the in-progress PCUpdate MBB. As the LSP is no longer controlled by the PCE, the router may take further actions depending on the state of the LSP. For example, if the LSP is up, and has FRR active or pre-emption, then the router starts a GlobalRevert or pre-emption MBB. If the LSP is down, the router starts the retry-timer to trigger setup.

9. The PE router updates the CPM and the data path with the new path.
10. The PE router sends a new PCRpt message to update PCE with the state of UP and the RRO as confirmation. The state of the LSP is now synchronized between the router and the PCE.
11. If the user makes any configuration change to the PCE-computed or PCE-controlled LSP, MPLS requests PCC to first revoke delegation in a PCRpt message (PCE-controlled only), and then MPLS and PCC follow the above steps to convey the changed constraint to PCE, which will result in a new path programmed into the data path, the LSP databases of PCC and PCE to be synchronized, and the delegation to be returned to PCE.

    In the case of an SR-TE LSP, MBB is not supported. Therefore, PCC first tears down the LSP and sends a PCRpt message to PCE with the Remove flag set to 1 before following this configuration change procedure.

Note: The preceding procedure is followed when the user performs a **no shutdown** on a PCE-controlled or PCE-computed LSP. The starting point is an administratively-down LSP with no active paths.

The following steps are followed for an LSP with an active path:

- If the user enabled the **path-computation-method pce** option on a PCC-controlled LSP which has an active path, no action is performed until the next time the router needs a path for the LSP following a network event of an LSP parameter change. At that point the procedures above are followed.
- If the user enabled the **pce-control** option on a PCC-controlled or PCE-computed LSP which has an active path, PCC will issue a PCRpt message to PCE with the state of UP and the RRO of the active path. It will set delegation control flag to delegate control to PCE. PCE will keep the active path of the LSP and will not update until the next network event or re-optimization. At that point the procedures above are followed.

The PCE supports the computation of disjoint paths for two different LSPs originating or terminating on the same or different PE routers. To indicate this constraint to PCE, the user must configure the PCE path profile ID and path group ID the LSP belongs to. These parameters are passed transparently by PCC to PCE and, so, opaque data to the router. The user can configure the path profile and path group using the **path-profile** *profile-id* [**path-group** *group-id*] command.

The association of the optional path group ID is to allow PCE determine which profile ID this path group ID must be used with. One path group ID is allowed per profile ID. The user can, however, enter the same path group ID with multiple profile IDs by executing this command multiple times. A maximum of five entries of **path-profile** [*path-group*] can be associated with the same LSP. More details of the operation of the PCE path profile are provided in the PCEP section of this guide.

2.8.3 SR-TE LSP Path Computation

For PCC-controlled SR-TE LSPs, CSPF is supported on the router using the **path-computation-method local-cspf** command. See SR-TE LSP Path Computation Using Local CSPF for details about the full CSPF path computation method. By default, the path is computed using the hop-to-label translation method. In the latter case, MPLS makes a request to the TE-DB to get the label corresponding to each hop entered by the user in the primary path of the SR-TE LSP. See SR-TE LSP Path Computation Using Hop-to-Label Translation for details of the hop-to-label translation.

The user can configure the path computation request of a CSPF-enabled SR-TE LSP to be forwarded to a PCE instead of the local router CSPF (**path-computation-method local-cspf** option enabled) by enabling the **path-computation-method pce** option, as explained in SR-TE LSP Instantiation. The user can further delegate the re-optimization of the LSP to the PCE by enabling the **pce-control** option. In both cases, PCE is responsible for determining the label required for each returned explicit hop and includes this in the SR-ERO.

In all cases, the user can configure the maximum number of labels which the ingress LER can push for a given SR-TE LSP by using the **max-sr-labels** command.

This command is used to set a limit on the maximum label stack size of the SR-TE LSP primary path so as to allow room to insert additional transport, service, and other labels when packets are forwarded in a given context.

CLI Syntax:
```
config>router>mpls>lsp>max-sr-labels label-stack-size
  [additional-frr-labels labels]
```

The **max-sr-labels** *label-stack-size* value should be set to account for the desired maximum label stack of the primary path of the SR-TE LSP. Its range is 1-11 and the default value is 6.

The value in **additional-frr-labels** *labels* should be set to account for additional labels inserted by remote LFA or Topology Independent LFA (TI-LFA) for the backup next-hop of the SR-TE LSP. Its range is 0-3 labels with a default value of 1.

The sum of both label values represents the worst case transport of SR label stack size for this SR-TE LSP and is populated by MPLS in the TTM such that services and shortcut applications can check it to decide if a service can be bound or a route can be resolved to this SR-TE LSP. More details of the label stack size check and requirements in various services and shortcut applications are provided in Service and Shortcut Application SR-TE Label Stack Check.

The maximum label stack supported by the router is discussed in Data Path Support and always signaled by PCC in the PCEP Open object as part of the SR-PCE-CAPABILITY TLV. It is referred to as the Maximum Stack Depth (MSD).

In addition, the per-LSP value for the **max-sr-labels** *label-stack-size* option, if configured, is signaled by PCC to PCE in the Segment-ID (SID) Depth value in a METRIC object for both a PCE-computed LSP and a PCE-controlled LSP. PCE will compute and provide the full explicit path with TE-links specified. If there is no path with the number of hops lower than the MSD value, or the SID Depth value if signaled, a reply with no path is returned to PCC.

For a PCC-controlled LSP, if the label stack returned by the TE-D exceeds the per LSP maximum SR label stack size, the LSP is brought down.

2.8.4 SR-TE LSP Path Computation Using Hop-to-Label Translation

MPLS passes the path information to the TE-DB, which converts the list of hops into a label stack as follows:

- A loose hop with an address matching any interface (loopback or not) of a router (identified by router-ID) is *always* translated to a node SID. If the prefix matching the hop address has a node SID in the TE database, it is selected by preference. If not, the node SID of any loopback interface of the same router that owns the hop address is selected. In the latter case, the lowest IP-address of that router that has a /32 Prefix-SID is selected.

- A strict hop with an address matching any interface (loopback or not) of a router (identified by router-ID) is always translated to an adjacency SID. If the hop address matches the host address reachable in a local subnet from the previous hop, then the adjacency SID of that adjacency is selected. If the hop address matches a loopback interface, it is translated to the adjacency SID of any link from the previous hop which terminates on the router owning the loopback. The adjacency SID label of the selected link is used.

  In both cases, it is possible to have multiple matching previous hops in the case of a LAN interface. In this case, the adjacency-SID with the lowest interface address is selected.
- In addition to the IGP instance that resolved the prefix of the destination address of the LSP in the RTM, all IGP instances are scanned from the lowest to the highest instance ID, beginning with IS-IS instances and then OSPF instances. For the first instance via which all specified path hop addresses can be translated, the label stack is selected. The hop-to-SID/label translation tool does not support paths that cross area boundaries. All SID/labels of a given path are therefore taken from the same IGP area and instance.
- Unnumbered network IP interfaces, which are supported in the router’s TE database, can be selected when converting the hops into an adjacency SID label when the user has entered the address of a loopback interface as a strict hop; however, the user cannot configure an unnumbered interface as a hop in the path definition.

Note: For the hop-to-label translation to operate, the user must enable TE on the network links, meaning to add the network interfaces to MPLS and RSVP. In addition, the user must enable the **traffic-engineering** option on all participating router IGP instances. Note that if any router has the **database-export** option enabled in the participating IGP instances to populate the learned IGP link state information into the TE-DB, then enabling of the **traffic-engineering** option is not required. For consistency purposes, it is recommended to have the **traffic-engineering** option always enabled.

2.8.5 SR-TE LSP Path Computation Using Local CSPF

This feature introduces full CSPF path computation for SR-TE LSP paths.

The hop-to-label translation, the local CSPF, or the PCE path computation methods for a SR-TE LSP can be user-selected with the following **path-computation-method** [**local-cspf** | **pce**] command. The **no** form of this command sets the computation method to the hop-to-label translation method, which is the default value. The **pce** option is not supported with the SR-TE LSP template.

2.8.5.1 Extending MPLS and TE Database CSPF Support to SR-TE LSP

The following are the MPLS and TE database features for extending CSPF support to SR-TE LSP:

- supports IPv4 SR-TE LSP
- supports local CSPF on both primary and secondary standby paths of an IPv4 SR-TE LSP
- supports local CSPF in LSP templates of types **mesh-p2p-srte** and **one-hop-p2p-srte** of SR-TE auto-LSP
- supports path computation in single area OSPFv2 and IS-IS IGP instances
- computes full explicit TE paths using TE links as hops and returning a list of SIDs consisting of adjacency SIDs and parallel adjacency set SIDs. SIDs of a non-parallel adjacency set is not used in CSPF. The details of the CSPF path computation are provided in SR-TE Specific TE-DB Changes. Loose-hop paths, using a combination of node SID and adjacency SID, are not required.
- uses random path selection in the presence of ECMP paths that satisfy the LSP and path constraints. Least-fill path selection is not required.
- provides an option to reduce or compress the label stack such that the adjacency SIDs corresponding to a segment of the explicit path are replaced with a node SID whenever the constraints of the path are met by all the ECMP paths to that node SID. The details of the label reduction are provided in SR-TE LSP Path Label Stack Reduction.
- uses legacy TE link attributes as in RSVP-TE LSP CSPF
- uses timer re-optimization of all paths of the SR-TE LSP that are in the operational UP state. This differs from RSVP-TE LSP resignal timer feature which re-optimizes the active path of the LSP only.

  MPLS provides the current path of the SR-TE LSP and TE-DB updates the total IGP or TE metric of the path, checking the validity of the hops and labels as per current TE-DB link information. CSPF then calculates a new path and provides both the new and metric updated current path back to MPLS. MPLS programs the new path only if the total metric of the new computed path is different than the updated metric of the current path, or if one or more hops or labels of the current path are invalid. Otherwise, the current path is considered one of the most optimal ECMP paths and is not updated in the data path.

  Timer resignal applies only to the CSPF computation method and not to the ip-to-label computation method.
- uses manual re-optimization of a path of the SR-TE LSP. In this case, the new computed path is always programmed even if the metric or SID list is the same.

- supports ad-hoc re-optimization. This SR-TE LSP feature for SR-TE LSP triggers the ad-hoc resignaling of all SR-TE LSPs if one or more IGP link down events are received in TE-DB.

  Once the re-optimization is triggered, the behavior is the same as the timer-based resignal or the delay option of the manual resignal. MPLS forces the expiry of the resignal timer and asks TE-DB to re-evaluate the active paths of all SR-TE LSPs. The re-evaluation consists of updating the total IGP or TE metric of the current path, checking the validity of the hops and labels, and computing a new CSPF for each SR-TE LSP. MPLS programs the new path only if the total metric of the new computed path is different than the updated metric of the current path, or if one or more hops or labels of the current path are invalid. Otherwise, the current path is considered one of the most optimal ECMP paths and is not updated in the data path.
- supports using unnumbered interfaces in the path computation. There is no support for configuring an unnumbered interface as a hop in the path of the LSP is not required. So, the path can be empty or include hops with the address of a system or loopback interface but path computation can return a path that uses TE links corresponding to unnumbered interfaces.
- supports **admin-group**, **hop-count**, IGP metric, and TE-metric constraints
- bandwidth constraint is not supported since SR-TE LSP does not have an LSR state to book bandwidth. Thus, the **bandwidth** parameter, when enabled on the LSP path, has no impact on local CSPF path calculation. However, the **bandwidth** parameter is passed to PCE when it is the selected path computation method. PCE reserves bandwidth for the SR-TE LSP path accordingly.

2.8.5.2 SR-TE Specific TE-DB Changes

When the **traffic-engineering** command is enabled in an OSPFv2 instance in the current implementation of SR OS, only local and remote TE-enabled links are added into the TE-DB. A TE-link is a link that has one or more TE attributes added to it in the MPLS interface context. Link TE attributes are TE metric, bandwidth, and membership in a SRLG or an Admin-Group.

In order to allow the SR-TE LSP path computation to use SR-enabled links which do not have TE attributes, the following changes are made:

- OSPFv2 is modified to pass all links, regardless if they are TE-enabled or SR-enabled, to TE-DB as currently performed by IS-IS.

- TE-DB relaxes the link back-check when performing a CSPF calculation to ensure that there is at least one link from the remote router to the local router. Since OSPFv2 advertises the remote link IP address or remote link identifier only when a link is TE-enabled, the strict check about the reverse direction of a TE-link cannot be performed if the link is SR-enabled but not TE-enabled.

As a consequence of this change, CSPF can compute an SR-TE LSP with SR-enabled links that do not have TE attributes. This means that if the user admin shuts down an interface in MPLS, an SR-TE LSP path which uses this interface will not go operationally down.

2.8.5.3 SR-TE LSP and Auto-LSP-Specific CSPF Changes

The local CSPF for a SR-TE LSP is performed in two phases. The first phase (Phase 1) computes a fully explicit path with all TE links to the destination specified as in the case of a RSVP-TE LSP.

If the user enabled label stack reduction or compression for this LSP, a second phase (Phase 2) is applied to reduce the label stack so that adjacency SIDs corresponding to a segment of the explicit path are replaced with a node SID whenever the constraints of the path are met by all the ECMP paths to that node SID. The details of the label reduction are provided in SR-TE LSP Path Label Stack Reduction.

The CSPF computation algorithm for the fully explicit path in the first phase remains mostly unchanged from its behavior in RSVP-TE LSP.

The meaning of a strict and loose hop in the path of the LSP are the same as in CSPF for RSVP-TE LSP. A strict hop means that the path from the previous hop must be a direct link. A loose hop means the path from the previous hop can traverse intermediate routers.

A loose hop may be represented by a set of back-to-back adjacency SIDs if not all paths to the node SID of that loose hop satisfy the path TE constraints. This is different from the ip-to-label path computation method where a loose hop always matches a node SID since no TE constraints are checked in the path to that loose hop.

When the label stack of the path is reduced or compressed, it is possible that a strict hop is represented by a node SID, if all the links from the previous hop satisfy the path TE constraints. This is different from the ip-to-label path computation method wherein a strict hop always matches an adjacency SID or a parallel adjacency set SID.

The first phase of CSPF returns a full explicit path with each TE link specified all the way to the destination and which label stack may contain protected adjacency SIDs, unprotected adjacency SIDs, and adjacency set SIDs. The user can influence the type of adjacency protection for the SR-TE LSP using a CLI command as explained in SR-TE LSP Path Protection.

The SR OS does not support the origination of a global adjacency SID. If received from a third-party router implementation, it is added into the TE database but is not used in any CSPF path computation.

2.8.5.3.1 SR-TE LSP Path Protection

Also introduced with SR-TE LSP is the indication by the user if the path of the LSP must use protected or unprotected adjacencies exclusively for all links of the path.

When SR OS routers form an IGP adjacency over a link and segment-routing context is enabled in the IGP instance, the static or dynamic label assigned to the adjacency is advertised in the link adjacency SID sub-TLV. By default, an adjacency is always eligible for LFA/RLFA/TI-LFA protection and the B-flag in the sub-TLV is set. The presence of a B-flag does not reflect the instant state of the availability of the adjacency LFA backup; it reflects that the adjacency is eligible for protection. The SR-TE LSP using the adjacency in its path still comes up if the adjacency does not have a backup programmed in the data path at that instant. Use the **configure>router>isis>interface> no sid-protection** command to disable protection. When protection is disabled, the B-flag is cleared and the adjacency is not eligible for protection by LFA/RLFA/TI-LFA.

SR OS also supports the adjacency set feature that treats a set of adjacencies as a single object and advertises a link adjacency sub-TLV for it with the S-flag (SET flag) set to 1. The adjacency set in the SR OS implementation is always unprotected, even if there is a single member link in it and therefore the B-flag is always clear. Only a parallel adjacency set, meaning that all links terminate on the same downstream router, are used by the local CSPF feature.

Be aware that the same P2P link can participate in a single adjacency and in one or more adjacency sets. Therefore, multiple SIDs can be advertised for the same link.

Third party implementations of Segment Routing may advertise two SIDs for the same adjacency: one protected with B-flag set and one unprotected with B-flag clear. SR OS can achieve the same behavior by adding a link to a single-member adjacency SET, in which case a separate SID is advertised for the SET and the B-flag is cleared while the SID for the regular adjacency over that link has its B-flag set by default. In all cases, SR OS CSPF can use all local and remote SIDs to compute a path for an SR-TE LSP based on the desired local protection property.

There are three different behaviors of CSPF introduced with SR-TE LSP with respect to local protection:

1. When the **local-sr-protection** command is not enabled (**no local-sr-protection**) or is set to **preferred**, the local CSPF prefers a protected adjacency over an unprotected adjacency whenever both exist for a TE link. This is done on a link-by-link basis after the path is computed based on the LSP path constraints. This means that the protection state of the adjacency is not used as a constraint in the path computation. It is only used to select an SID among multiple SIDs once the path is selected. Thus, the computed path can combine both types of adjacencies.

   If a parallel adjacency set exists between two routers in a path and all the member links satisfy the constraints of the path, a single protected adjacency is selected in preference to the parallel adjacency set which is selected in preference to a single unprotected adjacency.

   If multiples ECMP paths satisfy the constraints of the LSP path, one path is selected randomly and then the SID selection above applies. There is no check if the selected path has the highest number of protected adjacencies.
2. When the **local-sr-protection** command is set to a value of **mandatory**, CSPF uses it as an additional path constraint and selects protected adjacencies exclusively in computing the path of the SR-TE LSP. Adjacency sets cannot be used because they are always unprotected.

   If no path that satisfies the other LSP path constraints and consists of all TE links with protected adjacencies, the path computation returns no path.
3. Similarly, when the **local-sr-protection** command to **none**, CSPF uses it as an additional path constraint and selects unprotected adjacencies exclusively in computing the path of the SR-TE LSP.

   If a parallel adjacency set exists between two routers in a path and all the member links satisfy the constraints of the path, it is selected in preference to a single unprotected adjacency.

   If no path satisfies the other LSP path constraints and consists of all TE links with unprotected adjacencies, the path computation returns no path.

The **local-sr-protection** command impacts PCE-computed and PCE-controlled SR-TE LSP. When the **local-sr-protection** command is set to the default value **preferred**, or to the explicit value of **mandatory**, the local-protection-desired flag (L-flag) in the LSPA object in the PCReq (Request) message or in the PCRpt (Report) message is set to a value of 1.

When the **local-sr-protection** command is set to **none**, the local-protection-desired flag (L-flag) in the LSPA object is cleared. The PCE path computation checks this flag to decide if protected adjacencies are used in preference to unprotected adjacencies (L-flag set) or must not be used at all (L-flag clear) in the computation of the SR-TE LSP path.

2.8.5.3.2 SR-TE LSP Path Label Stack Reduction

The objective of the label stack reduction is twofold:

- It reduces the label stack so ingress PE routers with a lower Maximum SID Depth (MSD) can still work.
- It provides the ability to spray packets over ECMP paths to an intermediate node SID when all these paths satisfy the constraints of the SR-TE LSP path. Even if the resulting label stack is not reduced, this aspect of the feature is still useful.

If the user enables the **label-stack-reduction** command for this LSP, a second phase is applied attempting to reduce the label stack that resulted from the fully explicit path with adjacency SIDs and adjacency sets SIDs computed in the first phase.

This is to attempt a replacement of adjacency and adjacency set SIDs corresponding to a segment of the explicit path with a node SID whenever the constraints of the path are met by all the ECMP paths to that node SID.

This is the procedure followed by the label stack reduction algorithm:

1. Phase 1 of the CSPF returns up to three fully explicit ECMP paths that are eligible for label stack reduction. These paths are equal cost from the point of view of IGP metric or TE metric as configured for that SR-TE LSP.
2. Each fully explicit path of the SR-TE LSP that is computed in Phase 1 of the CSPF is split into a number of segments that are delimited by the user-configured loose or strict hops in the path of the LSP. Label stack reduction is applied to each segment separately.
3. Label stack reduction in Phase 2 consists of traversing the CSPF tree for each ECMP path returned in Phase 1 and then attempting to find the farthest node SID in a path segment that can be used to summarize the entire path up to that node SID. This requires that all links of ECMP paths are able to reach the node SID from the current node on the CSPF tree in order to satisfy all the TE constraints of the SR-TE LSP paths. ECMP is based on the IGP metric, in this case, since this is what routers use in the data path when forwarding a packet to the node SID.

   If the TE metric is enabled for the SR-TE LSP, then one of the constraints is that the TE metric must be the same value for all the IGP metric ECMP paths to the node SID.
4. CSPF in Phase 2 selects the first candidate ECMP path from Phase 1 which reduced label stack that satisfies the constraint carried in the **max-sr-labels** command.

5. The CSPF path computation in Phase 1 always avoids a loop over the same hop as is the case with the RSVP-TE LSP. In addition, the label stack reduction algorithm prevents a path from looping over the same hop due to the normal routing process. For example, it checks if the same node is involved in the ECMP paths of more than one segment of the LSP path and builds the label stack to avoid this situation.
6. During the MBB procedure of a timer or manual re-optimization of a SR-TE LSP path, the TE-DB performs additional steps as compared to the case of the initial path computation:
   - MPLS provides TE-DB with the current working path of the SR-TE LSP.
   - TE-DB updates the path's metric based on the IGP or TE link metric (if TE metric enabled for the SR-TE LSP).
     - For each adjacency SID, it verifies that the related link and SID are still in its database and that the link fulfills the LSP constraints. If so, it picks up the current metric.
     - For each node SID, it verifies that the related prefix and SID are still available, and if so, checks that all the links on the shortest IGP path to the node owning the node SID fulfill the SR-TE LSP path constraints. This is re-using the same checks detailed in Step 3 for the label compression algorithm.
   - CSPF computes a new path with or without label stack reduction as explained in Steps 1, 2, and 3.
   - TE-DB returns both paths to MPLS. MPLS always programs the new path in the case of a manual re-optimization. MPLS compares the metric of the new path to the current path and if different, programs the new path in the case of a timer re-optimization.
7. TE-DB returns to MPLS the following additional information together with the reduced path ERO and label stack:
   - a list of SRLGs of each hop in the ERO represented by a node SID and that includes SRLGs used by links in all ECMP paths to reach that node SID from the previous hop.
   - the cost of each hop in the ERO represented by an adjacency SID or adjacency set SID. This corresponds to the IGP metric or TE metric (if TE is metric-enabled for the SR-TE LSP) of that link or set of links. In the case of an adjacency set, all TE metrics of the links must be the same, otherwise CSPF does not select the set.
   - the cost of each hop in the ERO represented by a node SID and this corresponds to the cumulated IGP metric or TE metric (if TE metric is enabled for the SR-TE LSP) to reach the node SID from the previous hop using the fully explicit path computed in Phase 1.

  - the total cost or computed metric of the SR-TE LSP path. This consists of the cumulated IGP metric or TE metric (if TE metric enabled for the SR-TE LSP) of all hops of the fully explicit path computed in Phase 1 of the CSPF.

8. If label stack reduction is disabled, the values of the **max-sr-labels** and the **hop-limit** commands are applied to the full explicit path in Phase 1.

   The minimum of the two values is used as a constraint in the full explicit path computation.

   If the resulting ECMP paths net hop-count in Phase 1 exceeds this minimum value no path is returned by TE-DB to MPLS

9. If label stack reduction is enabled, the values of the **max-sr-labels** and the **hop-limit** commands are both ignored in Phase 1 and only the value of the max-sr-labels is used as a constraint in Phase 2.

   If the resulting net label stack size after reduction of all candidate paths in Phase 2 exceeds the value of parameter **max-sr-labels** then no path is returned by TE-DB to MPLS.

10. The label stack reduction does not support the use of an anycast SID, a prefix SID with N-flag clear, in order to replace a segment of the SR-TE LSP path. Only a node SID is used.

2.8.5.3.3 Interaction with SR-TE LSP Path Protection

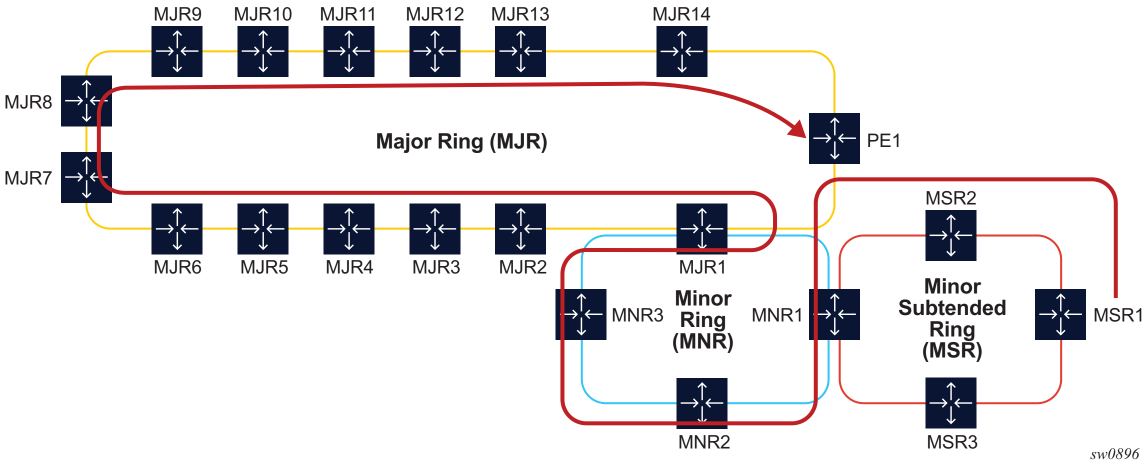
Label stack reduction is only attempted when the path protection **local-sr-protection** command is disabled or is configured to the value of **preferred**.

If **local-sr-protection** is configured to a value of **none** or **mandatory**, the command is ignored, and the fully explicit path computed out of Phase 1 is returned by the TE-DB CSPF routine to MPLS. This is because a node SID used to replace an adjacency SID or an adjacency set SID can be unprotected or protected by LFA and this is based on local configuration on each router which resolves this node SID but is not directly known in the information advertised into the TE-DB. Therefore, CSPF cannot enforce the protection constraint requested along the path to that node SID.

2.8.5.3.4 Examples of SR-TE LSP Path Label Stack Reduction

Figure 38 illustrates a metro aggregation network with three levels of rings for aggregating regional traffic from edge ring routers into a PE router.

Figure 38 **Label Stack Reduction in a 3-Tier Ring Topology**



The path of the highlighted LSP uses admin groups to force the traffic eastwards or westwards over the 3-ring topologies such that it uses the longest path possible. Assume all links in a bottom-most ring1 have admin-group=east1 for the eastward direction and admin-group=west1 for the westward direction.

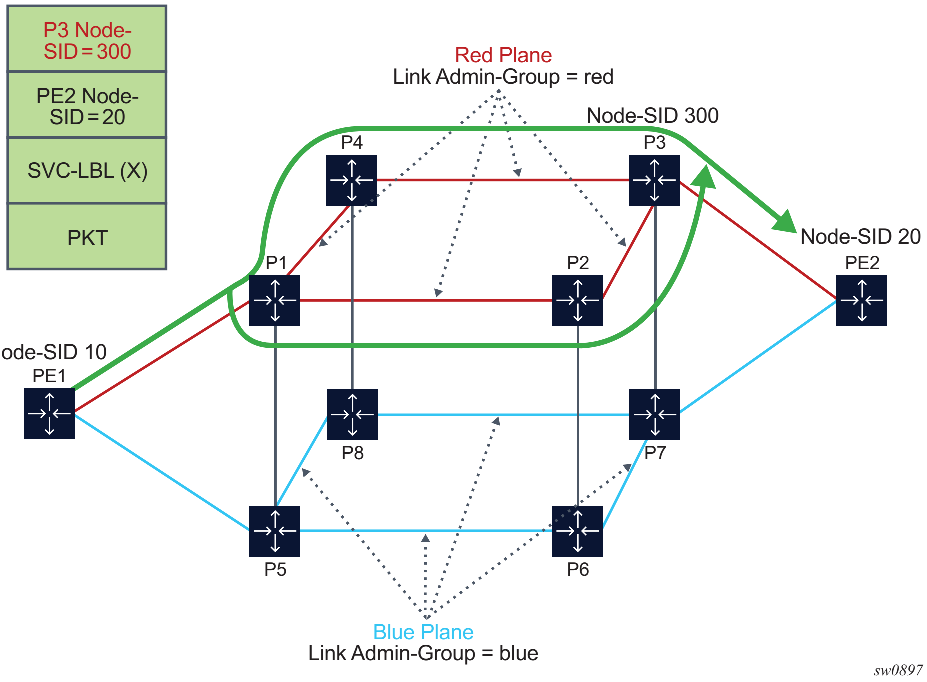
Similarly, links in middle ring2 have admin-group=east2 and admin-group=west2 and links in top-most ring3 have admin-group=east3 and admin-group=west3. To achieve the longest path shown, the LSP or path should have an include statement: include east1 west2 east3. The fully explicit path computed in Phase 1 of CSPF results in label stack of size 18.

The label stack reduction algorithm searches for the farthest node SID in that path which can replace a segment of the strict path while maintaining the stated admin-group constraints. The reduced label stack contains the SID adjacency MSR1-MSR2, the found node SIDs plus the node SID of the destination for a total of four labels to be pushed on the packet (the label for the adjacency MSR1-MSR2 is not pushed):

{N-SID MNR2, N-SID of MNR3, N-SID of MJR8, N-SID of PE1}

Figure 39 illustrates an example topology which creates two TE planes by applying a common admin group to all links of a given plane. There are a total of four ECMP paths to reach PE2 from PE1, two within the red plane and two within the blue plane.

Figure 39 **Label Stack Reduction in the Presence of ECMP Paths**



For a SR-TE LSP from PE1 to PE2 which includes the red admin-group as a constraint, Phase 1 of CSPF results in two fully explicit paths using adjacency SID of the red TE links:

path 1 = {PE1-P1, P1-P2, P2-P3, P3-PE2}

path 2 = {PE1-P1, P1-P4, P4-P3, P3-PE2}

Phase 2 of CSPF finds node SID of P3 as the farthest hop it can reach directly from PE1 while still satisfying the ‘include red’ admin-group constraint. If the node SID of PE2 is used as the only SID, then traffic would also be sent over the blue links.

Then, the reduced label stack is: {P3 Node-SID=300, PE2 Node-SID=20}.

The resulting SR-TE LSP path combines the two explicit paths out of Phase 1 into a single path with ECMP support.

2.8.6 SR-TE LSP Paths using Explicit SIDs

SR OS supports the ability for SR-TE primary and secondary paths to use a configured path containing explicit SID values. The SID value for an SR-TE LSP hop is configured using the **sid-label** command under **configure**>**router**>**mpls**>**path** as follows:

```
configure router mpls
         path <name>
            [no] hop <hop-index> sid-label <sid-value>
```

Where *sid-value* specifies an MPLS label value for that hop in the path.

When SIDs are explicitly configured for a path, the user must provide all of the necessary SIDs to reach the destination. The router does not validate whether the whole label stack provided is correct other than checking that the top SID is programmed locally. A path can come up even if it contains SIDs that are invalid. The user or controller programming the path should ensure that the SIDs are correct. A path must consist of either all SIDs or all IP address hops.

A path containing SID label hops is used even if **path-computation-method** {**local-cspf** | **pce**} is configured for the LSP. That is, the path computation method configured at the LSP level is ignored when explicit SIDs are used in the path. This means that the router can bring up the path if the configured path contains SID hops even if the LSP has path computation enabled.

Note: When an LSP consists of some SID label paths and some paths under local-CSPF computation, the router cannot guarantee SRLG diversity between the CSPF paths and the SID label paths because CSPF does not know of the existence of the SID label paths because they are not listed in the TE database.

Paths containing explicit SID values can only be used by SR-TE LSPs.

2.8.7 SR-TE LSP Protection

The router supports local protection of a given segment of an SR-TE LSP, and end-to-end protection of the complete SR-TE LSP.

Each path is locally protected along the network using LFA/remote-LFA next-hop whenever possible. The protection of a node SID re-uses the LFA and remote LFA features introduced with segment routing shortest path tunnels; the protection of an adjacency SID has been added to the SR OS in the specific context of an SR-TE LSP to augment the protection level. The user must enable the **loopfree-alternates** [**remote-lfa**] option in IS-IS or OSPF.

An SR-TE LSP has state at the ingress LER only. The LSR has state for the node SID and adjacency SID, whose labels are programmed in label stack of the received packet and which represent the part of the ERO of the SR-TE LSP on this router and downstream of this router. In order to provide protection for a SR-TE LSP, each LSR node must attempt to program a link-protect or node-protect LFA next-hop in the ILM record of a node SID or of an adjacency SID and the LER node must do the same in the LTN record of the SR-TE LSP. The following are details of the behavior:

- When the ILM record is for a node SID of a downstream router which is not directly connected, the ILM of this node SID points to the backup NHLFE computed by the LFA SPF and programmed by the SR module for this node SID. Depending on the topology and LFA policy used, this can be a link-protect or node-protect LFA next-hop.

  This behavior is already supported in the SR shortest path tunnel feature at both LER and LSR. As such, an SR-TE LSP that transits at an LSR and that matches the ILM of a downstream node SID automatically takes advantage of this protection when enabled. If required, node SID protection can be disabled under the IGP instance by excluding the prefix of the node SID from LFA.
- When the ILM is for a node SID of a directly connected router, then the LFA SPF only provides link protection. The ILM or LTN record of this node SID points to the backup NHLFE of this LFA next-hop. An SR-TE LSP that transits at an LSR and that matches the ILM of a neighboring node SID automatically takes advantage of this protection when enabled.

➡ **Note:** Only link protection is possible in this case because packets matching this ILM record can either terminate on the neighboring router owning the node SID or can be forwarded to different next-hops of the neighboring router; that is, to different next-next-hops of the LSR providing the protection. The LSR providing the connection does not have context to distinguish among all possible SR-TE LSPs and, as such, can only protect the link to the neighboring router.

- When the ILM or LTN record is for an adjacency SID, it is treated as in the case of a node SID of a directly connected router (as above).

When protecting an adjacency SID, the PLR first tries to select a parallel link to the node SID of the directly connected neighbor. That is the case when this node SID is reachable over parallel links. The selection is based on lowest interface ID. When no parallel links exist, then regular LFA/rLFA algorithms are applied to find a loopfree path to reach the node SID of the neighbor via other neighbors.

The ILM or LTN for the adjacency SID must point to this backup NHLFE and will benefit from FRR link-protection. As a result, an SR-TE LSP that transits at an LSR and matches the ILM of a local adjacency SID automatically takes advantage of this protection when enabled.

- At the ingress LER, the LTN record points to the SR-TE LSP NHLFE, which itself will point to the NHLFE of the SR shortest path tunnel to the node SID or adjacency SID of the first hop in the ERO of the SR-TE LSP. As such, the FRR link or node protection at ingress LER is inherited directly from the SR shortest path tunnel.

When an adjacency to a neighbor fails, IGP withdraws the advertisement of the link TLV information as well as its adjacency SID sub-TLV. However, the LTN or ILM record of the adjacency SID must be kept in the data path for a sufficient period of time to allow the ingress LER to compute a new path after IGP converges. If the adjacency is restored before the timer expires, the timer is aborted as soon as the new ILM or LTN records are updated with the new primary and backup NHLFE information. By default, the ILM/LTN and NHLFE information is kept for a period of 15 seconds.

The adjacency SID hold timer is configured using the **adj-sid-hold** command, and activated when the adjacency to neighbor fails due to the following conditions:

- The network IP interface went down due a link or port failure or due to the user performing a shutdown of the port.
- The user shuts down the network IP interface in the **config**>**router** or **config**>**router**>**ospf**/**isis** context.
- The adjacency SID hold timer is not activated if the user deleted an interface in the **config**>**router**>**ospf**/**isis** context.

Note:

- The adjacency SID hold timer does not apply to the ILM or LTN of a node SID, because NHLFE information is updated in the data path as soon as IGP is converged locally and a new primary and LFA backup next-hops have been computed.
- The label information of the primary path of the adjacency SID is maintained and re-programmed if the adjacency is restored before the above timer expires. However, the backup NHLFE may change when a new LFA SPF is run while the adjacency ILM is being held by the timer running. An update to the backup NHLFE is performed immediately following the LFA SPF and may cause packets to drop.
- A new PG-ID is assigned each time an adjacency comes back up. This PG-ID is used by the ILM of the adjacency SID and the ILMs of all downstream node SIDs which resolve to the same next-hop.

While protection is enabled globally for all node SIDs and local adjacency SIDs when the user enables the **loopfree-alternates** option in ISIS or OSPF at the LER and LSR, there are applications where the user wants traffic to never divert from the strict hop computed by CSPF for a SR-TE LSP. In that case, the user can disable protection for all adjacency SIDs formed over a given network IP interface using the **sid-protection** command.

The protection state of an adjacency SID is advertised in the B-FLAG of the IS-IS or OSPF Adjacency SID sub-TLV. No mechanism exists in PCEP for the PCC to signal to PCE the constraint to use only adjacency SIDs, which are not protected. The Path Profile ID is configured in PCE with the no-protection constraint.

2.8.7.1 Local Protection

Each path may be locally protected through the network using LFA/remote-LFA nexthop whenever possible. The protection of a SID node re-uses the LFA and remote LFA features introduced with segment routing shortest path tunnels; the protection of an adjacency SID has been added to the SR OS in the specific context of an SR-TE LSP to augment the protection level. The user must enable the **loopfree-alternates remote-lfa** option in IS-IS or OSPF.

This behavior is already supported in the SR shortest path tunnel feature at both LER and LSR. As such, an SR-TE LSP that transits at an LSR and that matches the ILM of a downstream SID node automatically takes advantage of this protection when enabled. If required, SID node protection can be disabled under the IGP instance by excluding the prefix of the SID node from LFA.

2.8.7.2 End to End Protection

This section provides a brief introduction to end to end protection for SR-TE LSPs. See Seamless BFD for SR-TE LSPs for more detailed description of protection switching using Seamless BFD and a configured failure-action.

End-to-end protection for SR-TE LSPs is provided using secondary or standby paths. Standby paths are permanently programmed in the data path, while secondary paths are only programmed when they are activated. S-BFD is used to provide end-to-end connectivity checking. The **failure-action failover-or-down** command under the **bfd** context of the LSP configures a switchover from the currently active path to an available standby or secondary path if the S-BFD session fails on the currently active path. If S-BFD is not configured, then the router that is local to a segment, can only detect failures of the top SID for that segment. End-to-end protection with S-BFD may be combined with local protection, but it is recommended that the S-BFD control packet timers be set to 1 second or more to allow sufficient time for any local protection action for a given segment to complete without triggering S-BFD to go down on the end to end LSP path.

To prevent failure between the paths of an SR-TE LSP, that is to avoid, for example, a failure of a primary path that affects its standby backup path, then disjoint paths should be configured or the **srlg** command configured on the secondary paths.

As with RSVP-TE LSPs, SR-TE standby paths support the configuration of a path preference. This value is used to select the standby path to be used when more than one available path exists.

For more details of end to end protection of SR-TE LSPs with S-BFD, see section Seamless BFD for SR-TE LSPs.

2.8.8 Seamless BFD for SR-TE LSPs

Seamless BFD (S-BFD) is a form of BFD that requires significantly less state and reduces the need for session bootstrapping as compared to LSP BFD. For more information, refer to “Seamless Bidirectional Forwarding Detection (S-BFD)” in *7450 ESS, 7750 SR, 7950 XRS, and VSR OAM and Diagnostics Guide*. S-BFD also requires centralized configuration for the reflector function, as well as a mapping at the head-end node between the remote session discriminator and the IP address for the reflector by each session. This configuration and the mapping are described in the *7450 ESS, 7750 SR, 7950 XRS, and VSR OAM and Diagnostics Guide*. This user guide describes the application of S-BFD to SR-TE LSPs, and the LSP configuration required for this feature.

S-BFD is supported in the following SR objects or contexts:

- PCC-Initiated:
  - SR-TE LSP level
  - SR-TE primary path
  - SR-TE secondary and standby path
- PCE-Initiated SR-TE LSPs
- SR-TE auto-LSPs

2.8.8.1 Configuration of S-BFD on SR-TE LSPs

For PCC-initiated or PCC-controlled LSPs, it is possible to configure an S-BFD session under the SR-TE LSP context, the primary path context, and the SR-TE secondary path by using the **config>router>mpls>lsp**, **config>router>mpls >lsp>primary**, and **config>router>mpls>lsp>secondary** commands.

The remote discriminator value is determined by passing the “to” address of the LSP to BFD, which then matches it to a mapping table of peer IP addresses to reflector remote discriminators, that are created by the centralized configuration under the IGP (refer to the *7450 ESS, 7750 SR, 7950 XRS, and VSR OAM and Diagnostics Guide*). If there is no match to the “to” address of the LSP, then a BFD session is not established on the LSP or path.

Note: A remote peer IP address to discriminator mapping must exist prior to bringing an LSP administratively up.

The referenced BFD template must specify parameters consistent with an S-BFD session. For example, the endpoint type is **cpm-np** for platforms supporting a CPM P-chip, otherwise a CLI error is generated. The same BFD template can be used for both S-BFD and any other type of BFD session requested by MPLS.

If S-BFD is configured at the LSP level, then sessions are created on all paths of the LSP.

```
config>router>mpls>lsp <name> sr-te
      bfd
        [no] bfd-enable
        [no] bfd-template
        [no] wait-for-up-timer <seconds>
        exit
```

S-BFD can alternatively be configured on the primary or a specific secondary path of the LSP, as follows:

```
config>router>mpls>lsp <name> sr-te
   primary <name>
      bfd
        [no] bfd-enable
        [no] bfd-template <name>
        [no] wait-for-up-timer <seconds>
        exit
```

```
config>router>mpls>lsp <name> sr-te
   secondary <name>
      bfd
        [no] bfd-enable
        [no] bfd-template <name>
        [no] wait-for-up-timer <seconds>
        exit
      standby
```

The wait-for-up-timer is only applicable if failure action is **failover-or-down**. For more information, see Support for BFD Failure Action with SR-TE LSPs.

For PCE-initiated LSPs and SR-TE auto LSPs, S-BFD session parameters are specified in the LSP template. The “to” address that is used for determining the remote discriminator is derived from the far end address of the auto LSP or PCE-initiated LSP.

```
config>router>mpls
   lsp-template <name> pce-init-p2p-sr-te <default | 1...4294967295>
      bfd
        [no] bfd-enable
        [no] bfd-template
        [no] wait-for-up-timer <seconds>
```

```
config>router>mpls
   lsp-template <name> mesh-sr-te <1...4294967295>
      bfd
        [no] bfd-enable
        [no] bfd-template
        [no] wait-for-up-timer <seconds>
```

```
config>router>mpls
   lsp-template <name> p2p-sr-te <1...4294967295>
      bfd
        [no] bfd-enable
        [no] bfd-template
        [no] wait-for-up-timer <seconds>
```

2.8.8.2 Support for BFD Failure Action with SR-TE LSPs

SR OS supports the configuration of a **failure-action** of type **failover-or-down** for SR-TE LSPs. The **failure-action** command is configured at the LSP level or in the LSP template. It can be configured whether S-BFD is applied at the LSP level or the individual path level.

For LSPs with a primary path and a standby or secondary path and **failure-action** of type **failover-or-down**:

- A path is held in an operationally down state when its S-BFD session is down.
- If all paths are operationally down, then the SR-TE LSP is taken operationally down and a trap is generated.
- If S-BFD is enabled at the LSP or active path level, a switchover from the active path to an available path is triggered on failure of the S-BFD session on the active path (primary or standby).
- If S-BFD is not enabled on the active path, and this path is shut down, then a switchover is triggered.
- If S-BFD is enabled on the candidate standby or secondary path, then this path is only selected if S-BFD is up.
- An inactive standby path with S-BFD configured is only considered as available to become active if it is not operationally down, for example, its S-BFD session, is up and all other criteria for it to become operational are true. It is held in an inactive state if the S-BFD session is down.
- The system does not revert to the primary path, nor start a reversion timer when the primary path is either administratively down or operationally down, because the S-BFD session is not up or down for any other reason.

For LSPs with only one path and **failure-action** of type **failover-or-down**:

- A path is held in an operationally down state when its S-BFD session is down.
- If the path is operationally down, then the LSP is taken operationally down and a trap is generated.

Note: S-BFD and other OAM packets can still be sent on an operationally down SR-TE LSP.

2.8.8.2.1 SR-TE LSP State Changes and Failure Actions Based on S-BFD

A path is first configured with S-BFD. This path is held operationally down and not added to the TTM until BFD comes up (subject to the BFD wait time).

The BFD **wait-for-up-timer** provides a mechanism that cleans up the LSP path state at the head end in both cases where S-BFD does not come up in the first place, and where S-BFD goes from up to down. This timer is started when BFD is first enabled on a path or an existing S-BFD session transitions from up to down. When this timer expires and if S-BFD is not up, the path is torn down by removing it from the TTM and the IOM and the LSP retry timer is started.

In the S-BFD up to down case, if there is only one path, the LSP is removed immediately from the TTM when S-BFD fails, and then deprogrammed when the **wait-for-up-timer** expires.

If all the paths of an LSP are operationally down due to S-BFD, then the LSP is taken operationally down and removed from the TTM and the BFD **wait-for-up-timer** is started for each path. If one or more paths do not have S-BFD configured on them, or are otherwise not down, then the LSP is not taken operationally down.

When an existing S-BFD session fails on a path and the failure action is **failover-or-down**, the path is put into the operationally down state. This state and reason code are displayed in a **show>router>bfd>seamless-bfd** command and a trap is raised. The configured failure action is then enacted.

2.8.8.3 S-BFD Operational Considerations

A minimum control packet timer transmit interval of 10 ms can be configured. To maximize the reliability of S-BFD connectivity checking in scaled scenarios with short timers, cases where BFD can go down due to normal changes of the next hop of an LSP path at the head end must be avoided. It is therefore recommended that LFA is not configured at the head end LER when using S-BFD with sub-second timers. When the LFA is not configured, protection of the SR-TE LSP is still provided end-to-end by the combination of S-BFD connectivity checking and primary or secondary path protection.

Similar to the case of LDP and RSVP, S-BFD uses a single path for a loose hop; multiple S-BFD sessions for each of the ECMP paths or spraying of S-BFD packets across the paths is not supported. S-BFD is not down until all the ECMP paths of the loose hop go down.

Note: With very short control packet timer values in scaled scenarios, S-BFD may bounce if the next-hop that the path is currently using goes down because it takes a finite time for BFD to be updated to use another next-hop in the ECMP set.

2.8.9 Static Route Resolution using SR-TE LSP

The user can forward packets of a static route to an indirect next-hop over an SR-TE LSP programmed in TTM by configuring the following static route tunnel binding command:

CLI Syntax:

```
config>router>static-route-entry {ip-prefix/prefix-
 length} [mcast] indirect {ip-address} tunnel-next-hop
  resolution {any | disabled | filter}
  resolution-filter
    [no] sr-te
      [no] [lsp name1]
      [no] [lsp name2]
      .
      .
      [no] [lsp name-N]
    exit
  [no] disallow-igp
  exit
exit
```

The user can select the **sr-te** tunnel type and either specify a list of SR-TE LSP names to use to reach the indirect next-hop of this static route or have the SR-TE LSPs automatically select the indirect next-hop in TTM.

2.8.10 BGP Shortcuts Using SR-TE LSP

The user can forward packets of BGP prefixes over an SR-TE LSP programmed in TTM by configuring the following BGP shortcut tunnel binding command:

CLI Syntax:

```
config>router>bgp>next-hop-resolution
  shortcut-tunnel
    [no] family {ipv4}
      resolution {any | disabled | filter}
      resolution-filter
        [no] sr-te
      exit
    exit
  exit
```

2.8.11 BGP Label Route Resolution Using SR-TE LSP

The user can enable SR-TE LSP, as programmed in TTM, for resolving the next-hop of a BGP IPv4 or IPv6 (6PE) label route by enabling the following BGP transport tunnel command:

CLI Syntax:

```
config>router>bgp>next-hop-res>
  labeled-routes
    transport-tunnel
      [no] family {label-ipv4 | label-ipv6 | vpn}
        resolution {any | disabled | filter}
        resolution-filter
          [no] sr-te
        exit
      exit
    exit
```

2.8.12 Service Packet Forwarding using SR-TE LSP

An SDP sub-type of the MPLS encapsulation type allows service binding to a SR-TE LSP programmed in TTM by MPLS:

Example:

```
*A:7950 XRS-20# configure service sdp 100 mpls create
*A:7950 XRS-20>config>service>sdp$ sr-te-lsp lsp-name
```

The user can specify up to 16 SR-TE LSP names. The destination address of all LSPs must match that of the SDP far-end option. Service data packets are sprayed over the set of LSPs in the SDP using the same procedures as for tunnel selection in ECMP. Each SR-TE LSP can, however, have up to 32 next-hops at the ingress LER when the first segment is a node SID-based SR tunnel. Consequently, service data packet will be forwarded over one of a maximum of 16x32 next-hops. The **tunnel-far-end** option is not supported. In addition, the **mixed-lsp-mode** option does not support the **sr-te** tunnel type.

The signaling protocol for the service labels for an SDP using a SR-TE LSP can be configured to static (**off**), T-LDP (**tldp**), or BGP (**bgp**).

An SR-TE LSP can be used in VPRN auto-bind with the following commands:

CLI Syntax:

```
config>service>vprn>
  auto-bind-tunnel
    resolution {any | disabled | filter}
    resolution-filter
      [no] sr-te
      exit
   exit
```

Both VPN-IPv4 and VPN-IPv6 (6VPE) are supported in a VPRN service using segment routing transport tunnels with the **auto-bind-tunnel** command.

This **auto-bind-tunnel** command is also supported with BGP EVPN service, as shown below:

CLI Syntax:

```
config>service>vpls>bgp-evpn>mpls>
   auto-bind-tunnel
      resolution {any | disabled | filter}
      resolution-filter
         [no] sr-te
      exit
   exit
```

The following service contexts are supported with SR-TE LSP:

- VLL, LDP VPLS, IES/VPRN spoke-interface, R-VPLS, BGP EVPN
- BGP-AD VPLS, BGP-VPLS, BGP VPWS when the **use-provisioned-sdp** option is enabled in the binding to the PW template
- intra-AS BGP VPRN for VPN-IPv4 and VPN-IPv6 prefixes with both auto-bind and explicit SDP
- inter-AS options B and C for VPN-IPv4 and VPN-IPv6 VPRN prefix resolution
- IPv4 BGP shortcut and IPv4 BGP label route resolution
- IPv4 static route resolution
- multicast over IES/VPRN spoke interface with spoke SDP riding a SR-TE LSP

2.8.13 Data Path Support

The support of SR-TE in the data path requires that the ingress LER pushes a label stack where each label represents a hop, a TE link, or a node, in the ERO for the LSP path computed by the router or the PCE. However, only the label and the outgoing interface to the first strict/loose hop in the ERO factor into the forwarding decision of the ingress LER. In other words, the SR-TE LSP only needs to track the reachability of the first strict/loose hop.

This actually represents the NHLFE of the SR shortest path tunnel to the first strict/loose hop. SR OS keeps the SR shortest path tunnel to a downstream node SID or adjacency SID in the tunnel table and so its NHLFE is readily available. The rest of the label stack is not meaningful to the forwarding decision. In this document, “super NHLFE” refers to this part of the label stack because it can have a much larger size.

As a result, an SR-TE LSP is modeled in the ingress LER data path as a hierarchical LSP with the super NHLFE is tunneled over the NHLFE of the SR shortest path tunnel to the first strict/loose hop in the SR-TE LSP path ERO.

Some characteristics of this design are as follows:

- The design saves on NHLFE usage. When many SR TE LSPs are going to the same first hop, they are riding the same SR shortest path tunnel, and will consume each one super NHLFE but they are pointing to a single NHLFE, or set of NHLFEs when ECMP exists for the first strict/loose hop, of the first hop SR tunnel.

  Also, the ingress LER does not need to program a separate backup super NHLFE. Instead, the single super NHLFE will automatically begin forwarding packets over the LFA backup path of the SR tunnel to the first hop as soon as the SR tunnel LFA backup path is activated.
- When the path of a SR-TE LSP contains a maximum of two SIDs, that is the destination SID and one additional loose or strict-hop SID, the SR-TE LSP will use a hierarchy consisting of a regular NHLFE pointing to the NHLFE of top SID corresponding to the first loose or strict hop.
- If the first segment is a node SID tunnel and multiple next-hops exist, then ECMP spraying is supported at the ingress LER.
- If the first hop SR tunnel, node or adjacency SID, goes down the SR module informs MPLS that outer tunnel down and MPLS brings the SR-TE LSP down and requests SR to delete the SR-TE LSP in IOM.

The data path behavior at LSR and egress LER for an SR-TE LSP is similar to that of shortest path tunnel because there is no tunnel state in these nodes. The forwarding of the packet is based on processing the incoming label stack consisting of a node SID and/or adjacency SID label. If the ILM is for a node SID and multiple next-hops exist, then ECMP spraying is supported at the LSR.

The link-protect LFA backup next-hop for an adjacency SID can be programmed at the ingress LER and LSR nodes (as explained in SR-TE LSP Protection).

A maximum of 12 labels, including all transport, service, hash, and OAM labels, can be pushed. The label stack size for the SR-TE LSP can be 1 to 11 labels, with a default value of 6.

The maximum value of 11 is obtained for an SR-TE LSP whose path is not protected via FRR backup and with no entropy or hash label feature enabled when such an LSP is used as a shortcut for an IGP IPv4/IPv6 prefix or as a shortcut for BGP IPv4/IPv6. In this case, the IPv6 prefix requires pushing the IPv6 explicit-null label at the bottom of the stack. This leaves 11 labels for the SR-TE LSP.

The default value of 6 is obtained in the worst cases, such as forwarding a vprn-ping packet for an inter-AS VPN-IP prefix in Option C:

6 SR-TE labels + 1 remote LFA SR label + BGP 3107 label + ELI (RFC 6790) + EL (entropy label) + service label + OAM Router Alert label = 12 labels.

The label stack size manipulation includes the following LER and LSR roles:

LER role:

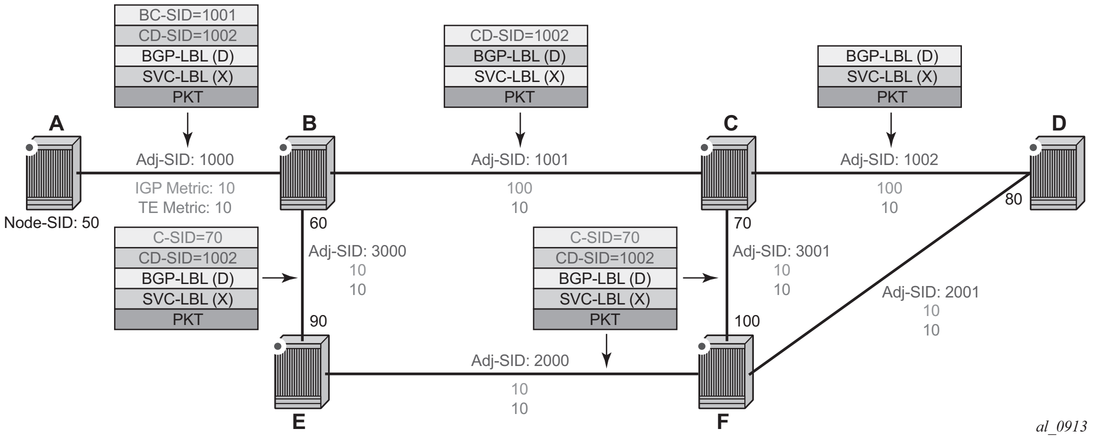
- Push up to 12 labels.
- Pop up to 8 labels of which 4 labels can be transport labels

LSR role:

- Pop up to 5 labels and swap one label for a total of 6 labels
- LSR hash of a packet with up to 16 labels

An example of the label stack pushed by the ingress LER and by a LSR acting as a PLR is illustrated in Figure 40.

Figure 40 **SR-TE LSP Label Stack Programming**



On node A, the user configures an SR-TE LSP to node D with a list of explicit strict hops mapping to the adjacency SID of links: A-B, B-C, and C-D.

Ingress LER A programs a super NHLFE consisting of the label for the adjacency over link C-D and points it to the already-programmed NHLFE of the SR tunnel of its local adjacency over link A-B. The latter NHLFE has the top label and also the outgoing interface to send the packet to.

Note: SR-TE LSP does not consume a separate backup super NHLFE; it only points the single super NHLFE to the NHLFE of the SR shortest path tunnel it is riding. When the latter activates its backup NHLFE, the SR-TE LSP will automatically forward over it.

LSR Node B already programmed the primary NHLFE for the adjacency SID over link C-D and has the ILM with label 1001 point to it. In addition, node B will pre-program the link-protect LFA backup next-hop for link B-C and point the same ILM to it.

Note: There is no super NHLFE at node B as it only deals with the programming of the ILM and primary/backup NHLFE of its adjacency SIDs and its local and remote node SIDs.

VPRN service in node A forwards a packet to the VPN-IPv4 prefix X advertised by BGP peer D. Figure 40 shows the resulting data path at each node for the primary path and for the FRR backup path at LSR B.

2.8.13.1 SR-TE LSP Metric and MTU Settings

The MPLS module assigns a TE-LSP the maximum LSP metric value of 16777215 when the local router provides the hop-to-label translation for its path. For a TE-LSP that uses the local CSPF or the PCE for path computation (**path-computation-method pce** option enabled) by PCE and/or which has its control delegated to PCE (**pce-control** enabled), the latter will return the computed LSP IGP or TE metric in the PCReq and PCUpd messages. In both cases, the user can override the returned value by configuring an admin metric using the command **config**>**router**>**mpls**>**lsp**>**metric**.

1. The MTU setting of a SR-TE LSP is derived from the MTU of the outgoing SR shortest path tunnel it is riding, adjusted with the size of the super NHLFE label stack size.

   The following are the details of this calculation:

   SR_Tunnel_MTU = MIN {Cfg_SR_MTU, IGP_Tunnel_MTU- (1+ *frr-overhead***)*4}**

   Where:

- **Cfg_SR_MTU** is the MTU configured by the user for all SR tunnels within a given IGP instance using **config**>**router**>**ospf/isis**>**segment-routing**>**tunnel-mtu**. If no value was configured by the user, the SR tunnel MTU is fully determined by the IGP interface calculation (explained below).
- **IGP_Tunnel_MTU** is the minimum of the IS-IS or OSPF interface MTU among all the ECMP paths or among the primary and LFA backup paths of this SR tunnel.
- *frr-overhead* is set to:
    - value of **ti-lfa** [**max-sr-frr-labels** *labels*] if **loopfree-alternates** and **ti-lfa** are enabled in this IGP instance
    - 1 if **loopfree-alternates** and **remote-lfa** are enabled but **ti-lfa** is disabled in this IGP instance
    - 0 for all other cases

This calculation is performed by IGP and passed to the SR module each time it changes due to an updated resolution of the node SID.

SR OS also provides the MTU for adjacency SID tunnel because it is needed in a SR-TE LSP if the first hop in the ERO is an adjacency SID. In that case, this calculation for SR_Tunnel_MTU, initially introduced for a node SID tunnel, is applied to get the MTU of the adjacency SID tunnel.

2. The MTU of the SR-TE LSP is derived as follows:

SRTE_LSP_MTU= SR_Tunnel_MTU- numLabels*4

Where:

- **SR_Tunnel_MTU** is the MTU SR tunnel shortest path the SR-TE LSP is riding. The formula is as given above.
- **numLabels** is the number of labels found in the super NHLFE of the SR-TE LSP. Note that at LER, the super NHLFE is pointing to the SR tunnel NHLFE, which itself has a primary and a backup NHLFEs.

This calculation is performed by the SR module and is updated each time the SR-TE LSP path changes or the SR tunnel it is riding is updated.

Note: The above calculated SR-TE LSP MTU is used for the determination of an SDP MTU and for checking the Layer 2 service MTU. For the purpose of fragmentation of IP packets forwarded in GRT or in a VPRN over a SR-TE LSP, the data path always deducts the worst case MTU (12 labels) from the outgoing interface MTU for the decision to fragment or not the packet. In this case, the above formula is not used.

2.8.13.2 LSR Hashing on SR-TE LSPs

The LSR supports hashing up to a maximum of 16 labels in a stack. The LSR is able to hash on the IP headers when the payload below the label stack is IPv4 or IPv6, including when a MAC header precedes it (**ethencap-ip** option). Alternatively, it is able to hash based only on the labels in the stack, which may include the entropy label (EL) or the hash label. See the MPLS Entropy Label and Hash Label section for more information about the hash label and entropy label features.

When the hash-label option is enabled in a service context, a hash label is always inserted at the bottom of the stack as per RFC 6391.

The EL feature, as specified in RFC 6790, indicates the presence of a flow on an LSP that should not be reordered during load balancing. It can be used by an LSR as input to the hash algorithm. The Entropy Label Indicator (ELI) is used to indicate the presence of the EL in the label stack. The ELI, followed by the actual EL, is inserted immediately below the transport label for which the EL feature is enabled. If multiple transport tunnels have the EL feature enabled, the ELI and EL are inserted below the lowest transport label in the stack.

The EL feature is supported with an SR-TE LSP. See the MPLS Entropy Label and Hash Label section for more information.

The LSR hashing operates as follows:

- If the **lbl-only** hashing option is enabled, or if one of the other LSR hashing options is enabled but an IPv4 or IPv6 header is not detected below the bottom of the label stack, the LSR parses the label stack and hashes only on the EL or hash label.
- If the **lbl-ip** option is enabled, the LSR parses the label stack and hashes on the EL or hash label and the IP headers.
- If the **ip-only** or **eth-encap-ip** is enabled, the LSR hashes on the IP headers only.

2.8.14 SR-TE Auto-LSP

The SR-TE auto-LSP feature allows the auto-creation of an SR-TE mesh LSP and for an SR-TE one-hop LSP.

The SR-TE mesh LSP feature specifically binds an LSP template of a new type, **mesh-p2p-srte**, with one more prefix list. When the TE database discovers a router, which has a router ID matching an entry in the prefix list, it triggers MPLS to instantiate an SR-TE LSP to that router using the LSP parameters in the LSP template.

The SR-TE one-hop LSP feature specifically activates an LSP template of a new type, **one-hop-p2p-srte**. In this case, the TE database keeps track of each TE link which comes up to a directly connected IGP neighbor. It then instructs MPLS to instantiate an SR-TE LSP with the following parameters:

- the source address of the local router
- an outgoing interface matching the interface index of the TE-link
- a destination address matching the router-id of the neighbor on the TE link

In both types of SR-TE auto-LSP, the router’s hop-to-label translation or local CSPF computes the label stack required to instantiate the LSP path.

➡ **Note:** An SR-TE auto-LSP can be reported to a PCE but cannot be delegated or have its paths computed by PCE.

2.8.14.1 Feature Configuration

This feature introduces two new LSP template types: **one-hop-p2p-srte** and **mesh-p2p-srte**. The configuration for these commands is the same as that of the RSVP-TE auto-lsp of type **one-hop-p2p** and **mesh-p2p** respectively.

The user first creates an LSP template of the one of the following types:

- **config**>**router**>**mpls**>**lsp-template** *template-name* **mesh-p2p-srte**
- **config**>**router**>**mpls**>**lsp-template** *template-name* **one-hop-p2p-srte**

In the template, the user configures the common LSP and path level parameters or options shared by all LSPs using this template.

These new types of LSP templates contain the SR-TE LSP-specific commands as well as all other LSP or path commands common to RSVP-TE LSP and SR-TE LSP, and which are supported by the existing RSVP-TE LSP template.

Next, the user either binds the LSP template of type **mesh-p2p-srte** with one or more prefix lists using the **config**>**router**>**mpls**>**lsp-template** *template-name* **policy** *peer-prefix-policy1* [*peer-prefix-policy2*] command, or binds the LSP template of type **one-hop-p2p-srte** with the **one-hop** option using the **config**>**router**>**mpls**>**lsp-template** *template-name* **one-hop** command.

See Configuring and Operating SR-TE for an example configuration of the SR-TE auto-LSP creation using an LSP template of type **mesh-p2p-srte**.

2.8.14.2 Automatic Creation of an SR-TE Mesh LSP

This feature behaves the same way as the RSVP-TE auto-LSP using an LSP template of type **mesh-p2p**.

The **auto-lsp** command binds an LSP template of type **mesh-p2p-srte** with one or more prefix lists. When the TE database discovers a router that has a router ID matching an entry in the prefix list, it triggers MPLS to instantiate an SR-TE LSP to that router using the LSP parameters in the LSP template.

The prefix match can be exact or longest. Prefixes in the prefix list that do not correspond to a router ID of a destination node will never match.

The path of the LSP is that of the default path name specified in the LSP template. The hop-to-label translation tool or the local CSPF determines the node SID and adjacency SID corresponding to each loose and strict hop in the default path definition respectively.

The LSP has an auto-generated name using the following structure:

"*TemplateName-DestIpv4Address-TunnelId*"

where:

- *TemplateName* = the name of the template
- *DestIpv4Address* = the address of the destination of the auto-created LSP
- *TunnelId* = the TTM tunnel ID

In SR OS, an SR-TE LSP uses three different identifiers:

- LSP Index is used for indexing the LSP in the MIB table shared with RSVP-TE LSP. Range:
  - provisioned SR-TE LSP: 65536 to 81920
  - SR-TE auto-LSP: 81921 to 131070

- LSP Tunnel Id is used in the interaction with PCC/PCE. Range: 1 to 65536
- TTM Tunnel Id is the tunnel-ID service, shortcut, and steering applications use to bind to the LSP. Range: 655362 to 720897

The path name is that of the default path specified in the LSP template.

Note: This feature is limited to SR-TE LSP, that is controlled by the router (PCC-controlled) and which path is provided using the hop-to-label translation or the local CSPF path computation method.

2.8.14.3 Automatic Creation of an SR-TE One-Hop LSP

This feature like the RSVP-TE auto-LSP using an LSP template of **one-hop-p2p** type. Although the provisioning model and CLI syntax differ from that of a mesh LSP by the absence of a prefix list, the actual behavior is quite different. When the **one-hop-p2p** command is executed, the TE database keeps track of each TE link that comes up to a directly connected IGP neighbor. It then instructs MPLS to instantiate an SR-TE LSP with the following parameters:

- the source address of the local router
- an outgoing interface matching the interface index of the TE link
- a destination address matching the router ID of the neighbor on the TE link

In this case, the hop-to-label translation or the local CSPF returns the SID for the adjacency to the remote address of the neighbor on this link. Therefore, the **auto-lsp** command binding an LSP template of type **one-hop-p2p-srte** with the **one-hop** option results in one SR-TE LSP instantiated to the IGP neighbor for each adjacency over any interface.

Because the local router installs the adjacency SID to a link independent of whether the neighbor is SR-capable, the TE-DB finds the adjacency SID and a one-hop SR-TE LSP can still come up to such a neighbor. However, remote LFA using the neighbor’s node SID will not protect the adjacency SID and so, will also not protect the one-hop SR-TE LSP because the node SID is not advertised by the neighbor.

The LSP has an auto-generated name using the following structure:

“*TemplateName-DestIpv4Address-TunnelId*”

where:

- *TemplateName* = the name of the template
- *DestIpv4Address* = the address of the destination of the auto-created LSP

- *TunnelId* = the TTM tunnel ID.

The path name is that of the default path specified in the LSP template.

Note: This feature is limited to an SR-TE LSP that is controlled by the router (PCC-controlled) and for which labels for the path are provided by the hop-to-label translation or the local CSPF path computation method.

2.8.14.4 Interaction with PCEP

A template-based SR-TE auto-LSP can only be operated as a PCC-controlled LSP. It can, however, be reported to the PCE using the **pce-report** command. It cannot be operated as a PCE-computed or PCE-controlled LSP. This is the same interaction with PCEP as that of a template-based RSVP-TE LSP.

2.8.14.5 Forwarding Contexts Supported with SR-TE Auto-LSP

The following are the forwarding contexts that can be used by an auto-LSP:

- resolution of IPv4 BGP label routes and IPv6 BGP label routes (6PE) in TTM
- resolution of IPv4 BGP route in TTM (BGP shortcut)
- resolution of IPv4 static route to indirect next-hop in TTM
- VPRN and BGP-EVPN auto-bind for both IPv4 and IPv6 prefixes

The auto-LSP is, however, not available to be used in a provisioned SDP for explicit binding by services. Therefore, an auto-LSP can also not be used directly for auto-binding of a PW template with the **use-provisioned-sdp** option in BGP-AD VPLS or FEC129 VLL service. However, an auto-binding of a PW template to an LDP LSP, which is then tunneled over an SR-TE auto-LSP is supported.

2.8.15 SR-TE LSP Traffic Statistics

As in RSVP-TE LSPs, it is possible to enable the collection of traffic statistics on SR-TE LSPs (using either a named LSP or SR-TE templates). However, traffic statistics are only available on egress or ingress LER. Also, traffic statistics cannot be recorded into an accounting file.

Unlike RSVP-TE LSP statistics, SR-TE LSP statistics are provided without any forwarding class or QoS profile distinction. However, traffic statistics are recorded and made available for each of the paths of the LSP (primary and backup). Statistic indexes are only allocated at the time the path is effectively programmed, are maintained across switch-over for primary and standbys only, and are released if egress statistics are disabled or the LSP is deleted.

Note: SR-TE LSP egress statistics are not supported on VSR.

2.8.16 SR-TE Label Stack Checks

2.8.16.1 Service and Shortcut Application SR-TE Label Stack Check

If a packet forwarded in a service or a shortcut application has resulted in the net label stack size being pushed on the packet to exceed the maximum label stack supported by the router, the packet is dropped on the egress. Each service and shortcut application on the router performs a check of the resulting net label stack after pushing all the labels required for forwarding the packet in that context.

To that effect, the MPLS module populates each SR-TE LSP in the TTM with the maximum transport label stack size, which consists of the sum of the values in **max-sr-labels** *label-stack-size* and **additional-frr-labels** *labels*.

Each service or shortcut application will then add the additional, context-specific labels, such as service label, entropy/hash label, and control-word, required to forward the packet in that context and to check that the resulting net label stack size does not exceed the maximum label stack supported by the router.

If the check succeeds, the service is bound or the prefix is resolved to the SR-TE LSP.

If the check fails, the service will not bind to this SR-TE LSP. Instead, it will either find another SR-TE LSP or another tunnel of a different type to bind to, if the user has configured the use of other tunnel types. Otherwise, the service will go down. When the service uses a SDP with one or more SR-TE LSP names, the spoke SDP bound to this SDP will remain operationally down as long as at least one SR-TE LSP fails the check. In this case, a new spoke SDP flag is displayed in the **show** output of the service: "labelStackLimitExceeded". Similarly, the prefix will not get resolved to the SR-TE LSP and will either be resolved to another SR-TE LSP or another tunnel type, or will become unresolved.

The value of **additional-frr-labels** *labels* is checked against the maximum value across all IGP instances of the parameter *frr-overhead*. This parameter is computed within a given IGP instance as described in Table 21.

Table 21 **frr-overhead Parameter Values**

Condition	frr-overhead Parameter Value
segment-routing is disabled in the IGP instance	0
segment-routing is enabled but **remote-lfa** is disabled	0
segment-routing is enabled and **remote-lfa** is enabled	1

When the user configures or changes the configuration of **additional-frr-labels**, MPLS ensures that the new value accommodates the *frr-overhead* value across all IGP instances.

Example:

1. The user configures the **config>router>isis>loopfree-alternates remote-lfa** command.
2. The user creates a new SR-TE LSP or changes the configuration of an existing as follows: **mpls>lsp>max-sr-labels** *10* **additional-frr-labels** *0*.
3. Performing a **no shutdown** of the new LSP or changing the existing LSP configuration is blocked because the IS-IS instance enabled remote LFA, which requires one additional label on top of the 10 SR labels of the primary path of the SR-TE LSP.

If the check is successful, MPLS adds **max-sr-labels** and **additional-frr-labels** and checks that the result is lower or equal to the maximum label stack supported by the router. MPLS then populates the value of {**max-sr-labels** + **additional-frr-labels**}, along with tunnel information in TTM, and also passes **max-sr-labels** to the PCEP module.

Conversely, if the user tries a configuration change that results in a change to the computed *frr-overhead*, IGP will check that all SR-TE LSPs can properly account for the overhead or the change is rejected. On the IGP, enabling **remote-lfa** may cause the *frr-overhead* to change.

Example:

- An MPLS LSP is administratively enabled and has **mpls**>**lsp**>**max-sr-labels** *10* **additional-frr-overhead** *0* configured.
- The current configuration in IS-IS has the **loopfree-alternates** command disabled.
- The user attempts to configure

  isis>**loopfree-alternates remote-lfa**. This changes *frr-overhead* to 1.

  This configuration change will be blocked.

2.8.16.2 Control Plane Handling of Egress Label Stack Limitations

As described in Data Path Support, the egress IOM can push a maximum of 12 labels; however, this number may be reduced if other fields are pushed into the packets. For example, for a VPRN service, the ingress LER can send an IP VPN packet with 12 labels in the stack, including one service label, one label for OAM, and 10 transport labels. However, if entropy is configured, the number of transport labels is reduced by two (Entropy Label (EL) and Entropy Label Indicator (ELI)). Similarly, for EVPN services, the egress IOM might push specific fields that reduce the total number of supported transport labels.

To avoid silent packet drops in cases where the egress IOM cannot push the required number of labels, SR OS implements a set of procedures that prevent the system from sending packets if it is determined that the SR-TE label stack to be pushed exceeds the number of bytes that the egress IOM can put on the wire.

Table 22 describes the label stack egress IOM restrictions on FP-based hardware for IPVPN and EVPN services.

Table 22 **Label Stack Egress IOM Restrictions on FP-based Hardware for IPVPN and EVPN Services**

Features that reduce the Label Stack		Source Service Type				
		IP-VPN (VPRN)	EVPN-IFL (VPRN)	EVPN VPLS or EVPN Epipe	EVPN B-VPLS (PBB-EVPN)	EVPN-IFF (R-VPLS)
Always Computed [1]	Service Label	1	1	1	1	1
	OAM Label	1	1	0	0	0
	Control Word	0	0	1	1	1
	ESI Label	0	0	1	0	0
Computed if configured [2]	Hash Label (mutex with EL)	1	1	0	0	0
	Entropy EL+ELI	2	2	2	2	2
Required Labels [3]		2	2	3	2	2
Required Labels + Options [3]		4	4	5	4	4
Maximum available labels [4]		12	12	10	6	9
Maximum available transport labels without options [5]		10	10	7	4	7
Maximum available transport labels with options [5]		8	8	5	2	5

1. These rows indicate the number of labels that the system assumes are always used on a given service. For example, the system always computes two labels to be reduced from the total number of labels for VPRN services with EVPN-IFL (EVPN Interface-less model enabled).
2. These rows indicate the number of labels that the system subtracts from the total only if they are configured on the service. For example, on VPRN services with EVPN-IFL, if the user configures hash-label, the system computes one additional label. If the user configures entropy-label, the system deducts two labels instead.
3. These rows indicate the number of labels that the system deducts from the total number.

4. This row indicates a different number depending on the service type and the inner encapsulation used by each service, which reduces the maximum number of labels to push on egress. For example, while the number of labels for VPRN services is 12, the maximum number for VPLS and Epipe services is 10 (to account for space for an inner Ethernet header).
5. This row indicates the maximum SR-TE labels that the system can push when sending service packets on the wire.

The total number of labels configured in the command **max-sr-labels** *label-stack-size* [**additional-frr-labels** *labels*] must not exceed the labels indicated in the "Maximum available transport labels with/without options" rows in Table 22. If the configured LSP labels exceed the available labels in the table, the BGP route next hop for the LSP is not resolved and the system does not even try to send packets to that LSP.

For example, for a VPRN service with EVPN-IFL where the user configures entropy-label, the maximum available transport labels is eight. If an IP Prefix route for next-hop X is received for the service and the SR-TE LSP to-X is the best tunnel to reach X, the system checks that (**max-sr-labels** + **additional-frr-labels**) is less than or equal to eight. Otherwise, the IP Prefix route is not resolved.

The same control plane check is performed for other service types, including IP shortcuts, spoke SDPs on IP interfaces, spoke SDPs on Epipes, VPLS, B-VPLS, R-VPLS, and R-VPLS in I-VPLS or PW-SAP. In all cases, the spoke SDP is brought down if the configured (**max-sr-labels** + **additional-frr-labels**) is greater than the maximum available transport labels. Table 23 indicates the maximum available transport labels for IP shortcuts and spoke SDP services.

Note: For PW-SAPs, the maximum available labels differ depending on the type of service PW-SAP used (Epipe or VPRN interface).

Table 23 **Maximum Available Transport Labels for IP Shortcuts and Spoke SDP services**

Features that reduce the Label Stack		**Source Service Type**							
		IP Shortcuts	**Spoke-sdp Interface**	**Spoke-sdp Epipe**	**Spoke-sdp VPLS**	**Spoke-sdp B-VPLS**	**Spoke-sdp R-VPLS**	**Spoke-sdp R-VPLS I-VPLS**	**PW-SAP Epipe/ Interface**
Always Computed [1]	Service Label	0	1	1	1	1	1	1	1/1
	OAM Label	0	1	1	1	1	0	0	0/0
	IPv6 label	1	0	0	0	0	0	0	0/0
Computed if configured [2]	Hash Label (mutex with EL)	0	1	1	1	1	1	0	0/0
	Entropy EL+ELI	2	2	2	2	2	2	2	0/0
	Control Word	0	1	1	1	1	1	1	0/0
Required Labels [3]		1	2	2	2	2	1	1	1/1
Required Labels + Options [3]		3	5	5	5	5	4	4	1/1
Maximum available labels [4]		12	9	10	10	6	8	4	10/7
Maximum available transport labels without options [5]		11	7	8	8	4	7	3	9/6
Maximum available transport labels with options [5]		9	4	5	5	1	4	0	9/6

1. Indicates the number of labels that the system assumes are always used on a specific service
2. Indicates the number of labels that the system subtracts from the total only if they are configured on the service
3. Number of labels that the system deducts from the total number

4. Indicates a different number depending on the service type and the inner encapsulation used by each service, which reduces the maximum number of labels to push on egress
5. Maximum SR-TE labels that the system can push when sending service packets on the wire

In general, the labels shown in Table 22 and Table 23 are valid for network ports that are null or dot1q encapsulated. For QinQ network ports, the available labels are deducted by one.

2.9 MPLS Service Usage

Nokia routers enable service providers to deliver VPNs and Internet access using Generic Routing Encapsulation (GRE) and/or MPLS tunnels, with Ethernet interfaces and/or SONET/SDH (on the 7750 SR and 7450 ESS) interfaces.

2.9.1 Service Distribution Paths

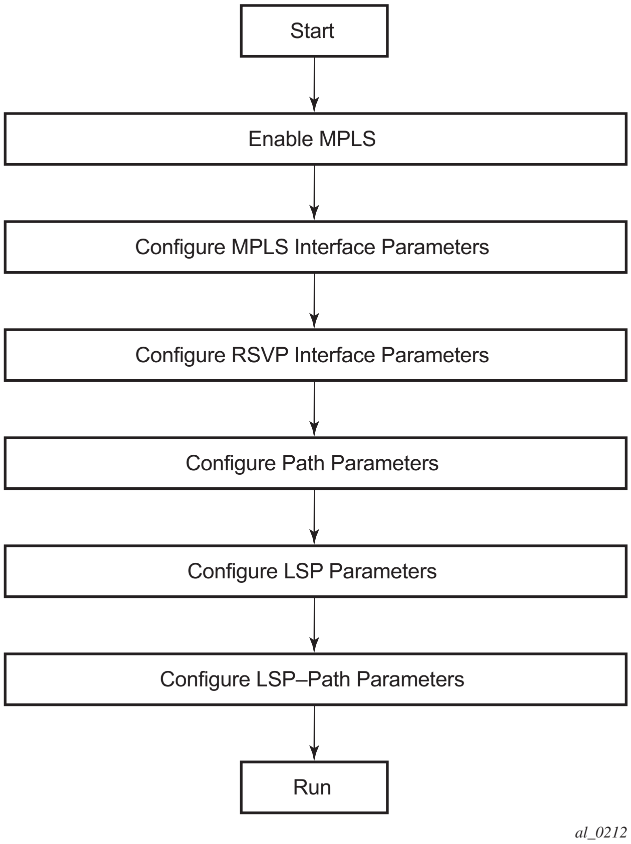
A service distribution path (SDP) acts as a logical way of directing traffic from one router to another through a uni-directional (one-way) service tunnel. The SDP terminates at the far-end router which directs packets to the correct service egress service access point (SAP) on that device. All services mapped to an SDP use the same transport encapsulation type defined for the SDP (either GRE or MPLS).

For information about service transport tunnels, refer to the Service Distribution Paths (SDPs) section in the *7450 ESS, 7750 SR, 7950 XRS, and VSR Services Overview Guide*. They can support up to eight forwarding classes and can be used by multiple services. Multiple LSPs with the same destination can be used to load-balance traffic.

2.10 MPLS/RSVP Configuration Process Overview

Figure 41 displays the process to configure MPLS and RSVP parameters.

Figure 41 **MPLS and RSVP Configuration and Implementation Flow**


Start
Enable MPLS
Configure MPLS Interface Parameters
Configure RSVP Interface Parameters
Configure Path Parameters
Configure LSP Parameters
Configure LSP–Path Parameters
Run
al_0212

2.11 Configuration Notes

This section describes MPLS and RSVP caveats.

- Interfaces must already be configured in the **config**>**router**>**interface** context before they can be specified in MPLS and RSVP.
- A router interface must be specified in the **config**>**router**>**mpls** context in order to apply it or modify parameters in the **config**>**router**>**rsvp** context.
- A system interface must be configured and specified in the **config**>**router**>**mpls** context.
- Paths must be created before they can be applied to an LSP.

2.12 Configuring MPLS and RSVP with CLI

This section provides information to configure MPLS and RSVP using the command line interface.

2.12.1 MPLS Configuration Overview

Multiprotocol Label Switching (MPLS) enables routers to forward traffic based on a simple label embedded into the packet header. A router examines the label to determine the next hop for the packet, saving time for router address lookups to the next node when forwarding packets. MPLS is not enabled by default and must be explicitly enabled.

In order to implement MPLS, the following entities must be configured:

2.12.1.1 LSPs

To configure MPLS-signaled label-switched paths (LSPs), an LSP must run from an ingress router to an egress router. Configure only the ingress router and configure LSPs to allow the software to make the forwarding decisions or statically configure some or all routers in the path. The LSP is set up by Resource Reservation Protocol (RSVP), through RSVP signaling messages. The router automatically manages label values. Labels that are automatically assigned have values ranging from 1,024 through 1,048,575 (see Label Values).

A static LSP is a manually set up LSP where the nexthop IP address and the outgoing label are explicitly specified.

2.12.1.2 Paths

To configure signaled LSPs, you must first create one or more named paths on the ingress router. For each path, the transit routers (hops) in the path are specified.

2.12.1.3 Router Interface

At least one router interface and one system interface must be defined in the **config**>**router**>**interface** context in order to configure MPLS on an interface.

2.12.1.4 Choosing the Signaling Protocol

In order to configure a static or a RSVP signaled LSP, you must enable MPLS on the router, which automatically enables RSVP and adds the system interface into both contexts. Any other network IP interface, other than loopbacks, added to MPLS is also automatically enabled in RSVP and becomes a TE link. When the interface is enabled in RSVP, the IGP instance will advertise the Traffic Engineering (TE) information for the link to other routers in the network in order to build their TE database and compute CSPF paths. Operators must enable the traffic-engineering option in the ISIS or OSPF instance for this. Operators can also configure under the RSVP context of the interface the RSVP protocol parameters for that interface.

If only static label switched paths are used in your configurations, operators must manually define the paths through the MPLS network. Label mappings and actions configured at each hop must be specified. Operators can disable RSVP on the interface if it is used only for incoming or outgoing static LSP label by shutting down the interface in the RSVP context. The latter causes IGP to withdraw the TE link from its advertisement which removes it from its local and neighbors TE database.

If dynamic LSP signaling is implemented in an operator’s network then they must keep RSVP enabled on the interfaces they want to use for explicitly defined or CSPF calculated LSP path.

2.12.2 Basic MPLS Configuration

This section provides information to configure MPLS and configuration examples of common configuration tasks. To enable MPLS, you must configure at least one MPLS interface. The other MPLS configuration parameters are optional. This follow displays an example of an MPLS configuration.

```
ALA-1>config>router>if-attr# info
----------------------------------------------
admin-group "green" 15
admin-group "yellow" value 20
admin-group "red" value 25
----------------------------------------------
A:ALA-1>config>router>mpls# info
------------------------------------------
```

```
            interface "system"
            exit
            interface "StaticLabelPop"
                admin-group "green"
                label-map 50
                    pop
                    no shutdown
                exit
            exit
            interface "StaticLabelPop"
                label-map 35
                    swap 36 nexthop 10.10.10.91
                    no shutdown
                exit
            exit
            path "secondary-path"
                no shutdown
            exit
            path "to-NYC"
                hop 1 10.10.10.104  strict
                no shutdown
            exit
            lsp "lsp-to-eastcoast"
                to 10.10.10.104
                from 10.10.10.103
                fast-reroute one-to-one
                exit
                primary "to-NYC"
                exit
                secondary "secondary-path"
                exit
                no shutdown
            exit
            static-lsp "StaticLabelPush"
                to 10.10.11.105
                push 60 nexthop 10.10.11.105
                no shutdown
            exit
            no shutdown
----------------------------------------------
A:ALA-1>config>router>mpls#
```

2.12.3 Common Configuration Tasks

This section provides a brief overview of the tasks to configure MPLS and provides the CLI commands.

The following protocols must be enabled on each participating router.

- MPLS
- RSVP (for RSVP-signaled MPLS only), which is automatically enabled when MPLS is enabled.

In order for MPLS to run, you must configure at least one MPLS interface in the **config**>**router**>**mpls** context.

- An interface must be created in the **config**>**router**>**interface** context before it can be applied to MPLS.
- In the **config**>**router**>**mpls** context, configure path parameters for configuring LSP parameters. A path specifies some or all hops from ingress to egress. A path can be used by multiple LSPs.
- When an LSP is created, the egress router must be specified in the **to** command and at least one primary or secondary path must be specified. All other statements under the LSP hierarchy are optional.

2.12.4 Configuring MPLS Components

Use the MPLS and RSVP CLI syntax in the following sections to configure MPLS components.

2.12.4.1 Configuring Global MPLS Parameters

Admin groups can signify link colors, such as red, yellow, or green. MPLS interfaces advertise the link colors it supports. CSPF uses the information when paths are computed for constrained-based LSPs. CSPF must be enabled in order for admin groups to be relevant.

To configure MPLS admin-group parameters, enter the following commands:

CLI Syntax:

```
if-attribute
admin-group group-name value group-value
mpls
frr-object
resignal-timer minutes
```

The following displays an admin group configuration example:

```
ALA-1>config>router>if-attr# info
----------------------------------------------
admin-group "green" value 15
admin-group "yellow" value 20
admin-group "red" value 25
----------------------------------------------
A:ALA-1>config>router>mpls# info
----------------------------------------------
            resignal-timer 500
...
```

```
--------------------------------------------
A:ALA-1>config>router>mpls#
```

2.12.4.2 Configuring an MPLS Interface

Configure the **label-map** parameters if the interface is used in a static LSP. To configure an MPLS interface on a router, enter the following commands:

CLI Syntax:

```
config>router>mpls
interface
   no shutdown
   admin-group group-name [group-name...(up to 32 max)]
   label-map
      pop
      swap
      no shutdown
   srlg-group group-name [group-name...(up to 5 max)]
   te-metric value
```

The following displays an interface configuration example:

```
A:ALA-1>config>router>mpls# info
--------------------------------------------
...
            interface "to-104"
                admin-group "green"
                admin-group "red"
                admin-group "yellow"
                label-map 35
                    swap 36 nexthop 10.10.10.91
                    no shutdown
                exit
            exit
            no shutdown
...
--------------------------------------------
A:ALA-1>config>router>mpls#
```

2.12.4.3 Configuring MPLS Paths

Configure an LSP path to use in MPLS. When configuring an LSP, the IP address of the hops that the LSP should traverse on its way to the egress router must be specified. The intermediate hops must be configured as either **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**).

Use the following CLI syntax to configure a path:

CLI Syntax:

```
config>router> mpls
path path-name
   hop hop-index ip-address {strict | loose}
   no shutdown
```

The following displays a path configuration example:

```
A:ALA-1>config>router>mpls# info
------------------------------------------
            interface "system"
            exit
            path "secondary-path"
                hop 1 10.10.0.121  strict
                hop 2 10.10.0.145  strict
                hop 3 10.10.0.1    strict
                no shutdown
            exit
            path "to-NYC"
                hop 1 10.10.10.103 strict
                hop 2 10.10.0.210  strict
                hop 3 10.10.0.215  loose
            exit
------------------------------------------
A:ALA-1>config>router>mpls#
```

2.12.4.4 Configuring an MPLS LSP

Configure an LSP path for MPLS. When configuring an LSP, you must specify the IP address of the egress router in the **to** statement. Specify the primary path to be used. Secondary paths can be explicitly configured or signaled upon the failure of the primary path. All other statements are optional.

The following displays an MPLS LSP configuration:

```
A:ALA-1>config>router>mplp# info
----------------------------------------------
...
            lsp "lsp-to-eastcoast"
                to 192.168.200.41
                rsvp-resv-style ff
                path-computation-method local-cspf
                include "red"
                exclude "green"
                adspec
                fast-reroute one-to-one
                exit
                primary "to-NYC"
                    hop-limit 10
                exit
                secondary "secondary-path"
                    bandwidth 50000
                exit
                no shutdown
            exit
            no shutdown
----------------------------------------------
A:ALA-1>config>router>mpls#
```

2.12.4.5 Configuring a Static LSP

An LSP can be explicitly (statically) configured. Static LSPs are configured on every node along the path. The label’s forwarding information includes the address of the next hop router.

Use the following CLI syntax to configure a static LSP:

CLI Syntax:

```
config>router>mpls
static-lsp lsp-name
  to ip-address
  push out-label nexthop ip-addr
  no shutdown
```

The following displays a static LSP configuration example:

```
A:ALA-1>config>router>mpls# info
----------------------------------------------
...
            static-lsp "static-LSP"
                to 10.10.10.124
                push 60 nexthop 10.10.42.3
                no shutdown
            exit
...
----------------------------------------------
A:ALA-1>config>router>mpls#
```

2.12.4.6 Configuring Manual Bypass Tunnels

Consider the following network setup.

```
A----B----C----D

     |    |

     E----F
```

The user first configures the option to disable the dynamic bypass tunnels on node B if required. The CLI for this configuration is:

config>router>mpls>dynamic-bypass [disable | enable]

By default, dynamic bypass tunnels are enabled.

Next, the user configures an LSP on node B, such as B-E-F-C to be used only as bypass. The user specifies each hop in the path, for example, the bypass LSP has a strict path.

Note that including the bypass-only keyword disables the following options under the LSP configuration:

- bandwidth
- fast-reroute
- secondary

The following LSP configuration options are allowed:

- adaptive
- adspec
- exclude
- hop-limit
- include
- metric-type
- path-computation-method local-cspf

The following example displays a bypass tunnel configuration:

```
A:ALA-48>config>router>mpls>path# info
-------------------------------------------
...
            path "BEFC"
                hop 10 10.10.10.11  strict
                hop 20 10.10.10.12  strict
                hop 30 10.10.10.13  strict
                no shutdown
            exit


            lsp "bypass-BC"
                to 10.10.10.15
                primary "BEFC"
                exit
                no shutdown
...
-------------------------------------------
A:ALA-48>config>router>mpls>path#
```

Next, the user configures an LSP from A to D and indicates fast-reroute bypass protection by selecting facility as the FRR method (**config>router>mpls>lsp>fast-reroute facility**). If the LSP goes through B, and bypass is requested, and the next hop is C, and there is a manually configured bypass-only tunnel from B to C, excluding link BC, then node B uses that.

2.12.4.7 Configuring RSVP Parameters

RSVP is used to set up LSPs. RSVP must be enabled on the router interfaces that are participating in signaled LSPs. The **keep-multiplier** and **refresh-time** default values can be modified in the RSVP context.

Initially, interfaces are configured in the **config>router>mpls>interface** context. Only these existing (MPLS) interfaces are available to modify in the **config>router>rsvp** context. Interfaces cannot be directly added in the RSVP context.

The following example displays an RSVP configuration example:

```
A:ALA-1>config>router>rsvp# info
----------------------------------------------
interface "system"
             no shutdown
          exit
          interface to-104
             hello-interval 4000
             no shutdown
          exit
          no shutdown
----------------------------------------------
A:ALA-1>config>router>rsvp#
```

2.12.4.8 Configure RSVP Message Pacing Parameters

RSVP message pacing maintains a count of the messages that were dropped because the output queue for the egress interface was full.

Use the following CLI syntax to configure RSVP parameters:

CLI Syntax:

```
config>router>rsvp
no shutdown
msg-pacing
   period milli-seconds
   max-burst number
```

The following example displays a RSVP message pacing configuration example:

```
A:ALA-1>config>router>rsvp# info
----------------------------------------------
            keep-multiplier 5
            refresh-time 60
            msg-pacing
                period 400
                max-burst 400
            exit
            interface "system"
                no shutdown
            exit
            interface to-104
                hello-interval 4000
                no shutdown
            exit
            no shutdown
----------------------------------------------
A:ALA-1>config>router>rsvp#
```

2.12.4.9 Configuring Graceful Shutdown

TE graceful shutdown can be enabled on a specific interface using the **config**>**router**>**rsvp**>**if**>**graceful-shutdown** command. This interface is referred to as the maintenance interface.

Graceful shutdown can be disabled by executing the **no** form of the command at the RSVP interface level or at the RSVP level. In this case, the user configured TE parameters of the maintenance links are restored and the maintenance node floods them.

2.13 MPLS Configuration Management Tasks

This section discusses MPLS configuration management tasks.

2.13.1 Deleting MPLS

NOTE: In order to remove the MPLS instance, MPLS must be disabled (shutdown) and all SDP bindings to LSPs removed. If MPLS is not shutdown first, when the **no mpls** command is executed, a warning message on the console displays indicating that MPLS is still administratively up.

When MPLS is shut down, the **no mpls** command deletes the protocol instance and removes all configuration parameters for the MPLS instance.
To disable MPLS, use the **shutdown** command.

To remove MPLS on a router, enter the following command:

CLI Syntax: `config>router# no mpls`

2.13.2 Modifying MPLS Parameters

Note: You must shut down MPLS entities in order to modify parameters. Re-enable (**no shutdown**) the entity for the change to take effect.

2.13.3 Modifying an MPLS LSP

Some MPLS LSP parameters such as primary and secondary, must be shut down before they can be edited or deleted from the configuration.

The following displays a MPLS LSP configuration example. Refer to the LSP configuration in Configuring an MPLS LSP.

```
A:ALA-1>>config>router>mpls>lsp# info
-----------------------------------------------
                shutdown
                to 10.10.10.104
                from 10.10.10.103
                rsvp-resv-style ff
                include "red"
                exclude "green"
                fast-reroute one-to-one
                exit
                primary "to-NYC"
                    hop-limit 50
                exit
                secondary "secondary-path"
                exit
----------------------------------------------
A:ALA-1>config>router>mpls#
```

2.13.4 Modifying MPLS Path Parameters

In order to modify path parameters, the **config>router>mpls>path** context must be shut down first.

The following displays a path configuration example. Refer to Configuring MPLS Paths.

```
A:ALA-1>config>router>mpls# info
#------------------------------------------
echo "MPLS"
#------------------------------------------
...
            path "secondary-path"
                hop 1 10.10.0.111  strict
                hop 2 10.10.0.222  strict
                hop 3 10.10.0.123  strict
                no shutdown
            exit
            path "to-NYC"
                hop 1 10.10.10.104 strict
                hop 2 10.10.0.210  strict
                no shutdown
            exit
...
----------------------------------------------
A:ALA-1>config>router>mpls#
```

2.13.5 Modifying MPLS Static LSP Parameters

In order to modify static LSP parameters, the **config>router>mpls>path** context must be shut down first.

The following displays a static LSP configuration example. Refer to the static LSP configuration in Configuring a Static LSP.

```
A:ALA-1>config>router>mpls# info
```

```
----------------------------------------------
...
            static-lsp "static-LSP"
                to 10.10.10.234
                push 102704 nexthop 10.10.8.114
                no shutdown
            exit
            no shutdown
----------------------------------------------
A:ALA-1>config>router>mpls#
```

2.13.6 Deleting an MPLS Interface

In order to delete an interface from the MPLS configuration, the interface must be shut down first.

Use the following CLI syntax to delete an interface from the MPLS configuration:

CLI Syntax:

```
mpls
[no] interface ip-int-name
   shutdown
```

```
ALA-1>config>router>if-attr# info
----------------------------------------------
admin-group "green" value 15
admin-group "yellow" value 20
admin-group "red" value 25
----------------------------------------------
A:ALA-1>config>router>mpls# info
----------------------------------------------
...
            interface "system"
            exit
            no shutdown
----------------------------------------------
A:ALA-1>config>router>mpls#
```

2.14 RSVP Configuration Management Tasks

This section discusses RSVP configuration management tasks.

2.14.1 Modifying RSVP Parameters

Only interfaces configured in the MPLS context can be modified in the RSVP context.

The **no rsvp** command deletes this RSVP protocol instance and removes all configuration parameters for this RSVP instance.

The **shutdown** command suspends the execution and maintains the existing configuration.

The following example displays a modified RSVP configuration example:

```
A:ALA-1>config>router>rsvp# info
----------------------------------------------
            keep-multiplier 5
            refresh-time 60
            msg-pacing
                period 400
                max-burst 400
            exit
            interface "system"
            exit
            interface "test1"
                hello-interval 5000
            exit
            no shutdown
----------------------------------------------
A:ALA-1>config>router>rsvp#
```

2.14.2 Modifying RSVP Message Pacing Parameters

RSVP message pacing maintains a count of the messages that were dropped because the output queue for the egress interface was full.

The following example displays command usage to modify RSVP parameters:

The following example displays a modified RSVP message pacing configuration example. Refer to Configure RSVP Message Pacing Parameters.

```
A:ALA-1>config>router>rsvp# info
----------------------------------------------
```

```
            keep-multiplier 5
            refresh-time 60
            msg-pacing
                period 200
                max-burst 200
            exit
            interface "system"
            exit
            interface "to-104"
            exit
            no shutdown
----------------------------------------------
A:ALA-1>config>router>rsvp#
```

2.14.3 Deleting an Interface from RSVP

Interfaces cannot be deleted directly from the RSVP configuration. An interface must have been configured in the MPLS context, which enables it automatically in the RSVP context. The interface must first be deleted from the MPLS context. This removes the association from RSVP.

See Deleting an MPLS Interface for information on deleting an MPLS interface.

2.15 Configuring and Operating SR-TE

This section provides information on the configuration and operation of the Segment Routing with Traffic Engineering (SR-TE) LSP.

2.15.1 SR-TE Configuration Prerequisites

To configure SR-TE, the user must first configure prerequisite parameters.

First, configure the label space partition for the Segment Routing Global Block (SRGB) for all participating routers in the segment routing domain by using the **mpls-labels**>**sr-labels** command.

Example:

```
mpls-labels
   sr-labels start 200000 end 200400
exit
```

Enable segment routing, traffic engineering, and advertisement of router capability in all participating IGP instances in all participating routers by using the **traffic-engineering**, **advertise-router-capability**, and **segment-routing** commands.

Example:

```
ospf 0
   traffic-engineering
   advertise-router-capability area
   loopfree-alternates remote-lfa
   area 0.0.0.202
      stub
         no summaries
      exit
      interface "system"
         node-sid index 194
         no shutdown
      exit
      interface "toSim199"
         interface-type point-to-point
         no shutdown
      exit
      interface "toSim213"
         interface-type point-to-point
         no shutdown
      exit
      interface "toSim219"
         interface-type point-to-point
         metric 2000
```

```
            no shutdown
         exit
      exit
      segment-routing
         prefix-sid-range global
         no shutdown
      exit
      no shutdown
   exit
```

Configure an segment routing tunnel MTU for the IGP instance, if required, by using the **tunnel-mtu** command.

Example:

```
prefix-sid-range global
tunnel-mtu 1500
no shutdown
```

Assign a node SID to each loopback interface that a router would use as the destination of a segment routing tunnel by using the **node-sid** command.

Example:

```
ospf 0
   area 0.0.0.202
      interface "system"
         node-sid index 194
         no shutdown
      exit
```

2.15.2 SR-TE LSP Configuration Overview

An SR-TE LSP can be configured as a label switched path (LSP) using the existing CLI command hierarchy under the MPLS context and specifying the new **sr-te** LSP type.

CLI Syntax:

```
config>router>mpls>lsp lsp-name | mpls-tp src-tunnel-num
 | sr-te
```

As for an RSVP LSP, the user can configure a primary path.

Use the following CLI syntax to associate an empty path or a path with strict or loose explicit hops with the primary paths of the SR-TE LSP:

CLI Syntax:

```
config>router>mpls>path>hop hop-index ip-address {strict
 | loose}
config>router>mpls>lsp>primary path-name
```

2.15.3 Configuring Path Computation and Control for SR-TE LSP

Use the following syntax to configure the path computation requests only (PCE-computed) or both path computation requests and path updates (PCE-controlled) to PCE for a specific LSP:

CLI Syntax:

```
config>router>mpls>lsp>path-computation-method pce
config>router>mpls>lsp>pce-control
```

The PCC LSP database is synchronized with the PCE LSP database using the PCEP PCRpt (PCE Report) message for LSPs that have the following commands enabled:

CLI Syntax:

```
config>router>mpls>pce-report sr-te {enable | disable}
config>router>mpls>lsp>pce-report {enable | disable |
  inherit}
```

2.15.3.1 Configuring Path Profile and Group for PCC-Initiated and PCE-Computed/Controlled LSP

The PCE supports the computation of disjoint paths for two different LSPs originating or terminating on the same or different PE routers. To indicate this constraint to PCE, the user must configure the PCE path profile ID and path group ID the LSP belongs to. These parameters are passed transparently by PCC to PCE and are thus opaque data to the router. Use the following syntax to configure the path profile and path group:

CLI Syntax:

```
config>router>mpls>lsp>path-profile profile-id [path-
  group group-id]
```

The association of the optional path group ID is to allow PCE determine which profile ID this path group ID must be used with. One path group ID is allowed per profile ID. The user can, however, enter the same path group ID with multiple profile IDs by executing this command multiple times. A maximum of five entries of **path-profile** [*path-group*] can be associated with the same LSP. More details of the operation of the PCE path profile are provided in the PCEP section of this guide.

2.15.4 Configuring SR-TE LSP Label Stack Size

Use the following syntax to configure the maximum number of labels which the ingress LER can push for a given SR-TE LSP:

CLI Syntax:

```
config>router>mpls>lsp>max-sr-labels label-stack-size
```

This command allows the user to reduce the SR-TE LSP label stack size by accounting for additional transport, service, and other labels when packets are forwarded in a given context. See Data Path Support for more information about label stack size requirements in various forwarding contexts. If the CSPF on the PCE or the router's hop-to-label translation could not find a path that meets the maximum SR label stack, the SR-TE LSP will remain on its current path or will remain down if it has no path. The range is 1-10 labels with a default value of 6.

2.15.5 Configuring Adjacency SID Parameters

Configure the adjacency hold timer for the LFA or remote LFA backup next-hop of an adjacency SID.

Use the following syntax to configure the length of the interval during which LTN or ILM records of an adjacency SID are kept:

CLI Syntax:

```
config>router>ospf>segment-routing>adj-sid-hold
  seconds[1..300, default 15]
config>router>isis>segment-routing>adj-sid-hold
  seconds[1..300, default 15]
```

Example:

```
adj-sid-hold 15
no entropy-label-capability
prefix-sid-range global
no tunnel-table-pref
no tunnel-mtu
no backup-node-sid
no shutdown
```

While protection is enabled globally for all node SIDs and local adjacency SIDs when the user enables the **loopfree-alternates** option in ISIS or OSPF at the LER and LSR, there are applications where the user wants traffic to never divert from the strict hop computed by CSPF for a SR-TE LSP. In that case, use the following syntax to disable protection for all adjacency SIDs formed over a given network IP interface:

CLI Syntax:

```
config>router>ospf>area>if>no sid-protection
config>router>isis>if>no sid-protection
```

Example:

```
node-sid index 194
no sid-protection
no shutdown
```

2.15.6 Configuring PCC-controlled, PCE-computed, and PCE-controlled SR-TE LSPs

The following example shows the configuration of PCEP PCC parameters on LER routers that require peering with the PCE server:

Example:

```
keepalive 30
dead-timer 120
no local-address
unknown-message-rate 10
report-path-constraints
peer 192.168.48.226
  no shutdown
exit
no shutdown
```

The following example shows the configuration of a PCC-controlled SR-TE LSP that is not reported to PCE:

Example:

```
lsp "to-SanFrancisco" sr-te
  to 192.168.48.211
  path-computation-method local-cspf
  pce-report disable
  metric 10
  primary "loose-anycast"
  exit
  no shutdown
exit
```

The following example shows the configuration of a PCC-controlled SR-TE LSP that is reported to PCE:

Example:

```
lsp "to-SanFrancisco" sr-te
  to 192.168.48.211
  path-computation-method local-cspf
  pce-report enable
  metric 10
  primary "loose-anycast"
  exit
  no shutdown
exit
```

The following example shows the configuration of a PCE-computed SR-TE LSP that is reported to PCE:

Example:

```
lsp "to-SanFrancisco" sr-te
  to 192.168.48.211
  path-computation-method local-cspf
  pce-report enable
  metric 10
  primary "loose-anycast"
  exit
  no shutdown
exit
```

The following example shows the configuration of a PCE-controlled SR-TE LSP with no PCE path profile:

Example:

```
lsp "from Reno to Atlanta no Profile" sr-te
  to 192.168.48.224
  path-computation-method local-cspf
  pce-report enable
  pce-control
  primary "empty"
  exit
  no shutdown
exit
```

The following example shows the configuration of a PCE-controlled SR-TE LSP with a PCE path profile and a maximum label stack set to a non-default value:

Example:

```
lsp "from Reno to Atlanta no Profile" sr-te
  to 192.168.48.224
  max-sr-labels 8 additional-frr-labels 1
  path-computation-method pce
  pce-report enable
  pce-control
  path-profile 10 path-group 2
  primary "empty"
    bandwidth 15
  exit
  no shutdown
exit
```

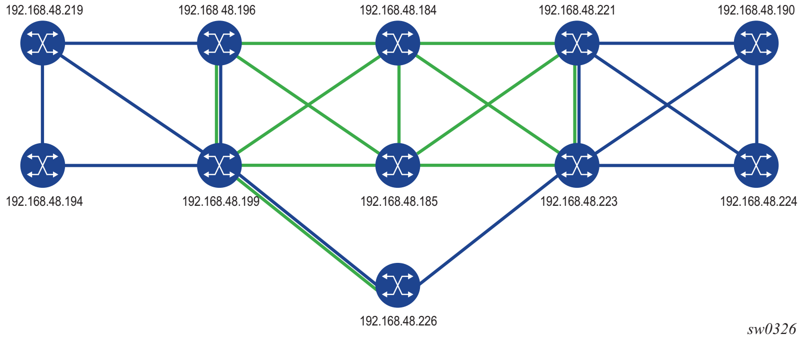
2.15.7 Configuring a Mesh of SR-TE Auto-LSPs

The following shows the detailed configuration for the creation of a mesh of SR-TE auto-LSPs. The network uses IS-IS with the backbone area being in Level 2 and the leaf areas being in Level 1.

The NSP is used for network discovery only and the NRC-P learns the network topology using BGP-LS.

Figure 42 shows the view of the multi-level IS-IS topology in the NSP GUI. The backbone L2 area is highlighted in green.

Figure 42 **Multi-level IS-IS Topology in the NSP GUI**



The mesh of SR-TE auto-LSPs is created in the backbone area and originates on an ABR node with address 192.168.48.199 (Phoenix 199). The LSP template uses a default path that includes an anycast SID prefix corresponding to a transit routers 192.168.48.184 (Dallas 184) and 192.168.48.185 (Houston 185).

The following is the configuration of transit router Dallas 184, which shows the creation of a loopback interface with the anycast prefix and the assignment of a SID to it. The same configuration must be performed on the transit router Houston 185. See lines marked with an asterisk (*).

```
*A:Dallas 184>config>router# info
----------------------------------------------
echo "IP Configuration"
#---------------------------------------------
        if-attribute
            admin-group "olive" value 20
            admin-group "top" value 10
            srlg-group "top" value 10
        exit
        interface "anycast-sid"                                  *
            address 192.168.48.99/32                             *
            loopback                                             *
            no shutdown                                          *
```

```
                exit
                interface "system"
                    address 192.168.48.184/32
                    no shutdown
                exit
                interface "toJun164"
                    address 10.19.2.184/24
                    port 1/1/4:10
                    no shutdown
                exit
                interface "toSim185"
                    address 10.0.3.184/24
                    port 1/1/2
                    no shutdown
                exit
                interface "toSim198"
                    address 10.0.2.184/24
                    port 1/1/3
                    if-attribute
                        admin-group "olive"
                    exit
                    no shutdown
                exit
                interface "toSim199"
                    address 10.0.13.184/24
                    port 1/1/5
                    no shutdown
                exit
                interface "toSim221"
                    address 10.0.4.184/24
                    port 1/1/1
                    no shutdown
                exit
                interface "toSim223"
                    address 10.0.14.184/24
                    port 1/1/6
                    no shutdown
                exit
#--------------------------------------------------

*A:Dallas 184>config>router>isis# info
----------------------------------------------
                    level-capability level-2
                    area-id 49.0000
                    database-export identifier 10 bgp-ls-identifier 10
                    traffic-engineering
                    advertise-router-capability area
                    level 2
                        wide-metrics-only
                    exit
                    interface "system"
                        ipv4-node-sid index 384
                        no shutdown
                    exit
                    interface "toSim198"
                        interface-type point-to-point
                        no shutdown
                    exit
                    interface "toSim185"
                        interface-type point-to-point
                        no shutdown
                    exit
                    interface "toSim221"
                        interface-type point-to-point
                        no shutdown
                    exit
                    interface "toSim199"
                        interface-type point-to-point
                        level 2
                            metric 100
                        exit
                        no shutdown
                    exit
                    interface "toSim223"
                        interface-type point-to-point
                        level 2
                            metric 100
                        exit
                        no shutdown
                    exit
                    interface "anycast-sid"                                                *
                        ipv4-node-sid index 99                                             *
                        no shutdown                                                        *
                    exit
                    segment-routing
                        prefix-sid-range global
                        no shutdown
                    exit
                    no shutdown
        ----------------------------------------------
```

In the ingress LER Phoenix 199 router, the anycast SID is learned from both transit routers, but is currently resolved in IS-IS to transit router Houston 185. See lines marked with an asterisk (*).

```
*A:Phoenix 199# show router isis prefix-sids
===============================================================================
Rtr Base ISIS Instance 0 Prefix/SID Table
===============================================================================
Prefix                            SID        Lvl/Typ    SRMS   AdvRtr
                                                         MT     Flags
-------------------------------------------------------------------------------
192.168.48.194/32                  399        1/Int.      N     Reno 194
                                                            0       NnP
192.168.48.194/32                  399        2/Int.      N     Salt Lake 198
                                                            0       RNnP
192.168.48.194/32                  399        2/Int.      N     Phoenix 199
                                                            0       RNnP
192.168.48.99/32                   99         2/Int.      N     Dallas 184         *
                                                            0       NnP            *
192.168.48.99/32                   99         2/Int.      N     Houston 185        *
                                                            0       NnP            *
192.168.48.184/32                  384        2/Int.      N     Dallas 184
                                                            0       NnP
192.168.48.185/32                  385        2/Int.      N     Houston 185
                                                            0       NnP
192.168.48.190/32                  390        2/Int.      N     Chicago 221
```

```
                                                            0       RNnP
192.168.48.190/32                  390        2/Int.      N     St Louis 223
                                                            0       RNnP
192.168.48.194/32                  394        1/Int.      N     Reno 194
                                                            0       NnP
192.168.48.194/32                  394        2/Int.      N     Salt Lake 198
                                                            0       RNnP
192.168.48.194/32                  394        2/Int.      N     Phoenix 199
                                                            0       RNnP
192.168.48.198/32                  398        1/Int.      N     Salt Lake 198
                                                            0       NnP
192.168.48.198/32                  398        2/Int.      N     Salt Lake 198
                                                            0       NnP
192.168.48.198/32                  398        2/Int.      N     Phoenix 199
                                                            0       RNnP
192.168.48.199/32                  399        2/Int.      N     Salt Lake 198
                                                            0       RNnP
192.168.48.199/32                  399        1/Int.      N     Phoenix 199
                                                            0       NnP
192.168.48.199/32                  399        2/Int.      N     Phoenix 199
                                                            0       NnP
192.168.48.219/32                  319        2/Int.      N     Salt Lake 198
                                                            0       RNnP
192.168.48.219/32                  319        2/Int.      N     Phoenix 199
                                                            0       RNnP
192.168.48.219/32                  319        1/Int.      N     Las Vegas 219
                                                            0       NnP
192.168.48.221/32                  321        2/Int.      N     Chicago 221
                                                            0       NnP
192.168.48.221/32                  321        2/Int.      N     St Louis 223
                                                            0       RNnP
192.168.48.223/32                  323        2/Int.      N     Chicago 221
                                                            0       RNnP
192.168.48.223/32                  323        2/Int.      N     St Louis 223
                                                            0       NnP
192.168.48.224/32                  324        2/Int.      N     Chicago 221
                                                            0       RNnP
192.168.48.224/32                  324        2/Int.      N     St Louis 223
                                                            0       RNnP
192.168.48.226/32                  326        2/Int.      N     PCE Server 226
                                                            0       NnP
3ffe::a14:194/128                  294        1/Int.      N     Reno 194
                                                            0       NnP
3ffe::a14:194/128                  294        2/Int.      N     Phoenix 199
                                                            0       RNnP
3ffe::a14:199/128                  299        1/Int.      N     Phoenix 199
                                                            0       NnP
3ffe::a14:199/128                  299        2/Int.      N     Phoenix 199
                                                            0       NnP
-------------------------------------------------------------------------------
No. of Prefix/SIDs: 32 (15 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
===============================================================================

*A:Phoenix 199# tools dump router segment-routing tunnel
====================================================================================
Legend: (B) - Backup Next-hop for Fast Re-Route
        (D) - Duplicate
====================================================================================
-----------------------------------------------------------------------------------+
 Prefix                                                                            |
 Sid-Type        Fwd-Type       In-Label  Prot-Inst                                |
                 Next Hop(s)                      Out-Label(s) Interface/Tunnel-ID |
-----------------------------------------------------------------------------------+
 192.168.48.99                                                                     *
 Node            Orig/Transit   200099    ISIS-0                                   *
                 10.0.5.185                       200099      toSim185             *
 3ffe::a14:194
 Node            Orig/Transit   200294    ISIS-0
                 fe80::62c2:ffff:fe00:0           200294      toSim194
 3ffe::a14:199
 Node            Terminating    200299    ISIS-0
 192.168.48.219
 Node            Orig/Transit   200319    ISIS-0
                 10.202.5.194                     200319      toSim194
 192.168.48.221
 Node            Orig/Transit   200321    ISIS-0
                 10.0.5.185                       200321      toSim185
 192.168.48.223
 Node            Orig/Transit   200323    ISIS-0
                 10.0.5.185                       200323      toSim185
 192.168.48.224
 Node            Orig/Transit   200324    ISIS-0
                 10.0.5.185                       200324      toSim185
 192.168.48.226
 Node            Orig/Transit   200326    ISIS-0
                 10.0.1.2                         100326      toSim226PCEServer
 192.168.48.184
 Node            Orig/Transit   200384    ISIS-0
                 10.0.5.185                       200384      toSim185
 192.168.48.185
 Node            Orig/Transit   200385    ISIS-0
                 10.0.5.185                       200385      toSim185
 192.168.48.190
 Node            Orig/Transit   200390    ISIS-0
                 10.0.5.185                       200390      toSim185
 192.168.48.194
 Node            Orig/Transit   200394    ISIS-0
                 10.202.5.194                     200394      toSim194
 192.168.48.198
 Node            Orig/Transit   200398    ISIS-0
                 10.0.9.198                       100398      toSim198
 192.168.48.199
 Node            Terminating    200399    ISIS-0
 10.0.9.198
 Adjacency       Transit        262122    ISIS-0
                 10.0.9.198                       3           toSim198
 10.202.1.219
 Adjacency       Transit        262124    ISIS-0
                 10.202.1.219                     3           toSim219
 10.0.5.185
 Adjacency       Transit        262133    ISIS-0
                 10.0.5.185                       3           toSim185
 fe80::62c2:ffff:fe00:0
 Adjacency       Transit        262134    ISIS-0
                 fe80::62c2:ffff:fe00:0           3           toSim194
 10.0.1.2
 Adjacency       Transit        262137    ISIS-0
                 10.0.1.2                         3           toSim226PCEServer
 10.0.13.184
 Adjacency       Transit        262138    ISIS-0
                 10.0.13.184                      3           toSim184
 10.0.2.2
 Adjacency       Transit        262139    ISIS-0
                 10.0.2.2                         3           toSim226PCEserver202
 10.202.5.194
 Adjacency       Transit        262141    ISIS-0
                 10.202.5.194                     3           toSim194
------------------------------------------------------------------------------------
No. of Entries: 22
------------------------------------------------------------------------------------
```

Next, a policy must be configured to add the list of prefixes to which the ingress LER Phoenix 199 must auto-create SR-TE LSPs.

```
*A:Phoenix 199>config>router>policy-options# info
----------------------------------------------
            prefix-list "sr-te-level2"
                prefix 192.168.48.198/32 exact
                prefix 192.168.48.221/32 exact
                prefix 192.168.48.223/32 exact
            exit
            policy-statement "sr-te-auto-lsp"
                entry 10
                    from
                        prefix-list "sr-te-level2"
                    exit
                    action accept
                    exit
                exit
                default-action drop
                exit
            exit
----------------------------------------------
```

Then, an LSP template of type **mesh-p2p-srte** must be configured, which uses a path with a loose-hop corresponding to anycast-SID prefix of the transit routers. The LSP template is then bound to the policy containing the prefix list. See lines marked with an asterisk (*).

```
*A:Phoenix 199>config>router>mpls# info
----------------------------------------------
            cspf-on-loose-hop
            interface "system"
                no shutdown
            exit
            interface "toESS195"
                no shutdown
            exit
            interface "toSim184"
                no shutdown
            exit
            interface "toSim185"
                admin-group "bottom"
                srlg-group "bottom"
                no shutdown
            exit
            interface "toSim194"
                admin-group "bottom"
                srlg-group "bottom"
                no shutdown
            exit
            interface "toSim198"
                no shutdown
            exit
            interface "toSim219"
                no shutdown
            exit
            path "loose-anycast-sid"                                               *
                hop 1 192.168.48.99 loose                                          *
                no shutdown                                                        *
            exit                                                                   *
            lsp-template "sr-te-level2-mesh" mesh-p2p-srte                         *
                default-path "loose-anycast-sid"                                   *
                max-sr-labels 8 additional-frr-labels 2                            *
                pce-report enable                                                  *
                no shutdown                                                        *
            exit                                                                   *
            auto-lsp lsp-template "sr-te-level2-mesh" policy "sr-te-auto-lsp"      *
            no shutdown                                                            *
----------------------------------------------
```

One SR-TE LSP should be automatically created to each destination matching the prefix in the policy as soon as the router with the router ID matching the address of the prefix appears in the TE database.

The following shows the three SR-TE auto-LSPs created. See lines marked with an asterisk (*).

```
*A:Phoenix 199# show router mpls sr-te-lsp
===============================================================================
MPLS SR-TE LSPs (Originating)
===============================================================================
LSP Name                           To               Tun     Protect   Adm  Opr
                                                    Id      Path
-------------------------------------------------------------------------------
Phoenix-SL-1                       192.168.48.223    1       N/A       Up   Up
Phoenix-SL-2-Profile               192.168.48.223    2       N/A       Up   Up
Phoenix-SL-3-Profile               192.168.48.223    3       N/A       Up   Up
Phoenix-SL-4-Profile               192.168.48.223    4       N/A       Up   Up
Phoenix-SL-1-Profile               192.168.48.223    5       N/A       Up   Up
Phoenix-SL-2                       192.168.48.223    6       N/A       Up   Up
Phoenix-SL-3                       192.168.48.223    7       N/A       Up   Up
```

```
Phoenix-SL-4                       192.168.48.223    8       N/A       Up   Up
sr-te-level2-mesh-192.168.48.198-  192.168.48.198    61442   N/A       Up   Up      *
716803                                                                             *
sr-te-level2-mesh-192.168.48.221-  192.168.48.221    61443   N/A       Up   Up      *
716804                                                                             *
sr-te-level2-mesh-192.168.48.223-  192.168.48.223    61444   N/A       Up   Up      *
716805                                                                             *
-------------------------------------------------------------------------------
LSPs : 17
===============================================================================
```

The auto-generated name uses the syntax convention “*TemplateName-DestIpv4Address-TunnelId*”, as explained in Automatic Creation of an SR-TE Mesh LSP. The tunnel ID used in the name is the TTM tunnel ID, not the MPLS LSP tunnel ID. See lines marked with an asterisk (*).

```
*A:Phoenix 199# show router mpls sr-te-lsp "sr-te-level2-mesh-192.168.48.223-
716805" detail
===============================================================================
MPLS SR-TE LSPs (Originating) (Detail)
===============================================================================
-------------------------------------------------------------------------------
Type : Originating
-------------------------------------------------------------------------------
LSP Name        : sr-te-level2-mesh-192.168.48.223-716805
LSP Type        : MeshP2PSrTe               LSP Tunnel ID        : 61444           *
LSP Index       : 126979                    TTM Tunnel Id        : 716805          *
From            : 192.168.48.199            To                   : 192.168.48.2*
Adm State       : Up                        Oper State           : Up
LSP Up Time     : 0d 00:02:12               LSP Down Time        : 0d 00:00:00
Transitions     : 3                         Path Changes         : 3
Retry Limit     : 0                         Retry Timer          : 30 sec
CSPF            : Enabled
Metric          : N/A                       Use TE metric        : Disabled
Include Grps    :                           Exclude Grps         :
None                                           None
VprnAutoBind    : Enabled
IGP Shortcut    : Enabled                   BGP Shortcut         : Enabled
IGP LFA         : Disabled                  IGP Rel Metric       : Disabled
BGPTransTun     : Enabled
Oper Metric     : 16777215
PCE Report      : Enabled
PCE Compute     : Disabled                  PCE Control          : Disabled
Max SR Labels   : 8                         Additional FRR Labels: 2
Path Profile    :
None
Primary(a)      : loose-anycast-sid         Up Time              : 0d 00:02:12
Bandwidth       : 0 Mbps
===============================================================================
```

These SR-TE auto-LSPs are also added into the tunnel table to be used by services and shortcut applications. See lines marked with an asterisk (*).

```
*A:Phoenix 199# show router tunnel-table
===============================================================================
IPv4 Tunnel Table (Router: Base)
```

```
===============================================================================
Destination       Owner     Encap TunnelId  Pref     Nexthop        Metric
-------------------------------------------------------------------------------
10.0.5.185/32     isis (0)  MPLS  524370    11       10.0.5.185     0
10.0.9.198/32     isis (0)  MPLS  524368    11       10.0.9.198     0
10.0.13.184/32    isis (0)  MPLS  524340    11       10.0.13.184    0
10.202.1.219/32   isis (0)  MPLS  524333    11       10.202.1.219   0
10.202.5.194/32   isis (0)  MPLS  524355    11       10.202.5.194   0
10.0.1.2/32       isis (0)  MPLS  524364    11       11.0.1.2       0
10.0.2.2/32       isis (0)  MPLS  524363    11       11.0.2.2       0
192.168.48.99/32  isis (0)  MPLS  524294    11       10.0.5.185     10
192.168.48.184/32 ldp       MPLS  65605     9        10.0.5.185     20
192.168.48.184/32 isis (0)  MPLS  524341    11       10.0.5.185     20
192.168.48.185/32 ldp       MPLS  65602     9        10.0.5.185     10
192.168.48.185/32 isis (0)  MPLS  524371    11       10.0.5.185     10
192.168.48.190/32 ldp       MPLS  65606     9        10.0.5.185     40
192.168.48.190/32 isis (0)  MPLS  524362    11       10.0.5.185     40
192.168.48.194/32 ldp       MPLS  65577     9        10.202.5.194   10
192.168.48.194/32 isis (0)  MPLS  524331    11       10.202.5.194   10
192.168.48.198/32 sr-te     MPLS  716803    8        192.168.48.99  16777215      *
192.168.48.198/32 ldp       MPLS  65601     9        10.0.9.198     10
192.168.48.198/32 isis (0)  MPLS  524369    11       10.0.9.198     10
192.168.48.219/32 ldp       MPLS  65579     9        10.202.5.194   20
192.168.48.219/32 isis (0)  MPLS  524334    11       10.202.5.194   20
192.168.48.221/32 sr-te     MPLS  716804    8        192.168.48.99  16777215      *
192.168.48.221/32 ldp       MPLS  65607     9        10.0.5.185     30
192.168.48.221/32 isis (0)  MPLS  524358    11       10.0.5.185     30
192.168.48.223/32 sr-te     MPLS  655362    8        10.0.13.184    200
192.168.48.223/32 sr-te     MPLS  655363    8        10.0.13.184    200
192.168.48.223/32 sr-te     MPLS  655364    8        10.0.5.185     40
192.168.48.223/32 sr-te     MPLS  655365    8        10.0.13.184    120
192.168.48.223/32 sr-te     MPLS  655366    8        10.0.5.185     120
192.168.48.223/32 sr-te     MPLS  655367    8        10.0.13.184    120
192.168.48.223/32 sr-te     MPLS  655368    8        10.0.13.184    200
192.168.48.223/32 sr-te     MPLS  655369    8        10.0.5.185     40
192.168.48.223/32 sr-te     MPLS  716805    8        192.168.48.99  16777215      *
192.168.48.223/32 ldp       MPLS  65603     9        10.0.5.185     20
192.168.48.223/32 isis (0)  MPLS  524306    11       10.0.5.185     20
192.168.48.224/32 ldp       MPLS  65604     9        10.0.5.185     30
192.168.48.224/32 isis (0)  MPLS  524361    11       10.0.5.185     30
192.168.48.226/32 isis (0)  MPLS  524365    11       11.0.1.2       65534
-------------------------------------------------------------------------------
Flags: B = BGP backup route available
       E = inactive best-external BGP route
===============================================================================
```

The details of the path of one of the SR-TE auto-LSPs now show the ERO transiting through the anycast SID of router Houston 185. See lines marked with an asterisk (*).

```
*A:Phoenix 199# show router mpls sr-te-lsp "sr-te-level2-mesh-192.168.48.223-
716805" path detail
===============================================================================
MPLS SR-TE LSP sr-te-level2-mesh-192.168.48.223-716805 Path  (Detail)
===============================================================================
Legend :
    S      - Strict                      L      - Loose
    A-SID  - Adjacency SID               N-SID  - Node SID
    +      - Inherited
```

```
===============================================================================
-------------------------------------------------------------------------------
SR-TE LSP sr-te-level2-mesh-192.168.48.223-716805 Path loose-anycast-sid
-------------------------------------------------------------------------------
LSP Name         : sr-te-level2-mesh-192.168.48.223-716805
Path LSP ID      : 20480
From             : 192.168.48.199       To                   : 192.168.48.223
Admin State      : Up                   Oper State           : Up
Path Name        : loose-anycast-sid    Path Type            : Primary
Path Admin       : Up                   Path Oper            : Up
Path Up Time     : 0d 02:30:28          Path Down Time       : 0d 00:00:00
Retry Limit      : 0                    Retry Timer          : 30 sec
Retry Attempt    : 1                    Next Retry In        : 0 sec
CSPF             : Enabled              Oper CSPF            : Enabled
Bandwidth        : No Reservation       Oper Bandwidth       : 0 Mbps
Hop Limit        : 255                  Oper HopLimit        : 255
Setup Priority   : 7                    Oper Setup Priority  : 7
Hold Priority    : 0                    Oper Hold Priority   : 0
Inter-area       : N/A
PCE Updt ID      : 0                    PCE Updt State       : None
PCE Upd Fail Code: noError
PCE Report       : Enabled              Oper PCE Report      : Disabled
PCE Control      : Disabled             Oper PCE Control     : Disabled
PCE Compute      : Disabled
Include Groups   :                      Oper Include Groups  :
None                                           None
Exclude Groups   :                      Oper Exclude Groups  :
None                                           None
IGP/TE Metric    : 16777215             Oper Metric          : 16777215
Oper MTU         : 1492                 Path Trans           : 1
Failure Code     : noError
Failure Node     : n/a
Explicit Hops    :
    192.168.48.99(L)
Actual Hops      :
    192.168.48.99 (192.168.48.185)(N-SID)          Record Label        : 200099    *
 -> 192.168.48.223 (192.168.48.223)(N-SID)         Record Label        : 200323    *
===============================================================================
```

3 GMPLS

3.1 GMPLS

The Generalized Multi-Protocol Label Switching (GMPLS) User to Network Interface (UNI) permits dynamic provisioning of optical transport connections between IP routers and optical network elements in order to reduce the operational time and administrative overhead required to provision new connectivity. The optical transport connections typically originate and terminate in an IP/MPLS controlled domain and traverse an intermediate optical transport network. The GMPLS UNI model is based on an overlay approach, whereby the IP/MPLS control plane is transported transparently over the intermediate transport network, which itself is controlled by a GMPLS control plane.

The UNI provides a clear demarcation point between the domains of responsibility of the parties involved in managing the overlying IP/MPLS network and the underlying optical network. For example, these parties could be two divisions in a service provider organization, or a subscriber/client of the service provider and the service provider itself.

The UNI has a client part, the UNI-C, and a network part, the UNI-N. In the Nokia solution, the UNI-C is an SR OS system, such as a 7750 SR or a 7950 XRS, while the UNI-N is an optical device; for example, an 1830 PSS.

Control plane related information is exchanged between the UNI-C and the UNI-N using a dedicated out of band communication channel. Note that the adjacent optical network element and the router assume that they are connected to a trusted peer, and thus assume a secure communication. This is achieved by physically securing the link carrying the control channel between the two.

Based on standardized UNI messaging (RFC 4208), the UNI-C indicates to the UNI-N which far-end peer UNI-C node (corresponding to a remote router) to make an optical transport connection to. This path request can include additional path attributes to indicate requirements such as bandwidth, priority and diversity/ resiliency parameters.

3.1.1 Example Applications

This section summarizes some of the use cases that the GMPLS UNI may be used to address.

3.1.1.1 Use Case 1: Dynamic Connection Setup with Constraints

This use case aims to solve inefficiencies between IP and transport teams within an operator for connectivity setup; for example:

- Process complexity, with complex database exchange, parsing and filtering
- Long-winded organizational communication prior to path establishment

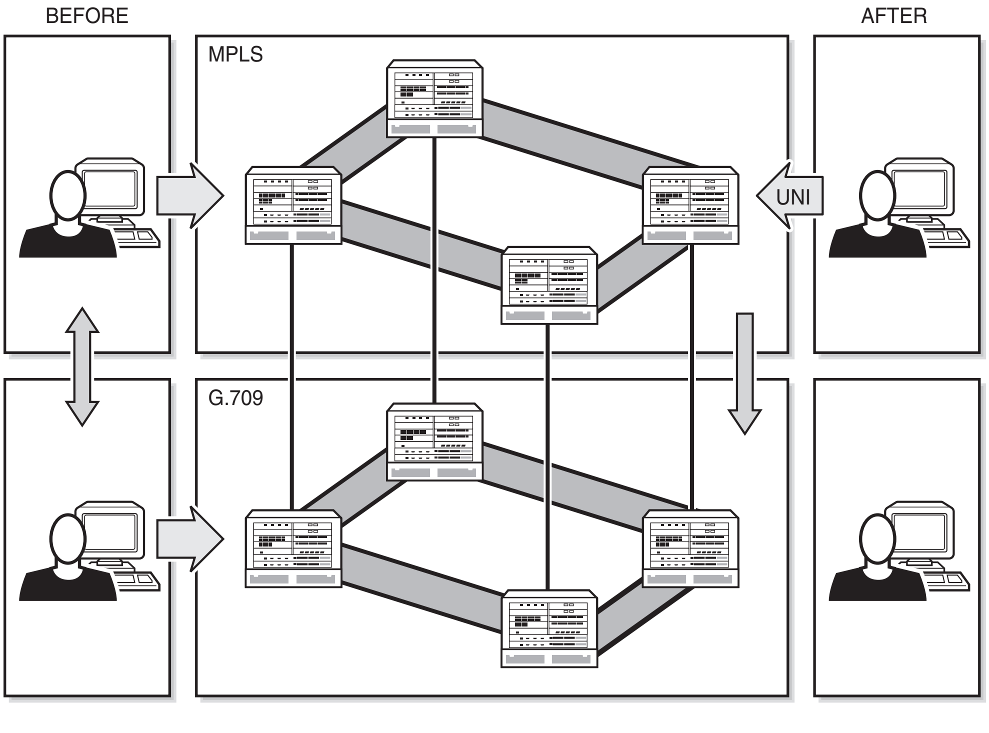
It therefore aims to optimize IP/Optical transport team interactions by removing complex processes, and reduces per-connection provisioning in the optical core.

The UNI should allow the setup/maintenance/release of connections across an intermediate optical transport network from a UNI-C router to another remote UNI-C router. The routers are connected to an optical network that consists of optical cross connects (OXCs), and the interconnection between the OXC and the router is based on the GMPLS UNI (RFC 4208). The UNI-C routers are 7750 SR or 7950 XRS nodes, while the UNI-N OXC is the Nokia1830 PSS. The UNI connection is instantiated using a GMPLS LSP (gLSP).

The UNI-C router is always the initiator of the connection. The only per-connection configuration occurs at the UNI-C, and it is operator initiated. Connections to any of the remote UNI-C routers are signaled over the UNI. The initiation of a connection request is via CLI or SNMP to the UNI-C router.

Signaling is based on RSVP-TE (RFC 4208). Constraints can be signaled with a connection setup request. These include bandwidth, protection type, and latency. In the event that a connection could not be established, a correct (descriptive) error code is returned to the initiator.

Figure 43 **Dynamic Connection Setup**



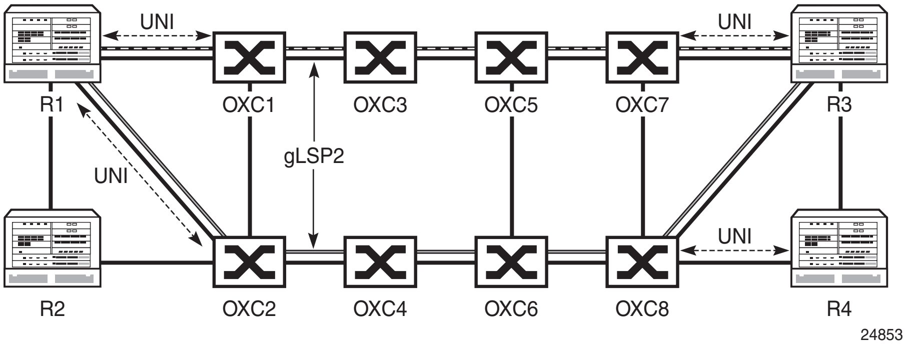
3.1.1.2 Use Case 2: Multi-Layer Resiliency

The objective of this application is to ensure optical network path diversity for primary/backup paths of an overlay IP network. It thus aims to resolve situations where the UNI-C router has no topological visibility of the optical network, and to allow the router to indicate that paths have to be either co-routed or avoid specific optical nodes or links along a path.

Route diversity for LSPs from single homed UNI-C router and dual-homed UNI-C router is a common requirement in optical transport networks. Dual homing is typically used to avoid a single point of failure (for example, the UNI link or OXC) or to allow two disjoint connections to form a protection group.

For the dual-homing case, it is possible to establish two connections from the source router to the same destination router where one connection is using one UNI link to, for example, OXC1 and the other connection is using the UNI link to OXC2. In order to avoid single points of failure within the optical network, it is necessary to also ensure path (gLSP) diversity within the provider network in order to achieve end-to-end diversity for the two gLSPs between the two routers.

Figure 44 **Multi-Layer Resiliency**

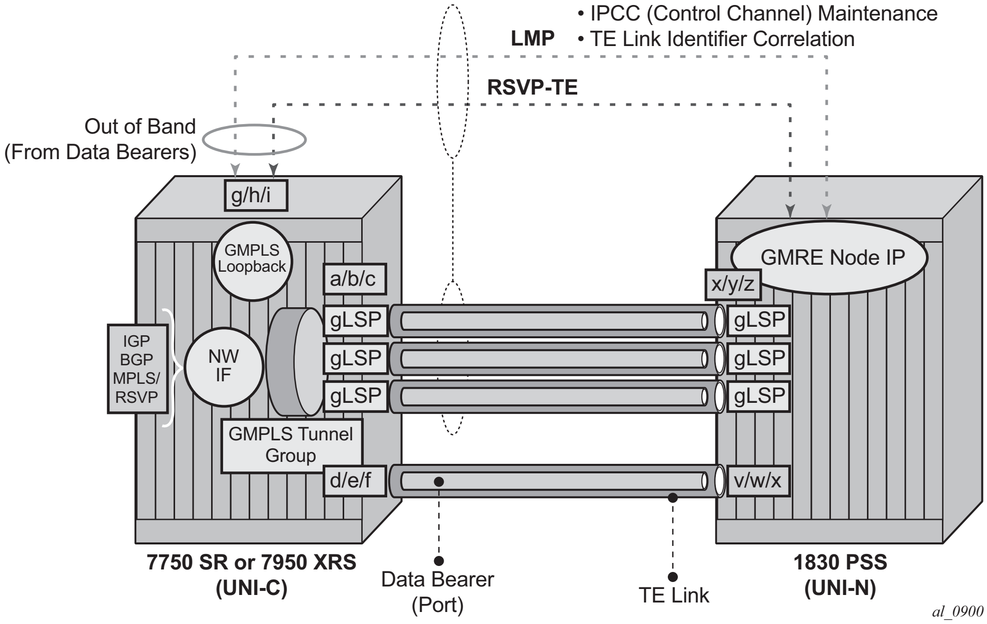


As the two connections are entering the provider network at different OXC devices, the OXC device that receives the connection request for the second connection needs to be capable of determining the additional path computation constraints such that the path of the second LSP is disjointed with respect to the already established first connection entering the network at a different PE device.

3.2 GMPLS UNI Architecture

This section specifies the architectural and functional elements of the GMPLS UNI on the 7750 SR or 7950 XRS nodes and the 1830 PSS node (which must be GMRE), and how they relate to one another. The architecture is illustrated in Figure 45.

Figure 45 **GMPLS UNI Architecture**



On the UNI-C side, the UNI consists of the following functional components:

- A set of one or more data bearers between the UNI-C and UNI-N. Each data bearer maps to a black and white Ethernet network port.
- A Traffic Engineering (TE) link (RFC 4202), represented by an identifier on the UNI-C and UNI-N nodes. This identifier is manually configured. A TE link maps to a single data bearer. There may be one or more TE links per UNI between a UNI-C and UNI-N pair.
- An IP Control Channel (IPCC) between the UNI-C and UNI-N. The IPCC carries GMPLS control plane traffic between the two nodes and is separate from the links carrying user plane traffic. The IPCC may be native unencapsulated traffic, or it may be encapsulated in a GRE tunnel, and may use either an IP interface bound to a network Ethernet port on an Ethernet MDA/IMM or an OES Ethernet port on the CPM. This port is separate from the TE links. The IPCC carries the following two control protocols:

  - LMP — This is responsible for checking the correlation between the UNI-C/UNI-N and the TE link/Data Bearer identifiers, and maintaining the IPCC adjacency between the UNI-C and UNI-N.
  - RSVP-TE — RSVP-TE runs on the same network interface as LMP. The next hop from an RSVP-TE perspective is the UNI-N. RSVP-TE is used to establish and maintain a GMPLS LSP.
- gLSP — The GMPLS LSP. At the UNI-C, this is a control plane object representing the TE-Link in the RSVP-TE control plane. Although this is an LSP, there is no explicit MPLS label in the data path at the UNI-C; the gLSP maps to a data bearer of the TE link to / from the UNI-N. When a gLSP is signaled to a far-end UNI-C node, the optical network establishes bidirectional connectivity between one of the data bearers in the TE link on the UNI-N at the ingress to the optical network, and one of the data bearers on the TE link on the egress UNI-N node connected to the far end UNI-C node.
- Network Interface — When a gLSP is successfully established, a network interface can be bound to the gLSP. The network interface then uses the data bearer associated with the gLSP to forward traffic. This network interface can be used by any applicable protocol associated with an overlying IP/MPLS network. The network interface is bound to the gLSPs via a GMPLS tunnel group.
- GMPLS Tunnel Group: A GMPLS tunnel group is a bundle of gLSPs providing an abstraction of the data bearers that are intended to be associated to one IP interface. A GMPLS tunnel group only exists on the UNI-C node and not on the 1830 PSS UNI-N node.

Although Figure 45 shows a single 7750 SR or 7950 XRS node connected to a single UNI-N (1830 PSS), it is possible to multi-home a router into more than one (usually two) UNI-Ns at the edge of the optical network. In this case, a separate IPCC, set of data bearers, and set of TE links, are required between the 7750 SR or 7950 XRS and each UNI-N.

3.2.1 Addressing and End-to-End gLSP Architecture

The GMPLS UNI assumed a flat addressing scheme between the UNI-C nodes and the optical network. In this model, a common addressing scheme is used between the UNI-C (IP router) and UNI-N (optical edge). The UNI-C and UNI-N must be in the same subnet. Also, none of the UNI-C addresses can overlap or clash with any of the GMPLS-aware nodes in the optical network. This does not mandate that the whole IP network share a common address space with the optical network, as a separate loopback address can be used for the GMPLS UNI on the UNI-C.

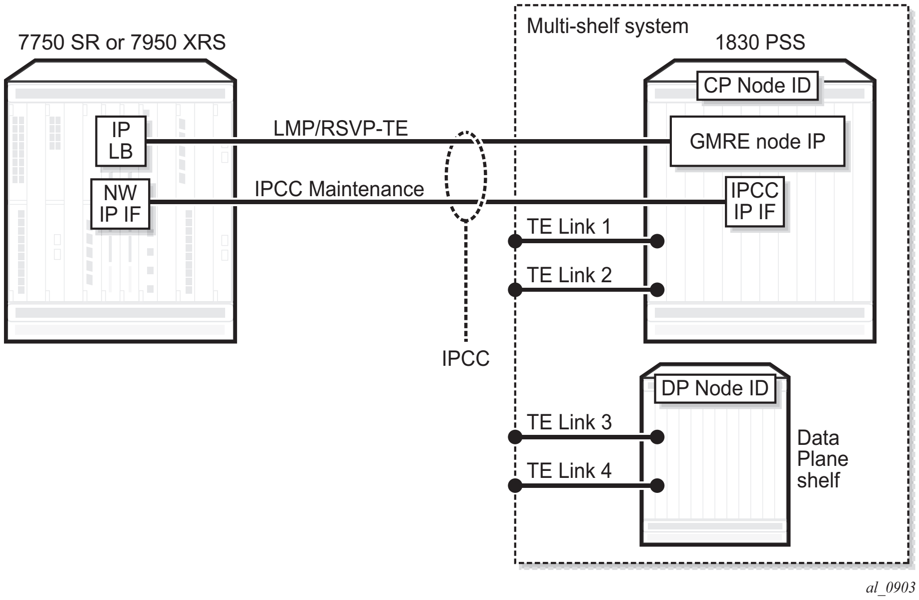
The ERO Expansion (RFC 5151) model is assumed for the GMPLS LSPs. The UNI-C is not exposed to the full ERO between the UNI-N nodes. Instead, the full ERO is inserted at the UNI-N. This model limits the sharing of topology information between the UNI-N and UNI-C.

3.3 1830 PSS Identifiers

This section describes the various identifiers used on the 1830 PSS that are relevant to configuring the GMPLS UNI on the 7750 SR or 7950 XRS node in conjunction with the 1830 PSS. Figure 46 illustrates the identifier architecture of a 1830 PSS multi-shelf system. The multi-shelf system consists of a control plane node and one or more data plane nodes. The following identifiers are used:

- GMRE node IP— This is the IP loopback address used for GMPLS protocols such as LMP and RSVP.
- IPCC IP address (also known as DcnGatewayAddress)— This is the source/destination address for IPCC maintenance messages such as LMP hellos and LMP config messages. When only one IPCC exists between a 7750 SR or 7950 XRS and 1830 PSS pair, this may be the same as the IP management loopback.
- CP Node ID — This is a non-routable identifier for the control plane node. It is used for identifying this node in the optical domain; for example, the session/sender template. It is also used in the RSVP ERO to identify the 1830 PSS node.
- DP Node ID — This is a non-routable identifier for a data plane node. This identifies a particular data plane shelf in the optical domain.
- TE Link IDs — The TE Link ID is unique across a set of DP and CP nodes forming an 1830 PSS.

Figure 46 **Identifier Architecture**

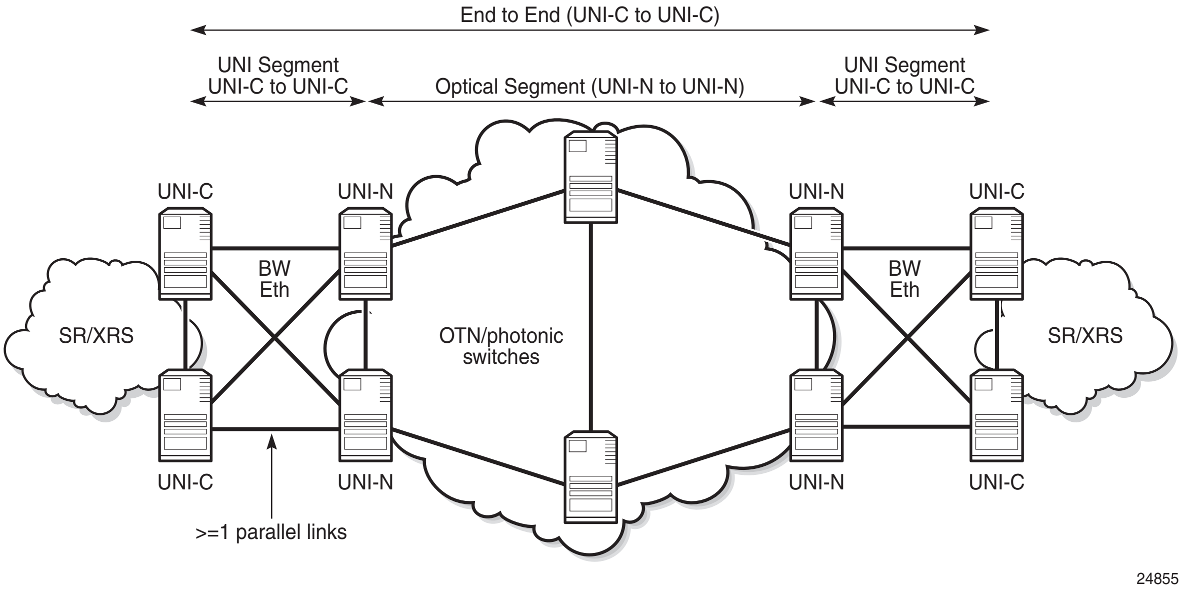


3.4 Recovery Reference Models

This section details the supported recovery reference models. These models are based on the mechanisms specified in RFC 4872 and RFC 4873.

Figure 47 presents a generalized reference model in which the 7750 SR or 7950 XRS UNI-C nodes are dual-homed at the link layer to the 1830 PSS UNI-N nodes. Not all elements of this architecture may be required in all deployment cases.

Figure 47 **General GMPLS UNI Interconnection Architecture**



This reference model includes two 7750 SR or 7950 XRS nodes, each hosting a UNI-C function, at the edge of each IP network facing two 1830 PSS nodes, each hosting a UNI-N function. A full mesh of black and while Ethernet links interconnects neighboring UNI-C nodes and UNI-N nodes. Parallel links may exist, so that a given 7750 SR or 7950 XRS UNI-C is connected to a neighbor 1830 PSS UNI-N by more than one Ethernet link.

Each router hosting a UNI-C has an IPCC to each of the two 1830 PSS UNI-Ns. Likewise, each 1830 PSS hosting UNI-N has an IPCC to both of the 7750 SR or 7950 XRS UNI-Cs that it is connected to. IPCCs only exist between UNI-C and UNI-N nodes, and not between UNI-C nodes. A control plane (LMP and RSVP) adjacency therefore exists between each UNI-C and it's corresponding UNI-Ns.

Recovery in the following domains is supported in the following locations:

- End to End — Between the 7750 SR or 7950 XRS UNI-C nodes at each end of a gLSP.
- Optical Segment — Between 1830 PSS UNI-N nodes at each edge of the optical network.

The following subsections detail some example recovery options that are possible using either GMPLS, or a combination of GMPLS mechanisms and mechanisms in the overlay IP network. Note that some of the functionality shown in one of the scenarios can be used in combination with functionality in another scenario, for example optical SRLG diversity.

The objective of GMPLS here is to minimize the disruption to the overlay IP network while simultaneously maximizing the utilization of both the gLSPs and the resources in the underlying optical network (or UNI links).

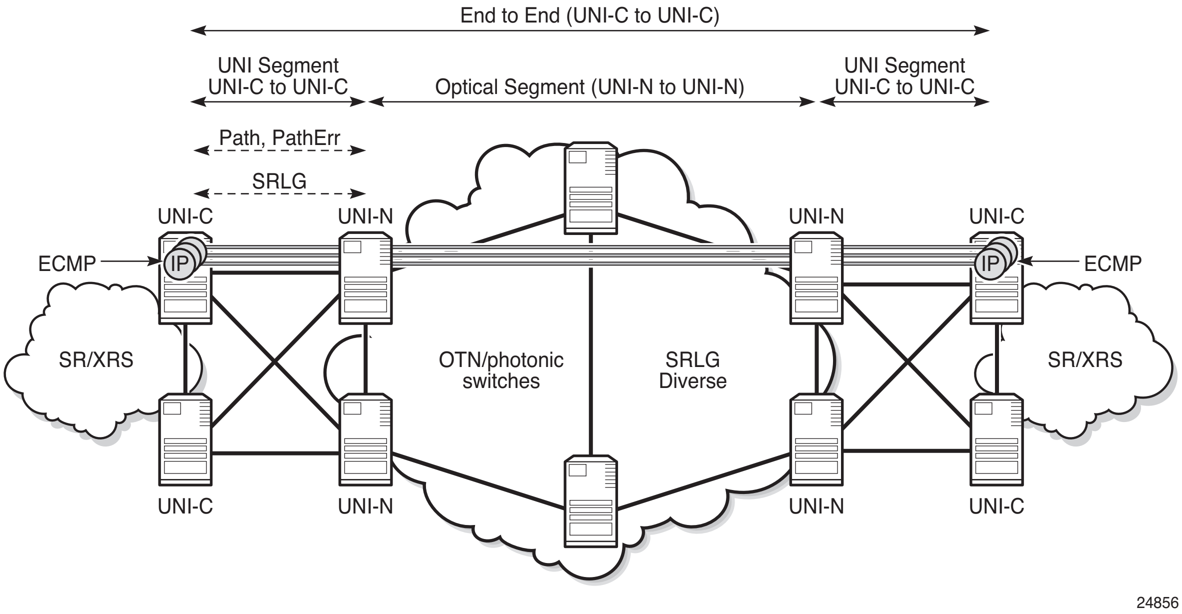
3.4.1 End to End Recovery (IP-layer)

End to end recovery applies to protection against failures at any point along the entire path between a local UNI-C and a far end UNI-C. In the context of the GMPLS UNI, recovery can be implemented in the overlay IP network either at Layer 3 or Layer 2, with assistance from the underlay optical network, with optional additional protection and/or restoration of gLSPs by GMPLS.

3.4.2 End to End ECMP

Figure 48 illustrates the first model. Multiple gLSPs are established between a pair of remote UNI-C nodes. Each gLSP is bound to a separate IP network interface at the UNI-C. RSVP signaling across the UNI is used to ensure that the gLSPs are SRLG diverse (by explicitly signaling the SRLG list to avoid in an XRO for every gLSP, or automatically collecting the SRLG list for a gLSP which does not have an XRO, and then signaling a subsequent gLSP including this collected list in its XRO). Protection is provided at the IP layer by hashing across the IP network interfaces associated with each gLSP. The operational state of each IP interface can be tied to the operational state of its gLSP (controlled using RSVP) or using mechanisms in the IP overlay such as BFD.

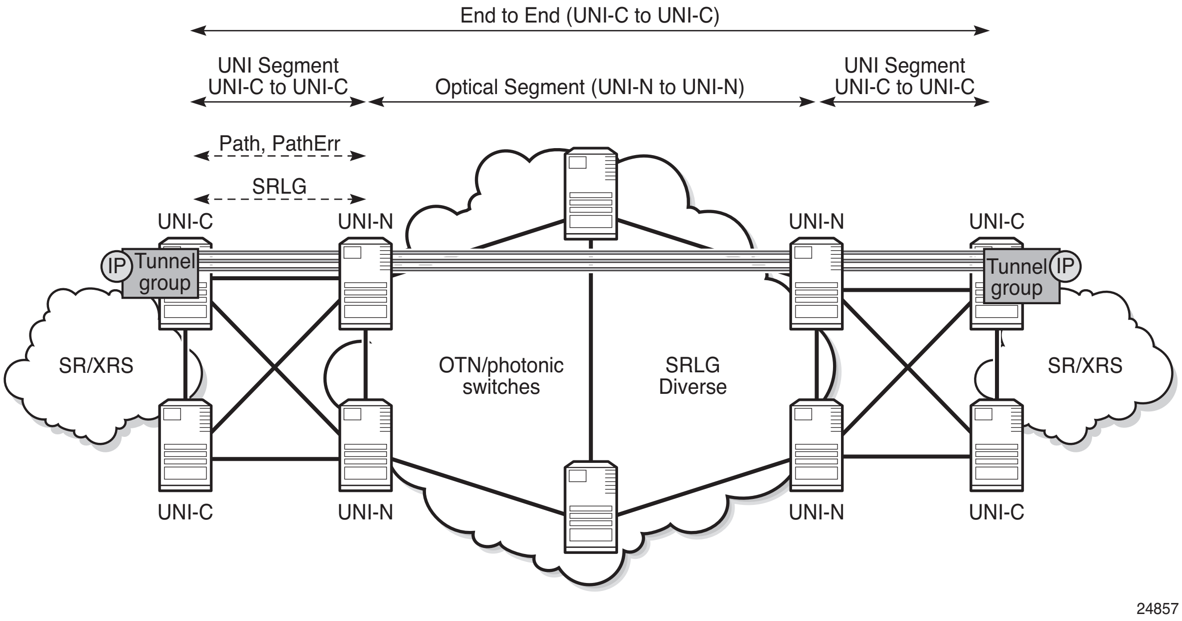
Figure 48 **End-to-End ECMP with gLSP Diversity Across Single UNI-C**



3.4.3 End to End Load Sharing Using a Load Sharing GMPLS Tunnel Group

Figure 49 shows the case where multiple gLSPs, instantiated as black and white Ethernet ports, are bundled together in a similar manner to LAG, using a GMPLS tunnel group. That is, each member gLSP of a tunnel group effectively maps to a member port, which runs end to end between remote UNI-Cs. Note that a LAG does not and cannot terminate on the neighboring 1830 PSS UNI-N. A single IP network interface is bound to the bundle of ports represented by the gLSPs. LACP does not run across the bundle; RSVP signaling is instead used to convey the state of the gLSP and thus the corresponding member port of the tunnel group. Traffic is load shared across the tunnel group members.

Figure 49 **End-to-End Load Sharing GMPLS Tunnel Group with gLSP Path Diversity**

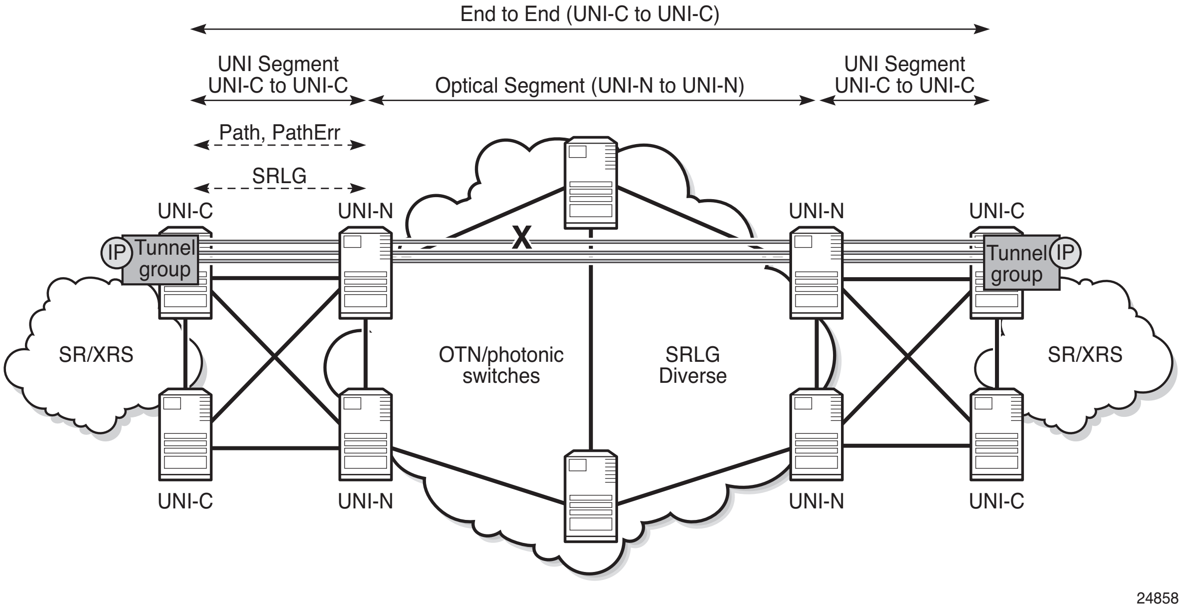


3.4.4 End to End Recovery (GMPLS Layer)

3.4.4.1 Unprotected gLSP

The default level of E2E recovery is unprotected. In this case, a gLSP can only recover from a failure when the downstream resource that failed is recovered. Figure 50 illustrates this. When a gLSP fails in the optical network, a failure notification is propagated to the source UNI-C node e.g. using a PathErr or a NotifyErr LSP Failure message. The source UNI-C node takes no action, but will continue to refresh the PATH message for this gLSP, which may be rerouted around the failure by the optical network e.g. if the IGP in the optical network reconverges. The gLSP is treated as operationally down until a message indicating that the gLSP has been restored is received by the router. For example, a Notify Error LSP Restored.

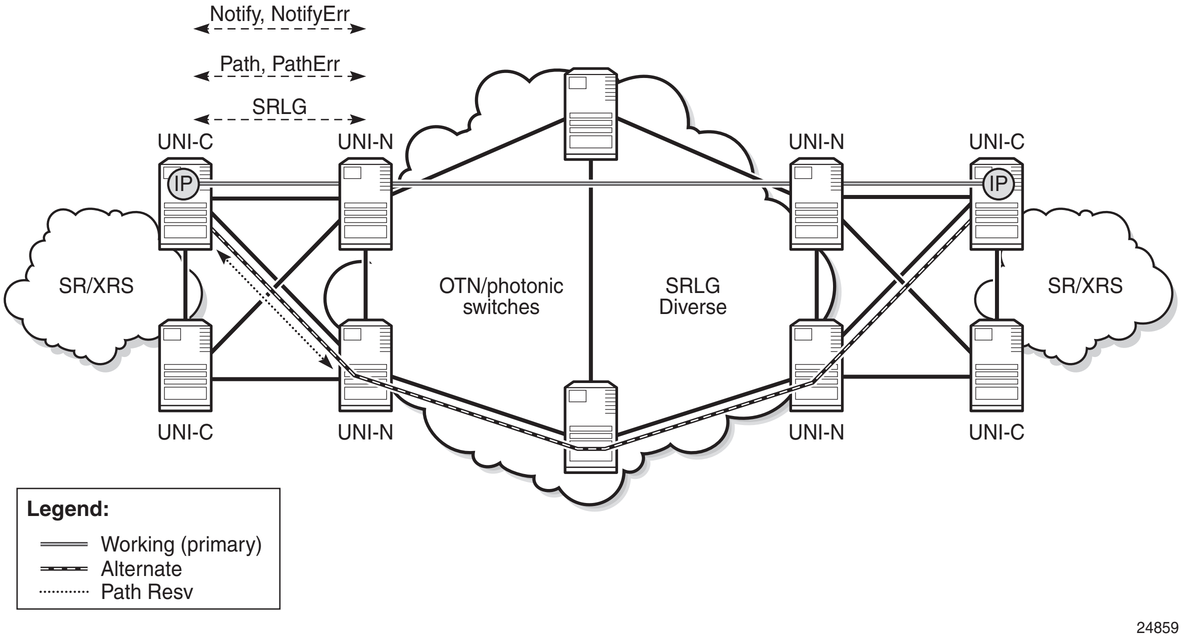
Figure 50 **gLSP Re-Establishment (PATH Refresh)**



3.4.4.2 Full LSP Rerouting

Full LSP rerouting (or restoration), specified in RFC 4872 section 11, switches normal traffic to an alternate LSP that is not even partially established until after the working LSP failure occurs. The new alternate route is selected at the LSP head-end node; it may reuse resources of the failed LSP at intermediate nodes and may include additional intermediate nodes and/or links.

Figure 51 **Full LSP Rerouting**



3.4.4.3 1: N Protection

In 1:N (N $\geq$1) protection, the protecting LSP path is a fully provisioned and resource-disjoint LSP path from the N working LSP paths. The N working LSP paths may also be mutually resource-disjoint. Coordination between end-nodes is required when switching from one of the working paths to the protecting path. Although RFC4872 allows extra traffic on the protecting path, this is not supported by the 7750 SR or 7950 XRS. Figure 52 illustrates this protection architecture when N=1, while Figure 53 shows the case for N>1.

Figure 52 **1:N Protection, with N=1 (RFC4872)**

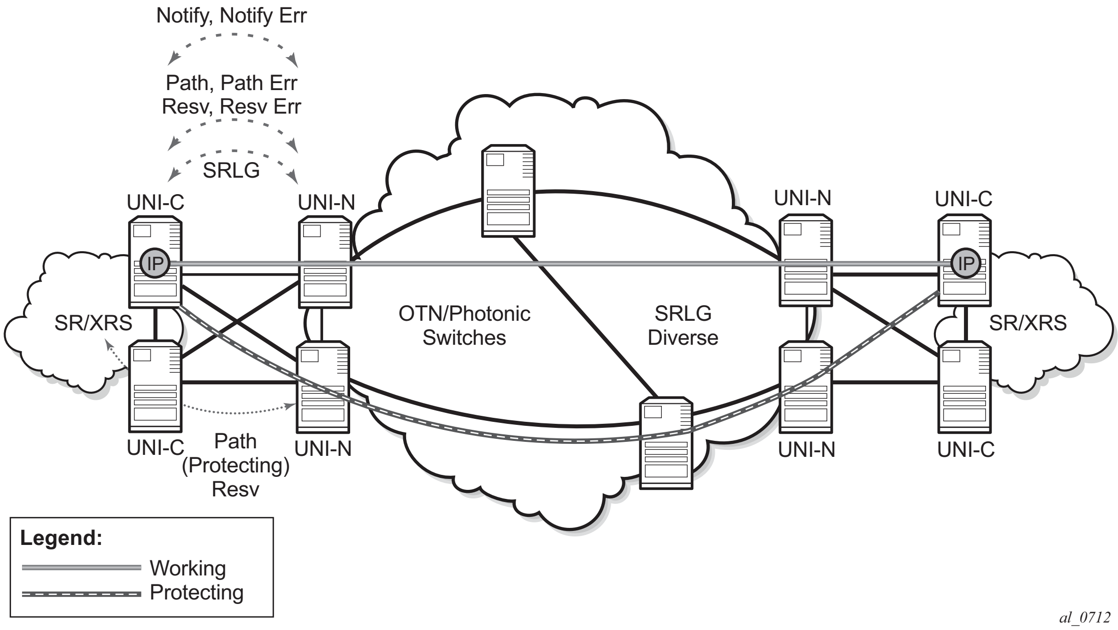
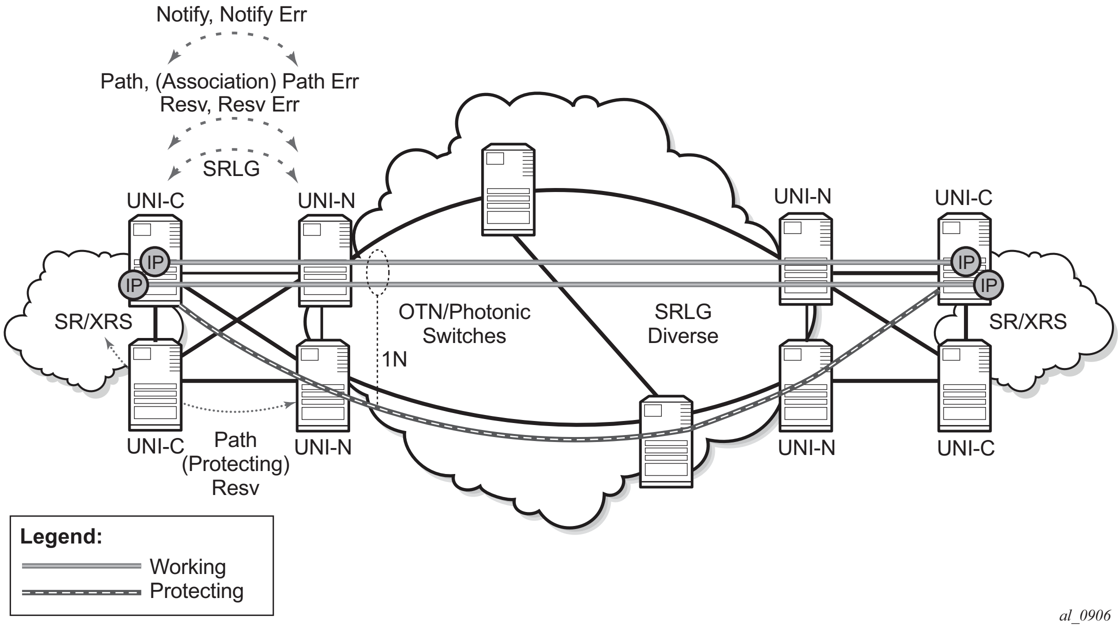


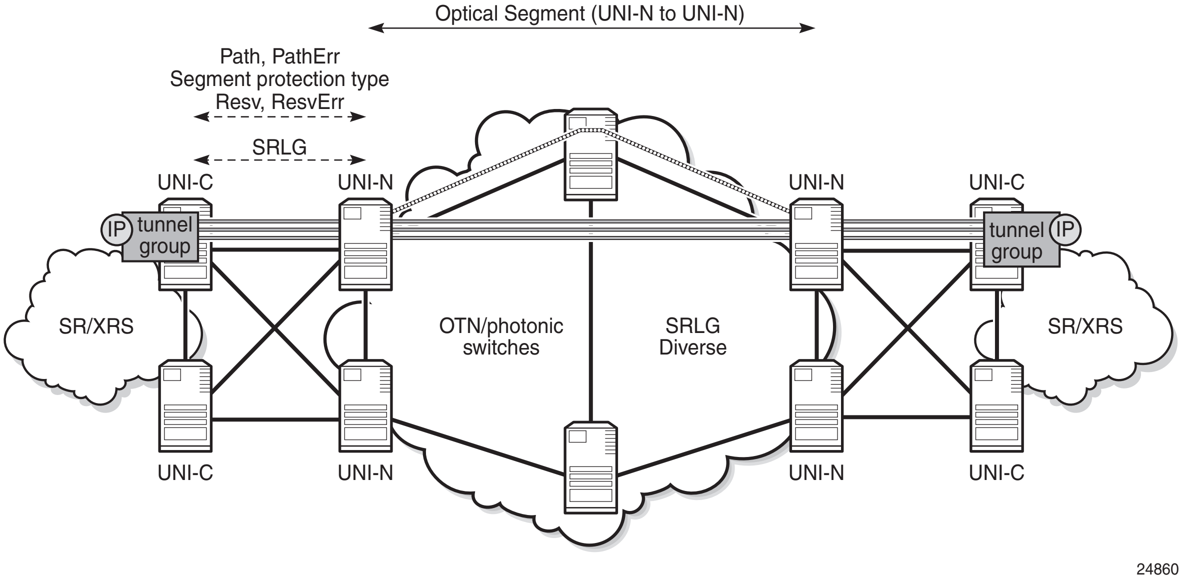
Figure 53 **N>1 Protection**



3.4.4.4 Optical Segment Recovery

Optical segment protection refers to the ability of the optical network to protect the span of a gLSP between ingress and egress UNI-N nodes. It does not require any protection switching on the UNI-C nodes. However, it does require the UNI-C to signal a request for a particular segment protection type towards the UNI-N in the PATH message for a gLSP. The optical network may either accept this request, reject it or respond with an alternative. Segment protection is defined in RFC 4873.

Figure 54 **Optical Segment Protection Domain**



Signaling of the following segment protection types is supported by the 7750 SR and 7950 XRS:

- Unprotected — The path is not protected against failure.
- Source-Based Reroute (SBR) — In this mechanism (also known as Full Rerouting), a path is restored after a failure, but the success of restoration depends on the available resources. This can reroute traffic in 200 ms or more.
- Guaranteed Restoration (GR) — A shared backup is assigned to the path, and recovery resources are reserved. If they cannot be reserved on a shared path, then this falls back to SBR. This can reroute traffic in 50 ms or less. This mechanism is also known as 1+shared standby. This is also known as Rerouting without extra traffic, or shared mesh restoration.
- Sub-network Connection Protection (SNCP) — This provides 50 ms protection in the case of a single failure. This is also known as 1+1 bidirectional path protection.

- Path Restoration Combined (PRC) — This provides 50 ms protection, even in the case of multiple failures. This is also known as SNCP with SBR.

3.5 GMPLS Configuration Overview

The Generalized Multi-Protocol Label Switching (GMPLS) User to Network Interface (UNI) permits dynamic provisioning of optical transport connections between IP routers and optical network elements in order to reduce the operational time and administrative overhead required to provision new connectivity.

3.6 LMP and IPCC Configuration

3.6.1 Configuration of IP Communication Channels for LMP and RSVP

Configuration starts with enabling the IP Communication Channel (IPCC) between the 7750 SR or 7950 XRS UNI-C and the adjacent UNI-N. The IPCC is a data communication channel for LMP and RSVP. For each different UNI-C and UNI-N adjacency, a different IPCC must be configured.

A numbered network IP interface is bound to the port connected to the DCN or directly to the 1830 PSS.

GMPLS protocols use a new loopback address type, called a GMPLS loopback, on the IPCC. The address of this loopback is termed the local GMPLS router ID. Packets that do not belong to a GMPLS protocol that are destined for this loopback address will be dropped. An interface is configured as a GMPLS loopback using the **interface** *interface-name* **gmpls-loopback** command.

CLI Syntax:

```
config
   router
      interface local-gmpls-router-id-name gmpls-loopback
         address local-gmpls-loopback-address
```

The destination address of the LMP and RSVP control plane packets should be set to the LMP/GMPLS loopback of the 1830 PSS. The 1830 PSS does that via a dedicated subnet on a VLAN interface on the management port. Another VLAN extends a separate subnet for management purposes. On the 7750 SR or 7950 XRS, the LMP and RSVP control plane packets should be sent to the next-hop for the GMPLS/LMP loopback address of the neighboring 1830 PSS. The 1830 PSS and the GMPLS router IDs must be in the same subnet. It is possible to operate over a routed DCN if the RSVP control plane messages do not set the IP router alert bit. Otherwise only direct IP connectivity over a Layer 2 network will work.

If the IPCC goes down, an existing TE Link or gLSP to a given peer UNI-N node is not torn down just because the IPCC is down. However, if the IPCC is down, then it is not possible to establish new gLSPs or TE Links, and a trap indicating a degraded state is raised.

The IPCC can use GRE encapsulation. This may be required in some network deployments where a routed DCN is used and is shared between multiple applications, in order to conceal GMPLS control plane traffic.

GRE encapsulation requires that a controlTunnel loopback interface representing the GRE tunnel is configured using the **interface** *interface-name* **control-tunnel** command. One IP tunnel can then be created on this interface. The local end tunnel IP address is configured using the interface primary IP address. The remote end tunnel IP address can be configured using the **ip-tunnel** command. GRE encapsulation is used by default for the IP tunnel.

Only the primary IPv4 interface address and only one IP tunnel per interface are allowed. Up to four tunnels can be configured using multiple controlTunnel loopback interfaces.

A static route is required to take the new controlTunnel interface as a next hop.

Note: GRE may be configured for IPCCs using a network interface or CPM port.

The following example illustrates the commands required to enable GRE tunneling on IP control channels to a given peer UNI-N.

In this example, an IPCC is established between the 7750 SR (10.20.40.40) and the 1830 PSS (10.20.50.50). Packets destined for 10.20.50.50 will match a static route pointing to interface “myTunnelItf1”, which is a controlTunnel loopback interface. When this interface is matched as a next hop, the system will add GRE encapsulation (in the CPM) to the packet and send it out using the source address 10.3.30.10 and destination address 10.3.30.1 for the tunnel (outer) IP header.

```
configure router "Base"|<cpm-vr-name>          -> cpm-vr-name "management" not
                                                  supported.
       interface "ipcc" gmpls-loopback
           address 10.20.40.40/32
       exit
       interface "myTunnelItf1" control-tunnel -> new ifType: controlTunnel(32)
           address 10.3.30.10/32               -> tunnel local address
           ip-tunnel
               remote-ip 10.3.30.1             -> tunnel remote address, gre encap
                                                  implicit
               ...                             -> future commands may be added,
                                                  e.g. encap (default will be gre)
           exit
           no shutdown                         -> interface is operationally up
                                                  only if remote-ip is reachable
       exit
       static-route-entry 10.20.50.50/32       -> static route pointing to
                                                  IPCC remote
end
           next-hop "myTunnelItf1"             -> interface of new controlTunnel
                                                  ifType can be configured as
                                                  next-hop

               no shutdown
           exit
```

```
        exit
        static-route-entry 10.3.30.1/32         -> eventually static route to reach
                                                   tunnel remote end may be needed
            next-hop 10.3.10.1
                no shutdown
            exit
        exit
    exit
```

3.6.2 Configuring LMP

LMP is used to establish and maintain an IPCC between adjacent peers, as well as to correlate the local and remote identifiers for the TE links that it controls. Some attributes must be configured locally on a per-peer basis, such as the LMP peer information, TE link information, and per-peer protocol related parameters.

The **config**>**router**>**lmp**>**peer** *peer-cp-node-id* command creates a context per LMP peer. The *peer-cp-node-id* parameter specifies the control plane identifier of the adjacent UNI-N. It is an IPv4 or unsigned integer-formatted address that is used by the UNI-C for LMP and RSVP-TE messages if a peer-loopback address is not subsequently configured. The local GMPLS router ID is used as the source address.

A static route must have previously been configured to this peer router ID. Dynamic routing (for example, using OSPF over the IPCC in order to resolve routes to the peer GMPLS router ID) is not supported. The local loopback address to use as the local GMPLS router ID should also be configured.

The LMP messages are sent over the interface corresponding to the IPCC that has been configured previously. The LMP session can be associated with one or more TE links that have been configured previously.

A control channel to an LMP peer is configured using the **config**>**router**>**lmp**>**lmp-peer** *peer-cp-node-id*>**control-channel** command. Control channels are indexed using the *lmp-cc-id* parameter, which corresponds to the lmpCcId object in the LMP MIB.

The following CLI tree illustrates the key commands for configuring LMP.

CLI Syntax:

```
config
   router
      [no] lmp
         [no] te-link te-link-id
            link-name te-link-name
            remote-id id
            [no] data-bearer data-bearer-id
               port port-id
      remote-id id
      [no] shutdown
    [no] shutdown
  gmpls-loopback-address local-gmpls-loopback-
    address
  [no] peer peer-cp-node-id
    peer-loopback-address peer-loopback-address
    retransmission-interval interval
    retry-limit limit
    [no] control-channel lmp-cc-id
      peer-interface-address ipcc-destination-
        addr
      hello interval interval dead-
        interval interval
      passive
      [no] shutdown
    te-link te-link-id
    [no] shutdown
  peer lmp-peer-address
  ...
  [no] shutdown
[no] shutdown
```

If **peer-loopback-address** is entered, then this is used as the routable peer address, otherwise the *peer-cp-node-id* is assumed to correspond to a routable peer loopback.

The **peer-interface-address** is mandatory and is the destination address of the IPCC on the peer UNI-N used to reach the GMPLS Router ID of the peer. It corresponds to the lmpCcRemoteIpAddr in RFC 4631. If the **peer-interface-address** is used as the destination IP address in the IP packet on the IPCC, then the router local interface address is used as the source IP address.

A **te-link** is configured under **config**>**router**>**lmp**>**te-link**. The **te-link** parameter under **config**>**router**>**lmp**>**peer** then assigns the control of the TE-links to the LMP protocol to a given peer. Each TE-Link can only be assigned to a single LMP peer.

The LMP protocol-specific attributes such as timers and retransmission retries are configured for each LMP peer under **config**>**router**>**lmp**>**peer**.

The **hello interval** ranges from 1000 to 65 535 ms. The default hello interval is 1000 ms.

The **hello dead-interval** ranges from 3000 to 65 535 ms. The default hello dead interval is 4000 ms.

The **retransmission-interval** ranges from 10 to 4 294 967 295 ms in 10-ms intervals, with a default of 500 ms.

The **ttl** command allows the user to configure the TTL of the IP control channel for RSVP and LMP packets to a value other than 1 (default). The range of values is 2 - 255. This enables multi-hop data communication networks between the UNI-C and UNI-N.

In order to configure an IPCC to a specific LMP peer to use an OES Ethernet port on the CPM, then the configuration must refer to a GMPLS loopback IP address that exists within a virtual management router that has an interface on that CPM Ethernet port. The IPCC to a specific LMP peer is created within a named management virtual router as follows:

CLI Syntax:

```
config>router>lmp
   peer peer-node-id
      control-channel-router router-name
            gmpls-loopback-address ipv4-address
```

The default router instance is “Base”.

The *router-name* parameter specifies the 64-byte name of a virtual router instance.

3.6.3 Configuring Traffic Engineering Links and Data Bearers

Traffic engineering (TE) links are configured under the **config**>**router**>**lmp** with a specific command, **te-link**, to create a specific context to hold TE specific configuration information pertinent to the local and remote identifiers, and physical resources assigned to the te-link. Only one data bearer per TE link is supported.

The te-link association is the creation of an association between a TE-link and data-bearing physical ports. Under the TE-link context, different data bearers can be configured via the data-bearer command. The data bearer is assigned a complete physical port, using
port<x/y/z> (slot-number/MDA-number/port-number) as input.

Note that a data bearer cannot be associated with a port in a LAG.

A TE-link has a unique *link-id*, which identifies it in RSVP-TE signaling.

The remote-id is the unnumbered link identifier at far-end of the TE link as advertised by the LMP peer that is the UNI-N.

The TE-link has associated physical resources which are assigned to the TE-link by configuring the data-bearer under the **config**>**router**>**te-link** context.

The operator must also configure the remote data-bearer link identifier under the data bearer subcontext.

Note that LMP does not correlate the local and remote Layer 2 interface identifiers (such as MAC addresses) for the data bearer. It only correlates the local and remote TE Link and Data Bearer link identifiers. The association between the Layer 2 interface address and the data bearer must be correctly configured at the UNI-C and UNI-N. The **show**>**router**>**lmp**>**te-link** command displays the local link ID, remote link ID, and associated port ID to assist with this.

The CLI tree for creating TE Links under LMP is as follows. Note that there are also some RSVP-specific TE Link parameters that are configured under a separate **gmpls** context (see below):

```
config
   router
      [no] lmp
        [no] te-link te-link-id
           link-name te-link-name
           remote-id id
           [no] data-bearer data-bearer-id
             port port-id
             remote-id id
             [no] shutdown
           [no] shutdown
        [no] shutdown
```

The *te-link-id* can take the form of an unsigned integer or 64 character (max) name: [1 to 2147483690] | *te-link-name*: 64 char max

Upon creation, only the unsigned integer needs to be specified. Once the link is created the user can configure the link name (for example, **link-name** *te-link-name*). From here, the user can refer to this te-link by either the unsigned integer or the ASCII name.

Note that LMP will normally assume a data bearer is operationally up, even if no MAC layer or a valid PCS IDLE stream is received. This is because a neighboring UNI-N may not generate a PCS IDLE stream and instead transparently transports the MAC layer from the far end, which won't be up unless a gLSP is configured. In order to prevent LMP from using a port for which there is a local fault on the data bearer, indicated by loss of light, a user must configure **report-alarm** on the Ethernet port, as follows:

config>**port**>**ethernet**>**report-alarm signal-fail**

Only ports with **report-alarm signal-fail** configured can be included in LMP, and that **report-alarm signal-fail** cannot be subsequently removed from a port in LMP.

RSVP requires that all traffic engineering attributes for TE Links are configured under the **config**>**router**>**gmpls**>**te-link** context.

```
config
   router
      [no] gmpls
            te-link te-link-id
               [no] shutdown
```

where *te-link-id*: [1..2147483690] | *te-link-name*: 32 char max

If a path (also refer to the description of a GMPLS path configuration, below) without an explicit te-link for the first hop is configured, the system will automatically select a TE Link to use for a gLSP path based on the lowest available TE Link ID with a matching bandwidth (if a bandwidth is configured for the gLSP). During a data-bearer link allocation request, an RSVP -requested gLSP BW could be either a non-zero value as per RFC 3471 signal-type (see below), or it could be zero. These are the following cases.

Case 1: Requested BW is non-zero as per RFC 3471 Signal-type configuration

- When a TE (or TE/DB) link is configured in the related hop LMP checks whether the related port BW is the same (exact match) as the requested BW, and allocates the port (provided any other checks are successful).
- When the related Hop is empty, LMP finds a db-link port to the peer with a matching the requested BW, and allocates it.

Case 2: Requested BW is Zero

- When TE (or TE/DB) link is configured in the related hop, LMP allocates the port (provided the other checks are OK), and provides the port BW to RSVP to use in signaling.
- When the related Hop is empty, LMP finds the first available db-link to the peer (based on lower db-link Id), and allocates it and provides the port BW to RSVP to use in signaling.

3.7 Configuring MPLS Paths for GMPLS

To establish an end-to-end connection between two 7750 SR or 7950 XRS nodes through a GMPLS network, a path is required, which is configured via the **config**>**router**>**gmpls**>**path** *path-name* context.

The path context consists of a set of numbered entries, each entry representing a resource that the gLSP must follow. The te-link ID is the ID allocated at the node referred to in the hop.

When interoperating with the Nokia 1830 PSS, at least the first and penultimate hops of the gLSP should be included.

The following CLI tree is used to configure a gLSP path:

```
config>router>gmpls
   path path-name
   no path path-name
     hop hop-index node-id node-id [te-link te-link-id]
               [strict | loose]
     no hop hop-index
     no shutdown
     shutdown
```

where:

node-id: IPv4 address a.b.c.d | 1830-data-plane-node-id 32-bit unsigned integer

In general, the 7750 SR or 7950 XRS node is able to populate the ERO with every hop along the gLSP path from ingress UNI-N to egress UNI-C. However, normally only a loose path across the optical network (from ingress UNI-N to egress UNI-N) is required because the optical network is responsible for path selection between ingress and egress UNI-N. Therefore the user will normally just configure hop 1 and hop 4 in the above example. For interoperability with the 1830 PSS, the user must configure a TE Link ID to use on the final hop in the ERO towards the destination UNI-C.

The following example shows how the Path should be configured for interoperability with the 1830 PSS.

Consider the following topology:

A B C D E F

[unic1]------[unin1]-----------[unin2]------[unic2]

where A-F are the TE Link IDs assigned at each end of a link.

Path configuration on unic1:

Hop 1 unic1 A strict

Hop 2 unin2 E loose

3.8 Configuring RSVP in GMPLS

RSVP-TE must be enabled on the SR OS towards the adjacent UNI-N in order to configure a GMPLS label-switched path (gLSP).

RSVP parameters specific to GMPLS are configured under the **config**>**router**>**gmpls** context.

This creates a new instance of RSVP for use in GMPLS signaling.

Global parameters for GMPLS are configured as follows:

```
config
   router
      gmpls
      no gmpls
         peer peer-cp-node-id
         gr-helper-time max-recovery recovery-interval max-restart restart-interval
         no gr-helper-time
         keep-multiplier number
         no keep-multiplier
         no rapid-retransmit-time
         rapid-retransmit-time hundred-milliseconds
         no rapid-retry-limit
         rapid-retry-limit limit
         no refresh-time
         refresh-time seconds
         no refresh-time
         lsp-init-retry-timeout seconds
         no lsp-init-retry-timeout
         no shutdown
         shutdown
```

The default max-restart interval for GMPLS is 180 s.

The LMP Peer is configured under **config**>**router**>**gmpls**>**peer** *peer-cp-node-id*, where the *peer-cp-node-id* is control plane identifier of the adjacent optical cross connect (UNI-N node). RSVP uses the destination address returned by LMP for this peer control plane node ID as the destination address, and the loopback address referenced under **config**>**router**>**lmp**>**gmpls-loopback-address** *local-gmpls-loopback-address* as the local router ID to use for the session.

RSVP will come up if at least one IPCC is up.

RSVP hellos and support for graceful restart helper functionality are supported. RSVP Graceful Restart Helper procedures implemented by the router also apply when the IPCC goes down and comes back up, or when the neighboring peer control plane restarts.

The following CLI tree is used for configuring RSVP parameters for each LMP peer:

```
config
    router
        gmpls
            peer peer-cp-node-id
            no peer peer-cp-node-id
               lsp-hold-timer hold-timer
               no lsp-hold-timer
               hello-interval milliseconds
               no shutdown
               shutdown
```

The per-peer **lsp-hold-timer** *hold-timer* parameter is used to configure a node-wide hold-down time. This timer is started when a RESV for a new gLSP is first received, or a failed gLSP path is restored (or the router is notified of a restoration following segment recovery) in order to give the optical network time to program its data path. The value range is 5 to 300 s, with a default of 60 s. A member of a GMPLS tunnel group is not considered up until the hold-timer has expired. Note that different optical network technologies have different data path programing/setup times.

Note that the **no hello-interval** command sets the hello-interval to the default value of 3000 ms. Configuring **hello-interval 0** will disable hellos in GMPLS.

3.9 Configuring a GMPLS LSP on the UNI

A GMPLS LSP is configured under **config**>**router**>**gmpls**>**lsp** *name* **gmpls-uni**. The optional **gmpls-uni** keyword indicates that the LSP is an RSVP signaled GMPLS LSP, which is profiled for the GMPLS UNI that is it uses the set of functions and CLI commands applicable to an overlay gLSP, rather than a peer model gLSP. Only overlay model gLSPs are supported; this is the default type of GMPLS LSP. The router can only act as an LER terminating a gLSP, and cannot switch a GMPLS that is it cannot act as a GMPLS LSR

GMPLS LSPs use the working path and protect path terminology from RFC 4872. Each gLSP configuration is composed of a working path and an optional protect path if end-to-end recovery is used.

Note that on-the-fly changes to an LSP or LSP path configuration are not allowed. This is because MBB is not supported for gLSPs. The LSP or LSP Path must be shut down to make configuration changes.

A GMPLS LSP (gLSP) is configured using the following CLI tree:

```
config
   router
      gmpls
         lsp lsp-name [gmpls-uni]
         no lsp lsp-name
            to remote-uni-c-gmpls-router-id
            switching-type {dcsc}
            no switching-type
            encoding-type {line}
            no encoding-type
            generalized-pid {ethernet}
            no generalized-pid
            e2e-protection-type {unprotected | 1toN | sbr}
            no e2e-protection-type
            protect-path path-name
            no protect-path path-name
                peer peer-gmpls-router-id
                no peer
                bandwidth signal-type rfc3471-name
                no bandwidth exclude-srlg group-name [group-name...(upto 5 max)]
                no exclude-srlg
                segment-protection-type {unprotected | sbr | gr | sncp | prc}
                no segment-protection-type
                no shutdown
                shutdown
            revert-timer timer-value //1 to 1800 seconds, default 0
            no revert-timer
            retry-limit limit
            no retry-limit
            no shutdown
            shutdown
            working-path path-name
            no working-path path-name
```

```
        bandwidth signal-type rfc3471-name
        no bandwidth
        exclude-srlg group-name [group-name...(upto 5 max)]
        no exclude-srlg
        peer peer-gmpls-router-id
        no peer bandwidth
        segment-protection-type {unprotected | sbr | gr | sncp | prc}
        no segment-protection-type
        no shutdown
        shutdown
    no shutdown
    shutdown
```

The loopback address of the remote router (UNI-C) must be configured after the **to** keyword and takes an IPv4 address as input.

The **switching-type** indicates the type of switching required for the gLSP. This can take a number of values, as defined in RFC 3471, and extended in RFC 6004 and RFC 7074 for Ethernet VPL (EVPL) services. The default CLI value is **DCSC**. This is the only supported value.

The **encoding-type** configuration specifies the encoding type of the payload carried by the gLSP. **line**, indicating 8B/10B encoding, is the only supported type.

The **generalized-pid** parameter specifies the type of payload carried by the gLSP. Standard ethertype values are used for packet and Ethernet LSPs (see RFC 3471). Only Ethernet (value 33) is supported.

Note that gLSPs are inherently bidirectional. That is, both directions of the gLSP are bound together. The destination UNI-C node will automatically bind an incoming gLSP PATH message to the corresponding egress direction based on the session name in the session object.

Any gLSP that needs to be bound to a specific TE Link (as referred to in the pPATH), will only be allowed if the corresponding TE Link exists under **config**>**router**>**gmpls**. Constraints such as HOP definition, SRLG, BW, and so on, will be checked before signaling the gLSP.

Since RSVP signaling operates out of band, refresh reduction is not supported. RSVP authentication is not supported on the 1830 PSS UNI-N, but MD5 authentication is implemented.

A configurable **retry-timer** is supported.

A configurable **retry-limit** for each gLSP is supported, with a range of 0 to 10000, and a default of 0.

The **working-path** and **protect-path** command allows paths to be configured for the gLSP. At least a **working-path** must be configured, although the path-name that it references may contain an empty path. The optional **working-path>peer** and **protect-path>peer** commands allow the user to specify a first hop UNI-N node to use for the gLSP path. The protect path is only configurable for 1:N recovery option.

Reversion from the protect path to the working path is supported.

RSVP uses the Fixed Filter (FF) style of RESV. The signaled MTU is hard-coded to 9212 bytes, as appropriate for Ethernet gLSPs.

The default **setup** and **hold** priorities are 5 and 1, respectively, and cannot be configured. gLSP preemption is not supported.

Record and **record-label** are enabled by default and no user configurable command is therefore provided.

3.9.1 gLSP Constraints

Each gLSP can be configured with the following constraints:

- Bandwidth
- SRLG
- Protection

3.10 Bandwidth

The bandwidth associated with a gLSP is configured with the bandwidth command, and can take the RFC 3471 signal type name as input.

The signaled bandwidth is then used for path computation and admission in the GMPLS domain.

By default, the actual interface bandwidth is used. If the user configures a bandwidth greater than the local data bearer bandwidth, then the gLSP establishment will be blocked. If the user configures a bandwidth less than or equal to the local data bearer bandwidth, then that bandwidth is signaled to the UNI-N.

The bandwidth required for the LSP is configured under the path context as follows. Note that the system will do an exact match check of the gLSP bandwidth against the data bearer bandwidth:

```
config
   router
      gmpls
         lsp gmpls-tunnel-name [gmpls-uni]
            to remote-uni-c-gmpls-router-id
            working-path path-name
               bandwidth signal-type rfc3471-name
```

The possible signal-type values are:

ds0 | ds1 | e1 | ds2 | e2 | ethernet | e3 | ds3 | sts-1 | fast-ethernet | e4 | fc-0-133m | oc-3/stm-1 | fc-0-266m | fc-0-531m | oc-12/stm-4 | gige | fc-0-1062m | oc-48/stm-16 | oc-192/stm-64 | 10gige-ieee | oc-768/stm-256 | 100gige-ieee

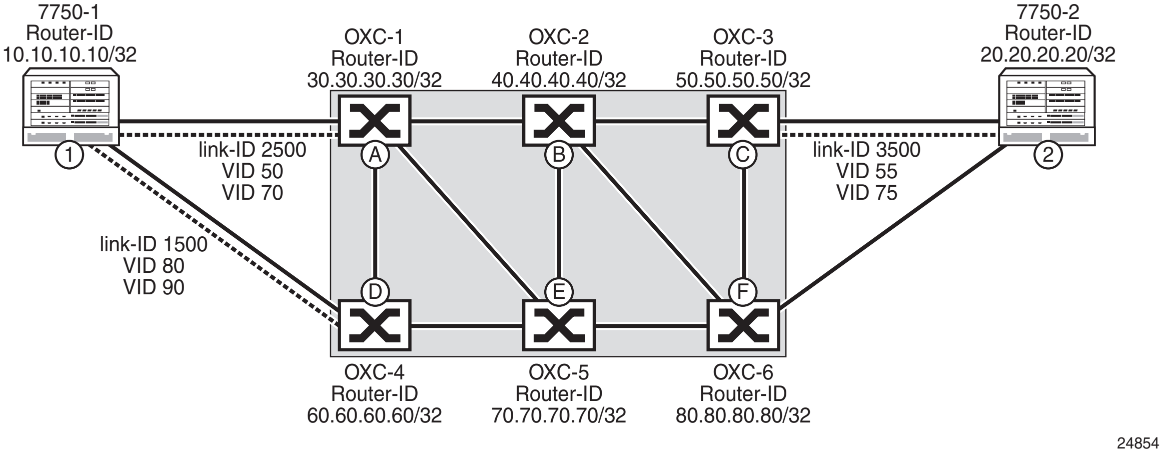
The code points to use for 10gige-ieee and 100gige-ieee are not yet registered with IANA. The following values are therefore used:

- 10G IEEE: 0x4E9502F9
- 100G IEEE: 0x503A43B7

3.11 Shared Risk Link Groups

Shared Risk Link Groups (SRLG) are used in the context of a gLSP to ensure that diverse paths can be taken for different gLSPs through the optical network. For example, consider the network shown in Figure 55:

Figure 55 **SRLG Example**



In this dual-homing scenario, the primary gLSP takes TE-Link 1-A, and C-2, while the secondary gLSP path takes TE-Links 1-D and F-2. In order to ensure that a failure in the underlying optical network does not affect both the primary and secondary paths for the gLSP, the SRLG list used by the optical network for the primary path is shared with the UNI-C (1) by the UNI-N (A) at the time the gLSP is established along the primary path. When the secondary path is signaled, the UNI-C (1) will signal the SRLG list to avoid to the UNI-N (D). Note that a similar procedure is beneficial even if a UNI-C is not dual homed to the optical network, but diverse primary and secondary paths are required through the optical network.

The 7750 SR and 7950 XRS routers support two methods for indicating a set of SRLGs to exclude:

- Explicit configuration of an SRLG list for a gLSP path. These are signaled in the XRO of the RSVP PATH message towards the optical network
- Automatic SRLG collection for a gLSP, using the procedures specified in draft-ietf-ccamp-rsvp-te-srlg-collect-04.txt, and operate as follows:
  - Retrieving SRLG information from a UNI-N for an existing gLSP Path — When a dual-homed UNI-C device intends to establish a gLSP path to the same destination UNI-N device via another UNI-N node, it can request the SRLG information for an already established gLSP path by setting the SRLG information flag in the LSP attributes sub-object of the RSVP PATH

message using a new SRLG flag. This path would be the primary path for a gLSP established by the router UNI-C. As long as the SRLG information flag is set in the PATH message, the UNI-N node inserts the SRLG sub-object as defined in draft-ietf-ccamp-rsvp-te-srlg-collect-04.txt into the RSVP RESV message that contains the current SRLG information for the gLSP path. Note that the provider network's policy may have been configured so as not to share SRLG information with the client network. In this case the SRLG sub-object is not inserted in the RESV message even if the SRLG information flag was set in the received PATH message. Note that the SRLG information is assumed to be always up-to-date by the UNI-C.

- Establishment of a new gLSP path with SRLG diversity constraints — When a dual-homed UNI-C device sends an LSP setup requests to a UNI-N for a new gLSP path that is required to be SRLG diverse with respect to an existing gLSP path that is entering the optical network via another UNI-N, the UNI-C sets a new SRLG diversity flag in the LSP attributes sub-object of the PATH message that initiates the setup of this new gLSP path. This path would be the protect path of a gLSP established by the router. When the UNI-N receives this request it calculates a path to the given destination and uses the received SRLG information as path computation constraints.

The router collects SRLG by default. SRLG collection occurs on all paths of the gLSP. The collected SRLG list is visible to the user via a **show** command. The recorded SRLGs are then used to populate the XRO. Only best effort (loose) SRLG diversity is supported.

Automated SRLG diversity is supported for the working and protect paths of the following end to end protection types:

- 1:N
- LSPs that form a part of a load sharing tunnel group

Already-established gLSPs within a load-sharing tunnel group or for which 1:N recovery is configured can be made mutually diverse by applying a **shutdown** / **no shutdown** operation. GMPLS LSPs with other types of protection can be made mutually SRLG-diverse by performing a shutdown of the gLSP, reconfiguring the SRLG list to exclude using the **exclude-srlg** command, and then applying a **no shutdown** of the gLSP.

3.12 Optical Network Segment Recovery

The router may request a particular GMPLS recovery type for a gLSP path segment that spans the optical network. This refers to the protection afforded to the gLSP path between the UNI-N nodes. The router supports the following segment protection types (code points are also shown):

- Unprotected: 0x00
- Source-Based Reroute (SBR) (Known as Full Rerouting in the IETF): 0x01
- Guaranteed Restoration (GR) (Also known as shared mesh restoration): 0x02
- Sub-network Connection Protection (SNCP) (1+1 bidirectional protection): 0x10
- Path Restoration Combined (PRC): 0x11

These resiliency options are configured under the **segment-protection-type** command for a given path.

```
config
   router
      gmpls
         lsp gmpls-tunnel-name [gmpls-uni]
            to remote-uni-c-gmpls-router-id
            working-path path-name
               [no] segment-protection-type {unprotected | sbr | gr | sncp | prc}
               ...
               [no] shutdown
```

The default **segment-protection-type** setting is **unprotected**.

If the requested protection type cannot be satisfied by the optical network, the router will generate a CLI warning and an SNMP trap.

Table 24 lists the recommended combinations of segment protection type and end-to-end protection type.

Table 24 **Combinations of End-to-End and Segment Protection**

E2E/Segment	Unprotected	SBR	GR	SNCP	PRC
Unprotected	Yes	Yes	Yes	Yes	Yes
1:1/1:N	Yes	Yes	Yes	Yes	—
Full Rerouting	Yes	—	—	Yes	—

3.13 Configuration of End-to-End GMPLS Recovery

End-to-end GMPLS recovery is configured at the LSP level using the **e2e-protection-type** command, as follows:

```
config
   router
      gmpls
         lsp gmpls-tunnel-name [gmpls-uni]
            to remote-uni-c-gmpls-router-id
            e2e-protection-type [unprotected | 1toN | sbr]
            revert-timer timer-value
```

The protection type names are common to those used in the optical network. The protection types are as follows:

- **unprotected** — 0x00
- **1toN** — 1:N protection. Extra traffic is not supported. Note that 1:1 protection is a special case of 1:N. 0x04
- **sbr** — Full LSP rerouting; 0x01

The default end-to-end protection type is **unprotected**.

It is possible to configure segment protection on a path independently of the type of end-to-end protection that is configured.

1toN protection requires the configuration of multiple working paths and a protect path for a GMPLS LSP. The working paths are then associated with different GMPLS Tunnel Groups. Configuration is as follows:

```
config
  router
     gmpls
        lsp lsp-name gmpls-uni
            to remote-uni-c-gmpls-router-id
            e2e-protection-type 1toN // Only these types are allowed for gmpls-uni
            switching-type ethernet
            encoding-type ethernet
            generalized-pid ethernet
            revert-timer timer-value
            retry-limit limit
        working-path path-name1 [lmp-peer <peer-gmpls-router-id>] ...
            [no] shutdown
            working-path path-name2 [lmp-peer peer-gmpls-router-id] ...
            [no] shutdown
        working-path path-name3 [lmp-peer peer-gmpls-router-id] ...
            [no] shutdown
        protect-path path-name4 [lmp-peer peer-gmpls-router-id] ...
            [no] shutdown
```

The LSP is then bound to one or more GMPLS tunnel groups. Load sharing or 1:N protection may be used across the working paths. The load sharing case is described below.

For the non-load sharing 1:N case, a single LSP is assigned to each tunnel group, as follows:

For the head end node:

```
config > gmpls-tunnel-group 2 create
   type head-end
   far-end remote-uni-c-router-id
   mode protection
   member 1 create
      glsp session-name lsp-name:path-name1
      no shutdown
   no shutdown
config > gmpls-tunnel-group 3
   type head-end
   far-end remote-uni-c-router-id
   mode protection
   member 1 create
      glsp session-name lsp-name:path-name1
      no shutdown
   no shutdown
config > gmpls-tunnel-group 4
   type head-end
   far-end remote-uni-c-router-id
   mode protection
   member 1 create
      glsp session-name lsp-name:path-name1
      no shutdown
   no shutdown
```

For the tail end node:

```
config > gmpls-tunnel-group 2
   type tail-end
   far-end remote-uni-c-router-id
   mode protection
   member 1 create
      glsp session-name lsp-name:path-name1
      no shutdown
   no shutdown
config > gmpls-tunnel-group 3
   type tail-end
   far-end remote-uni-c-router-id
   mode protection
   member 1 create
      glsp session-name lsp-name:path-name1
      no shutdown
   no shutdown
config > gmpls-tunnel-group 4
   type tail-end
   far-end remote-uni-c-router-id
   mode protection
```

```
member 1 create
   glsp session-name lsp-name:path-name1
   no shutdown
no shutdown
```

A shutdown of a working path does not trigger a switchover to the protect path. The user should either use the **tools**>**perform**>**router**>**gmpls force** or **manual** commands, or shut down the TE-Link, data bearer, or port associated with the gLSP path.

3.14 GMPLS Tunnel Groups

A GMPLS tunnel group is a bundle of gLSPs providing an abstraction of the data bearers that are intended to be associated to one IP interface. This object allows, for example, end-to-end load balancing across the set of data bearers corresponding to a set of gLSPs. A gLSP is bound to a GMPLS tunnel group by a gLSP tunnel (session) name at both the head end and the tail end UNI-C nodes of the gLSP. A sender address (the far-end) may optionally be configured for the tail end of a gLSP in case different head end nodes use overlapping gLSP tunnel names.

```
config
   gmpls-tun-grp gmpls-tun-grp-id
      type {head-end | tail-end}
      far-end remote-uni-c-router-id
      mode {load-sharing | active-standby}
       no mode
      [no] member-threshold threshold [action down]
      member mem-id [create]
         glsp session-name name
         no glsp session-name name
         [no] shutdown
      ...
      [no] shutdown
```

gmpls-tun-grp-id is an unsigned integer from 1 to 1024, shared with the Ethernet tunnel ID range.

The GMPLS Tunnel Group must be configured as either at both the **head-end** or **tail-end** of a set of member gLSPs (identified using the **head-end** or **tail-end** keywords). These keywords are mutually exclusive.

Nodes at the head-end initiate signaling of gLSPs. The **far-end** is the far end of the GMPLS tunnel group. If this node is a head end, then the far end address is taken as the to address for the member gLSPs. Each gLSP that is bound to the tunnel group must have a to address matching the far end address. A binding is held down if a gLSP to and the tunnel group to do not match.

Nodes at the tail end wait for the first path message for a gLSP. The **far-end-address** address must be configured at the tail end. It is the GMPLS Router ID of the head-end UNI-C (the *remote-uni-c-node-id*), and must be configured at the tail end UNI-C of a gLSP. The combination of *session-name* and *remote-uni-c-node-id* provides a unique key to bind an incoming gLSP setup request to a tunnel group. A binding to the tunnel group is held down at the tail end until a gLSP PATH message with a matching *session-name* and source address that matches the tunnel group's far-end address is received.

At the tail end, the **session-name** is composed of the LSP name and Path name as configured at the head end

If **load-sharing** is configured, then all of the gLSPs must terminate on the same far-end node. All of the ports used by gLSPs in a load-sharing must be equivalent in that they must have the same named QoS policy, bandwidth, and so on. Once more than one gLSP is associated with a tunnel group, the QoS policy/scheduler policy cannot be changed in any of the ports. All gLSPs must be unprotected end-to-end in load-sharing mode. Segment protection is allowed for gLSPs associated in load sharing mode to a GMPLS tunnel group.

In **active-standby** mode, only one member gLSP can be associated with the tunnel group.

All members of a tunnel group must be of the same bandwidth.

The **member-threshold** is the number of member gLSPs that must be operationally up before the gmpls tunnel group is considered operationally up.

A member of a GMPLS tunnel group may be treated as down for one of the following reasons. These reason codes are recorded in the tmnxGmplsTunGrpMemberTable in the MIB:

- adminDn — The member or the related tunnel-grp is administratively down.
- wpLspDn — The associated working lsp-path is down.
- wpPortDn — The data-bearer port associated with the working lsp-path is down.
- wpPortNoRsrc — The data-bearer port associated with the working lsp-path has no resource to support the services over the gmpls-tunnel-grp logical port.
- ppLspDn — The associated protect lsp-path is down.
- ppPortDn — The data-bearer port associated with the protect lsp-path is down.
- ppPortNoRsrc — The data-bearer port associated with the protect lsp-path has no resource to support the services over the gmpls-tunnel-grp logical port.

Note that in the case of wpPortNoRsrc and ppPortNoRsrc, the term 'resources' relates to QoS or ACL related resources. For example, this can happen when a subsequent physical or data bearing port is added to a GMPLS tunnel group, which already has services running over it. If the new-complex does not have the resources to support those services over that GMPLS tunnel group, the related member operState would be down with reasonCode PortNoRsrc. If a gLSP is already established on a data bearer when a resource failure is experienced, the RSVP PATH message A-Bit is updated so that both ends ensure that the LSP Path is held down.

The user should free resources from the complex, and shutdown/no shutdown the GMPLS tunnel group member. This repeats the resource check, which will bring the member operUp if it passes.

A gLSP associated with a tunnel group member will be down if the member is operationally down, or a fault is detected on the associated data bearer.

If a member is in the admin down state, a gLSP will not be set-up. If a gLSP is already up, the RSVP Path message A-Bit is updated so that both ends of the gLSP path are kept down.

3.15 Configuring IP and MPLS in an Overlay Network to Use a GMPLS LSP

IP and MPLS is able to use GMPLS LSPs as transport by bringing a numbered or unnumbered IP interface to an endpoint of one or more gLSPs. This IP interface appears as any other IP interface bound to a network port. The IP interface is bound to the GMPLS tunnel group by a GMPLS tunnel group number configured in the **port** command.

The GMPLS tunnel group number must correspond to a locally configured GMPLS tunnel group.

The following CLI tree illustrates where the GMPLS tunnel group is referenced. This must be done at the nodes at the tunnel groups at both ends of the transport service.

```
config
   router
      interface if-name
         address a.b.c.d | ipv6-address
         port gmpls-tunnel-group gmpls-tunnel-group-id
```

3.16 Configuration Notes

This section describes GMPLS caveats.

- Interfaces must already be configured in the config>router>interface context before they can be specified in GMPLS.
- A router interface must be specified in the config>router>mpls context in order to apply it or modify parameters in the config>router>rsvp context.
- A system interface must be configured and specified in the config>router>mpls context.
- Paths must be created before they can be applied to an LSP.

4 MPLS Forwarding Policy

The MPLS forwarding policy provides an interface for adding user-defined label entries into the label FIB of the router and user-defined tunnel entries into the tunnel table.

The endpoint policy allows the user to forward unlabeled packets over a set of user-defined direct or indirect next hops with the option to push a label stack on each next hop. Routes are bound to an endpoint policy when their next hop matches the endpoint address of the policy.

The user defines an endpoint policy by configuring a set of next-hop groups, each consisting of a primary and a backup next hops, and binding an endpoint to it.

The label-binding policy provides the same capability for labeled packets. In this case, labeled packets matching the ILM of the policy binding label are forwarded over the set of next hops of the policy.

The user defines a label-binding policy by configuring a set of next-hop groups, each consisting of a primary and a backup next hops, and binding a label to it.

This feature is targeted for router programmability in SDN environments.

4.1 Introduction to MPLS Forward Policy

This section provides information about configuring and operating a MPLS forwarding policy using CLI.

There are two types of MPLS forwarding policy:

- endpoint policy
- label-binding policy

The endpoint policy allows the user to forward unlabeled packets over a set of user-defined direct or indirect next hops, with the option to push a label stack on each next hop. Routes are bound to an endpoint policy when their next hop matches the endpoint address of the policy.

The label-binding policy provides the same capability for labeled packets. In this case, labeled packets matching the ILM of the policy binding label are forwarded over the set of next hops of the policy.

The data model of a forwarding policy represents each pair of {primary next hop, backup next hop} as a group and models the ECMP set as the set of Next-Hop Groups (NHGs). Flows of prefixes can be switched on a per NHG basis from the primary next hop, when it fails, to the backup next hop without disturbing the flows forwarded over the other NHGs of the policy. The same can be performed when reverting back from a backup next hop to the restored primary next hop of the same NHG.

4.2 Feature Validation and Operation Procedures

The MPLS forwarding policy follows a number of configuration and operation rules which are enforced for the lifetime of the policy.

There are two levels of validation:

- The first level validation is performed at provisioning time. The user can bring up a policy (**no shutdown** command) once these validation rules are met. Afterwards, the policy is stored in the forwarding policy database.
- The second level validation is performed when the database resolves the policy.

4.2.1 Policy Parameters and Validation Procedure Rules

The following policy parameters and validation rules apply to the MPLS forwarding policy and are enforced at configuration time:

- A policy must have either the **endpoint** or the **binding-label** command to be valid or the **no shutdown** will not be allowed. These commands are mutually exclusive per policy.
- The **endpoint** command specifies that this policy is used for resolving the next hop of IPv4 or IPv6 packets, of BGP prefixes in GRT, of static routes in GRT, of VPRN IPv4 or IPv6 prefixes, or of service packets of EVPN prefixes. It is also used to resolve the next hop of BGP-LU routes.

  The resolution of prefixes in these contexts matches the IPv4 or IPv6 next-hop address of the prefix against the address of the endpoint. The family of the primary and backup next hops of the NHGs within the policy are not relevant to the resolution of prefixes using the policy.

  See Tunnel Table Handling of MPLS Forwarding Policy for information about CLI commands for binding these contexts to an endpoint policy.
- The **binding-label** command allows the user to specify the label for binding to the policy such that labeled packets matching the ILM of the binding label can be forwarded over the NHG of the policy.

  The ILM entry is created only when a label is configured. Only a provisioned binding label from a reserved label block is supported. The name of the reserved label block using the **reserved-label-block** command must be configured.

  The payload of the packet forwarded using the ILM (payload underneath the swapped label) can be IPv4, IPv6, or MPLS. The family of the primary and backup next hops of the NHG within the policy are not relevant to the type of payload of the forwarded packets.

- Changes to the values of the **endpoint** and **binding-label** parameters require a **shutdown** of the specific forwarding policy context.
- A change to the name of the **reserved-label-block** requires a **shutdown** of the **forwarding-policies** context. The **shutdown** is not required if the user extends or shrinks the range of the **reserved-label-block**.
- The **preference** parameter allows the user to configure multiple endpoint forwarding policies with the same endpoint address value or multiple label-binding policies with the same binding label; providing the capability to achieve a 1:N backup strategy for the forwarding policy. Only the most preferred, lowest numerical preference value, policy is activated in data path as explained in Policy Resolution and Operational Procedures.
- Changes to the value of parameter **preference** requires a shutdown of the specific **forwarding-policy** context.
- A maximum of eight label-binding policies, with different preference values, are allowed for each unique value of the binding label.

  Label-binding policies with exactly the same value of the tuple {**binding label** | **preference**} are duplicate and their configuration is not allowed.

  The user can not perform **no shutdown** on the duplicate policy.
- A maximum eight endpoint policies, with different preference values, are allowed for each unique value of the tuple {**endpoint**}.

  Endpoint policies with exactly the same value of the tuple {**endpoint**, **reference**} are duplicate and their configuration is not allowed.

  The user can not perform **no shutdown** on the duplicate policy.
- The **metric** parameter is supported with the endpoint policy only and is inherited by the routes which resolve their next hop to this policy.
- The **revert-timer** command configures the time to wait before switching back the resolution from the backup next hop to the restored primary next hop within a given NHG. By default, this timer is disabled meaning that the NHG will immediately revert to the primary next hop when it is restored.

  The revert timer is restarted each time the primary next hop flaps and comes back up again while the previous timer is still running. If the revert timer value is changed while the timer is running, it is restarted with the new value.
- The MPLS forwarding policy feature allows for a maximum of 32 NHGs consisting of, at most, one primary next hop and one backup next hop.
- The **next-hop** command allows the user to specify a direct next-hop address or an indirect next-hop address.
- A maximum of ten labels can be specified for a primary or backup direct next hop using the **pushed-labels** command. The label stack is programmed using a super-NHLFE directly on the outgoing interface of the direct primary or backup next hop.

Note: This policy differs from the SR-TE LSP or SR policy implementation which can push a total of 11 labels due to the fact it uses a hierarchical NHLFE (super-NHLFE with maximum 10 labels pointing to the top SID NHLFE).

- The **resolution-type** {**direct**| **indirect**} command allows a limited validation at configuration time of the NHGs within a policy. The **no shutdown** command fails if any of these rules are not satisfied. The following are the rules of this validation:
  - NHGs within the same policy must be of the same resolution type.
  - A forwarding policy can have a single NHG of resolution type **indirect** with a primary next hop only or with both primary and backup next hops. An NHG with backup a next hop only is not allowed.
  - A forwarding policy will have one or more NHGs of resolution type **direct** with a primary next hop only or with both primary and backup next hops. An NHG with a backup next hop only is not allowed.
  - A check is performed to make sure the address value of the primary and backup next hop, within the same NHG, are not duplicates. No check is performed for duplicate primary or backup next-hop addresses across NHGs.
  - A maximum of 64,000 forwarding policies of any combination of label binding and endpoint types can be configured on the system.
- The IP address family of an endpoint policy is determined by the family of the **endpoint** parameter. It is populated in the TTMv4 or TTMv6 table accordingly. A label-binding policy does not have an IP address family associated with it and is programmed into the label (ILM) table.

  The following are the IP type combinations for the primary and backup next hops of the NHGs of a policy:
  - A primary or a backup indirect next hop with no pushed labels (label-binding policy) can be IPv4 or IPv6. A mix of both IP types is allowed within the same NHG.
  - A primary or backup direct next hop with no pushed labels (label-binding policy) can be IP types IPv4 or IPv6. A mix of both families is allowed within the same NHG.
  - A primary or a backup direct next hop with pushed labels (both endpoint and label binding policies) can be IP types IPv4 or IPv6. A mix of both families is allowed within the same NHG.

4.2.2 Policy Resolution and Operational Procedures

This section describes the validation of parameters performed at resolution time, as well as the details of the resolution and operational procedures.

- The following parameter validation is performed by the forwarding policy database at resolution time; meaning each time the policy is re-evaluated:
  - If the NHG primary or backup next hop resolves to a route whose type does not match the configured value in **resolution-type**, that next hop is made operationally “down”.

    A DOWN reason code shows in the state of the next hop.
  - The primary and backup next hops of an NHG are looked up in the routing table. The lookups can match a direct next hop in the case of the direct resolution type and therefore the next hop can be part of the outgoing interface primary or secondary subnet. They can also match a static, IGP, or BGP route for an indirect resolution type, but only the set of IP next hops of the route are selected. Tunnel next hops are not selected and if they are the only next hops for the route, the NHG will be put in operationally “down” state.
  - The first 32, out of a maximum of 64, resolved IP next hops are selected for resolving the primary or backup next hop of a NHG of **resolution-type indirect**.
  - If the primary next hop is operationally “down”, the NHG will use the backup next hop if it is UP. If both are operationally DOWN, the NHG is DOWN. See Data Path Support for details of the active path determination and the failover behavior.
  - If the binding label is not available, meaning it is either outside the range of the configured **reserved-label-block**, or is used by another MPLS forwarding policy or by another application, the label-binding policy is put operationally “down” and a retry mechanism will check the label availability in the background.

    A policy level DOWN reason code is added to alert users who may then choose to modify the binding label value.
  - No validation is performed for the pushed label stack of or a primary or backup next hop within a NHG or across NHGs. Users are responsible for validating their configuration.
- The forwarding policy database activates the best endpoint policy, among the named policies sharing the same value of the tuple {**endpoint**}, by selecting the lowest preference value policy. This policy is then programmed into the TTM and into the tunnel table in the data path.

If this policy goes DOWN, the forwarding policy database performs a re-evaluation and activates the named policy with the next lowest preference value for the same tuple {**endpoint**}.

If a more preferred policy comes back up, the forwarding policy database reverts to the more preferred policy and activates it.

- The forwarding policy database similarly activates the best label-binding policy, among the named policies sharing the same binding label, by selecting the lowest preference value policy. This policy is then programmed into the label FIB table in the data path as detailed in Data Path Support.

  If this policy goes DOWN, the forwarding policy database performs a re-evaluation and activates the named policy with the next lowest preference value for the same binding label value.

  If a more preferred policy comes back up, the forwarding policy database reverts to the more preferred policy and activates it.
- The active policy performs ECMP, weighted ECMP, or CBF over the active (primary or backup) next hops of the NHG entries.
- When used in the PCEP application, each LSP in a label-binding policy is reported separately by PCEP using the same binding label. The forwarding behavior on the node is the same whether the binding label of the policy is advertised in PCEP or not.
- A policy is considered UP when it is the best policy activated by the forwarding policy database and when at least one of its NHGs is operationally UP. A NHG of an active policy is considered UP when at least one of the primary or backup next hops is operationally UP.
- When the **config**>**router**>**mpls** or **config**>**router**>**mpls**>**forwarding-policies** context is set to **shutdown**, all forwarding policies are set to DOWN in the forwarding policy database and deprogrammed from IOM and data path.

  Prefixes which were being forwarded using the endpoint policies revert to the next preferred resolution type configured in the specific context (GRT, VPRN, or EVPN).
- When an NHG is set to **shutdown**, it is deprogrammed from the IOM and data path. Flows of prefixes which were being forwarded to this NHG are re-allocated to other NHGs based on the ECMP, Weighted ECMP, or CBF rules.
- When a policy is set to **shutdown**, it is deleted in the forwarding policy database and deprogrammed from the IOM and data path. Prefixes which were being forwarded using this policy will revert to the next preferred resolution type configured in the specific context (GRT, VPRN, or EVPN).
- The **no forwarding-policies** command deletes all policies from the forwarding policy database provided none of them are bound to any forwarding context (GRT, VPRN, or EVPN). Otherwise, the command fails.

4.3 Tunnel Table Handling of MPLS Forwarding Policy

An endpoint forwarding policy once validated as the most preferred policy for given endpoint address is added to the TTMv4 or TTMv6 according to the address family of the address of the **endpoint** parameter. A new owner of **mpls-fwd-policy** is used. A tunnel-id is allocated to each policy and is added into the TTM entry for the policy. For more information about the **mpls-fwd-policy** command, used to enable MPLS forwarding policy in different services, refer to the following guides:

- *7450 ESS, 7750 SR, 7950 XRS, and VSR Layer 2 Services and EVPN Guide: VLL, VPLS, PBB, and EVPN*
- *7450 ESS, 7750 SR, 7950 XRS, and VSR Layer 3 Services Guide: IES and VPRN*
- *7450 ESS, 7750 SR, 7950 XRS, and VSR Router Configuration Guide*
- *7450 ESS, 7750 SR, 7950 XRS, and VSR Unicast Routing Protocols Guide*

The TTM preference value of a forwarding policy is configurable using the parameter **tunnel-table-pref**. The default value of this parameter is 255.

Each individual endpoint forwarding policy can also be assigned a preference value using the **preference** command with a default value of 255. When the forwarding policy database compares multiple forwarding policies with the same endpoint address, the policy with the lowest numerical preference value is activated and programmed into TTM. The TTM preference assigned to the policy is its own configured value in the **tunnel-table-pref** parameter.

If an active forwarding policy preference has the same value as another tunnel type for the same destination in TTM, then routes and services which are bound to both types of tunnels use the default TTM preference for the two tunnel types to select the tunnel to bind to as shown in Table 25.

Table 25 **Route Preferences**

Route Preference	Value	Release Introduced
ROUTE_PREF_RIB_API	3	new in 16.0.R4 for RIB API IPv4 and IPv6 tunnel table entry
ROUTE_PREF_MPLS_FWD_POLICY	4	new in 16.0.R4 for MPLS forwarding policy of endpoint type
ROUTE_PREF_RSVP	7	—
ROUTE_PREF_SR_TE	8	new in 14.0
ROUTE_PREF_LDP	9	—

Table 25 **Route Preferences (Continued)**

Route Preference	Value	Release Introduced
ROUTE_PREF_OSPF_TTM	10	new in 13.0.R1
ROUTE_PREF_ISIS_TTM	11	new in 13.0.R1
ROUTE_PREF_BGP_TTM	12	modified in 13.0.R1 (pref was 10 in R12)
ROUTE_PREF_UDP	254	introduced with 15.0 MPLS-over-UDP tunnels
ROUTE_PREF_GRE	255	—

An active endpoint forwarding policy populates the highest pushed label stack size among all its NHGs in the TTM. Each service and shortcut application on the router will use that value and perform a check of the resulting net label stack by counting all the additional labels required for forwarding the packet in that context.

This check is similar to the one performed for SR-TE LSP and SR policy features. If the check succeeds, the service is bound or the prefix is resolved to the forwarding policy. If the check fails, the service will not bind to this forwarding policy. Instead, it will bind to a tunnel of a different type if the user configured the use of other tunnel types. Otherwise, the service will go down. Similarly, the prefix will not get resolved to the forwarding policy and will either be resolved to another tunnel type or will become unresolved.

For more information about the **resolution-filter** CLI commands for resolving the next hop of prefixes in GRT, VPRN, and EVPN MPLS into an endpoint forwarding policy, refer to the following guides:

- *7450 ESS, 7750 SR, 7950 XRS, and VSR Layer 2 Services and EVPN Guide: VLL, VPLS, PBB, and EVPN*
- *7450 ESS, 7750 SR, 7950 XRS, and VSR Layer 3 Services Guide: IES and VPRN*
- *7450 ESS, 7750 SR, 7950 XRS, and VSR Router Configuration Guide*
- *7450 ESS, 7750 SR, 7950 XRS, and VSR Unicast Routing Protocols Guide*

BGP-LU routes can also have their next hop resolved to an endpoint forwarding policy.

4.4 Data Path Support

Note: The data path model for both the MPLS forwarding policy and the RIB API is the same. Unless explicitly stated, the selection of the active next hop within each NHG and the failover behavior within the same NHG or across NHGs is the same.

4.4.1 NHG of Resolution Type Indirect

Each NHG is modeled as a single NHLFE. The following are the specifics of the data path operation:

- Forwarding over the primary or backup next hop is modeled as a swap operation from the binding label to an implicit-null label over multiple outgoing interfaces (multiple NHLFEs) corresponding to the resolved next hops of the indirect route.
- Packets of flows are sprayed over the resolved next hops of an NHG with resolution of type indirect as a one-level ECMP spraying. See Spraying of Packets in a MPLS Forwarding Policy.
- An NHG of resolution type **indirect** uses a single NHLFE and does not support uniform failover. It will have CPM program only the active, the primary or backup, and the indirect next hop at any given point in time.
- Within a given NHG, the primary next hop is the preferred active path in the absence of any failure of the NHG of resolution type **indirect**.
- The forwarding database tracks the primary or backup next hop in the routing table. A **route delete** of the primary indirect next hop causes CPM to program the backup indirect next hop in the data path.

  A **route modify** of the indirect primary or backup next hop causes CPM to update the its resolved next hops and to update the data path if it is the active indirect next hop.
- When the primary indirect next hop is restored and is added back into the routing table, CPM waits for an amount of time equal to the user programmed revert-timer before updating the data path. However, if the backup indirect next hop fails while the timer is running, CPM updates the data path immediately.

4.4.2 NHG of Resolution Type Direct

The following rules are used for a NHG with a resolution type of **direct**:

- Each NHG is modeled as a pair of {primary, backup} NHLFEs. The following are the specifics of the label operation:
  - For a label-binding policy, forwarding over the primary or backup next hop is modeled as a swap operation from the binding label to the configured label stack or to an implicit-null label (if the **pushed-labels** command is not configured) over a single outgoing interface to the next hop.
  - For an endpoint policy, forwarding over the primary or backup next hop is modeled as a push operation from the binding label to the configured label stack or to an implicit-null label (if the **pushed-labels** command is not configured) over a single outgoing interface to the next hop.
  - The labels, configured by the **pushed-labels** command, are not validated.
- By default, packets of flows are sprayed over the set of NHGs with resolution of type **direct** as a one-level ECMP spraying. See Spraying of Packets in a MPLS Forwarding Policy.
- The user can enable weighted ECMP forwarding over the NHGs by configuring weight against all the NHGs of the policy. See Spraying of Packets in a MPLS Forwarding Policy.
- Within a given NHG, the primary next hop is the preferred active path in the absence of any failure of the NHG of resolution type direct.

Note: The RIB API feature can change the active path away from the default. The gRPC client can issue a next-hop switch instruction to activate any of the primary or backup path at any time.

- The NHG supports uniform failover. The forwarding policy database assigns a Protect-Group ID (PG-ID) to each of the primary next hop and the backup next hop and programs both of them in the data path. A failure of the active path switches traffic to the other path following the uniform failover procedures as described in Active Path Determination and Failover in a NHG of Resolution Type Direct.
- The forwarding database tracks the primary or backup next hop in the routing table. A **route delete** of the primary or backup direct next hop causes CPM to send the corresponding PG-ID switch to the data path.

  A **route modify** of the direct primary or backup next hop causes CPM to update the MPLS forwarding database and to update the data path since both next hops are programmed.

- When the primary direct next hop is restored and is added back into the routing table, CPM waits for an amount of time equal to the user programmed **revert-timer** before activating it and updating the data path. However, if the backup direct next hop fails while the timer is running, CPM activates it and updates the data path immediately. The latter failover to the restored primary next hop is performed using the uniform failover procedures as described in Active Path Determination and Failover in a NHG of Resolution Type Direct.

➡ **Note:** RIB API does not support the revert timer. The gRPC client can issue a next-hop switch instruction to activate the restored primary next hop.

- CPM keeps track and updates the IOM for each NHG with the state of active or inactive of its primary and backup next hops following a failure event, a reversion to the primary next hop, or a successful next-hop switch request instruction (RIB API only).

4.4.2.1 Active Path Determination and Failover in a NHG of Resolution Type Direct

An NHG of resolution type **direct** supports uniform failover either within an NHG or across NHGs of the same policy. These uniform failover behaviors are mutually exclusive on a per-NHG basis depending on whether it has a single primary next hop or it has both a primary and backup next hops.

When an NHG has both a primary and a backup next hop, the forwarding policy database assigns a Protect-Group ID (PG-ID) to each and programs both in data path. The primary next hop is the preferred active path in the absence of any failure of the NHG.

During a failure affecting the active next hop, or the primary or backup next hop, CPM signals the corresponding PG-ID switch to the data path which then immediately begins using the NHLFE of the other next hop for flow packets mapped to NHGs of all forwarding polices which share the failed next hop.

An interface down event sent by CPM to the data path causes the data path to switch the PG-ID of all next hops associated with this interface and perform the uniform failover procedure for NHGs of all policies which share these PG-IDs.

Any subsequent network event causing a failure of the newly active next hop while the originally active next hop is still down, blackholes traffic of this NHG until CPM updates the policy to redirect the affected flows to the remaining NHGs of the forwarding policy.

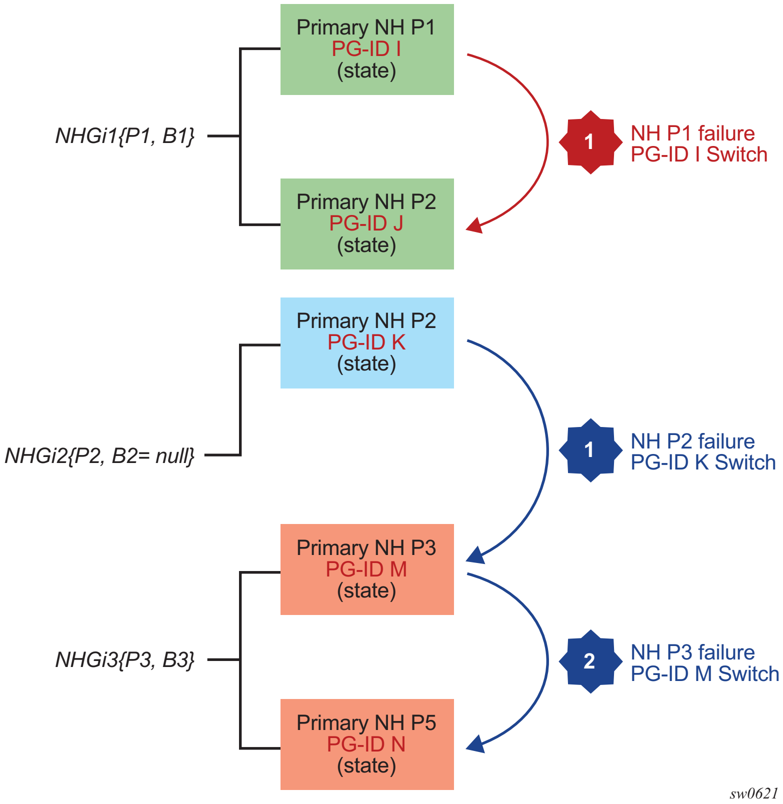
When the NHG has only a primary next hop and it fails, CPM signals the corresponding PG-ID switch to the data path which then uses the uniform failover procedure to immediately re-assign the affected flows to the other NHGs of the policy.

A subsequent failure of the active next hop of a NHG the affected flow was re-assigned to in the first failure event, causes the data path to use the uniform failover procedure to immediately switch the flow to the other next hop within the same NHG.

Figure 56 illustrates the failover behavior for the flow packets assigned to an NHG with both a primary and backup next hop and to an NHG with a single primary next hop.

The notation NHGi{Pi,Bi} refers to NHG "i" which consists of a primary next hop (Pi) and a backup next hop (Bi). When an NHG does not have a backup next hop, it is referred to as NHGi{Pi,Bi=null}.

Figure 56 **NHG Failover Based on PG-ID Switch**



4.4.3 Spraying of Packets in a MPLS Forwarding Policy

When the node operates as an LER and forwards unlabeled packets over an endpoint policy, the spraying of packets over the multiple NHGs of type **direct** or over the resolved next hops of a single NHG of type **indirect** follows prior implementation. Refer to the *7450 ESS, 7750 SR, 7950 XRS, and VSR Interface Configuration Guide.*

When the node operates as an LSR, it forwards labeled packets matching the ILM of the binding label over the label-binding policy. An MPLS packet, including a MPLS-over-GRE packet, received over any network IP interface with a binding label in the label stack, is forwarded over the primary or backup next hop of either the single NHG of type **indirect** or of a selected NHG among multiple NHGs of type **direct**.

The router performs the following procedures when spraying labeled packets over the resolved next hops of a NHG of resolution type **indirect** or over multiple NHGs of type **direct**.

1. The router performs the GRE header processing as described in *MPLS-over-GRE termination* if the packet is MPLS-over-GRE encapsulated. Refer to the *7450 ESS, 7750 SR, 7950 XRS, and VSR Router Configuration Guide*.
2. The router then pops one or more labels and if there is a match with the ILM of a binding label, the router swaps the label to implicit-null label and forwards the packet to the outgoing interface. The outgoing interface is selected from the set of primary or backup next hops of the active policy based on the LSR hash on the headers of the received MPLS packet.
   a. The hash calculation follows the method in the user configuration of the command **lsr-load-balancing** {**lbl-only** | **lbl-ip** | **ip-only**} if the packet is MPLS-only encapsulated.
   b. The hash calculation follows the method described in *LSR Hashing of MPLS-over-GRE Encapsulated Packet* if the packet is MPLS-over-GRE encapsulated. Refer to the *7450 ESS, 7750 SR, 7950 XRS, and VSR Interface Configuration Guide*.

4.4.4 Outgoing Packet Ethertype Setting and TTL Handling in Label Binding Policy

The following rules determine how the router sets the Ethertype field value of the outgoing packet:

- If the swapped label is not the Bottom-of-Stack label, the Ethertype is set to the MPLS value.

- If the swapped label is the Bottom-of-Stack label and the outgoing label is not implicit-null, the Ethertype is set to the MPLS value.
- If the swapped label is the Bottom-of-Stack label and the outgoing label is implicit-null, the Ethertype is set to the IPv4 or IPv6 value when the first nibble of the exposed IP packet is 4 or 6 respectively.

The router sets the TTL of the outgoing packet as follows:

- The TTL of a forwarded IP packet is set to MIN(MPLS_TTL-1, IP_TTL), where MPLS_TTL refers to the TTL in the outermost label in the popped stack and IP_TTL refers to the TTL in the exposed IP header.
- The TTL of a forwarded MPLS packet is set to MIN(MPLS_TTL-1, INNER_MPLS_TTL), where MPLS_TTL refers to the TTL in the outermost label in the popped stack and INNER_MPLS_TTL refers to the TTL in the exposed label.

4.4.5 Ethertype Setting and TTL Handling in Endpoint Policy

The router sets the Ethertype field value of the outgoing packet to the MPLS value.

The router checks and decrements the TTL field of the received IPv4 or IPv6 header and sets the TTL of all labels of the label stack specified in the **pushed-labels** command according to the following rules:

1. The router propagates the decremented TTL of the received IPv4 or IPv6 packet into all labels of the pushed label stack for a prefix in GRT.
2. The router then follows the configuration of the TTL propagation in the case of a IPv4 or IPv6 prefix forwarded in a VPRN context:

```
config>router>ttl-propagate>vprn-local {none | vc-only |
 all}
config>router>ttl-propagate>vprn-transit {none | vc-only
 | all}
config>service>vprn>ttl-propagate>local {inherit | none
 | vc-only | all}
config>service>vprn>ttl-propagate>transit {inherit |
 none | vc-only | all}
```

When a IPv6 packet in GRT is forwarded using an endpoint policy with an IPv4 endpoint, the IPv6 explicit null label is pushed first before the label stack specified in the **pushed-labels** command.

4.5 Weighted ECMP Enabling and Validation Rules

Weighted ECMP is supported within an endpoint or a label-binding policy when the NHGs are of resolution type **direct**. Weighted ECMP is not supported with an NHG of type **indirect**.

Weighted ECMP is performed on labeled or unlabeled packets forwarded over the set of NHGs in a forwarding policy when all NHG entries have a **load-balancing-weight** configured. If one or more NHGs have **no load-balancing-weight** configured, the spraying of packets over the set of NHGs reverts to plain ECMP.

Also, the **weighted-ecmp** command in GRT (**config**>**router**>**weighted-ecmp**) or in a VPRN instance (**config**>**service**>**vprn**>**weighted-ecmp**) are not required to enable the weighted ECMP forwarding in an MPLS forwarding policy. These commands are used when forwarding over multiple tunnels or LSPs. Weighted ECMP forwarding over the NHGs of a forwarding policy is strictly governed by the explicit configuration of a weight against each NHG.

The weighted ECMP normalized weight calculated for a NHG index causes the data path to program this index as many times as the normalized weight dictates for the purpose of spraying the packets.

4.6 Statistics

4.6.1 Ingress Statistics

The user enables ingress statistics for an MPLS forwarding policy using the CLI commands provided in Introduction to MPLS Forward Policy.

The ingress statistics feature is associated with the binding label, that is the ILM of the forwarding policy, and provides aggregate packet and octet counters for packets matching the binding label.

The per-ILM statistic index for the MPLS forwarding policy features is assigned at the time the first instance of the policy is programmed in the data path. All instances of the same policy, for example, policies with the same binding-label, regardless of the **preference** parameter value, share the same statistic index.

The statistic index remains assigned as long as the policy exists and the **ingress-statistics** context is not shutdown. If the last instance of the policy is removed from the forwarding policy database, the CPM frees the statistic index and returns it to the pool.

If ingress statistics are not configured or are shutdown in a specific instance of the forwarding policy, identified by a unique value of pair {**binding-label**, **preference**} of the forwarding policy, an assigned statistic index is not incremented if that instance of the policy is activated

If a statistic index is not available at allocation time, the allocation fails and a retry mechanism will check the statistic index availability in the background.

4.6.2 Egress Statistics

Egress statistics are supported for both binding-label and endpoint MPLS forwarding policies; however, egress statistics are only supported in case where the next-hops configured within these policies are of resolution type **direct**. The counters are attached to the NHLFE of each next hop. Counters are effectively allocated by the system at the time the instance is programmed in the data-path. Counters are maintained even if an instance is deprogrammed and values are not reset. If an instance is reprogrammed, traffic counting resumes at the point where it last stopped. Traffic counters are released and thus traffic statistics are lost when the instance is removed from the database when the egress statistic context is deleted, or when egress statistics are disabled (**egress-statistics shutdown**).

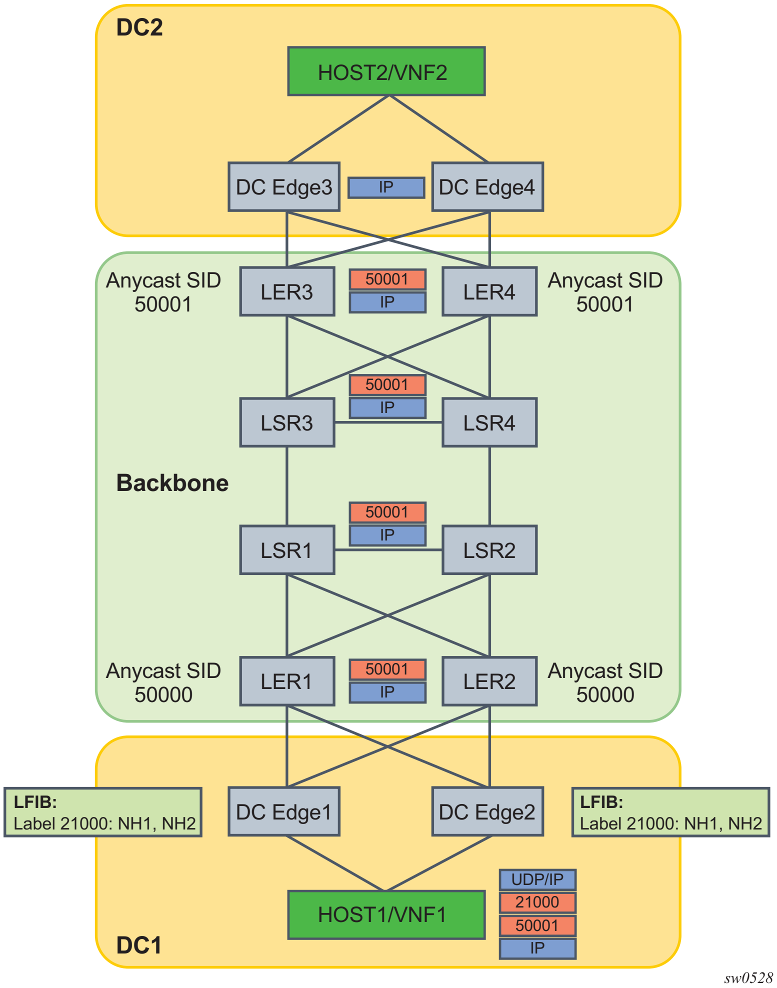
No retry mechanism is available for egress statistics. The system maintains a state per next-hop and per-instance regarding whether or not the allocation of statistic indices is successful. If the system is not able to allocate all the desired indices on a specified instance due to a lack of resources, the user should disable egress statistics on that instance, free the required number of statistics indices, and re-enable egress statistics on the desired entry. The selection of which other construct to release statistic indices from is beyond the scope of this document.

4.7 Configuring Static Label Routes using MPLS Forwarding Policy

4.7.1 Steering Flows to an Indirect Next-Hop

Figure 57 illustrates the traffic forwarding from a Virtual Network Function (VNF1) residing in a host in a Data Center (DC1) to VNF2 residing in a host in DC2 over the segment routing capable backbone network. DC1 and DC2 do not support segment routing and MPLS while the DC Edge routers do not support segment routing. Hence, MPLS packets of VNF1 flows are tunneled over a UDP/IP or GRE/IP tunnel and a static label route is configured on DC Edge1/2 to steer the decapsulated packets to the remote DC Edge3/4.

Figure 57 **Traffic Steering to an Indirect Next-hop using a Static Label Route**



The following are the data path manipulations of a packet across this network:

a. Host in DC1 pushes MPLS-over-UDP (or MPLS-over-GRE) header with outer IP destination address matching its local DC Edge1/2. It also pushes a static label 21000 which corresponds to the binding label of the MPLS forwarding policy configured in DC Edge1/2 to reach remote DC Edge3/4 (anycast address). The bottom of the label stack is the anycast SID for the remote LER3/4.

b. The label 21000 is configured on both DC Edge1 and DC Edge2 using a label-binding policy with an indirect next-hop pointing to the static route to the destination prefix of DC Edge3/4. The backup next-hop will point to the static route to reach some DC Edge5/6 in another remote DC (not shown).

c. There is EBGP peering between DC Edge1/2 and LER1/2 and between DC Edge3/4 and LER3/4.

d. DC Edge1/2 removes the UDP/IP header (or GRE/IP header) and swaps label 21000 to implicit-null and forwards (ECMP spraying) to all resolved next-hops of the static route of the primary or backup next-hop of the label-binding policy.

e. LER1/2 forwards based on the anycast SID to remote LER3/4.

f. LER3/4 removes the anycast SID label and forwards the inner IP packet to DC Edge3/4 which will then forward to Host2 in DC2.

The following CLI commands configure the static label route to achieve this use case. It creates a label-binding policy with a single NHG that is pointing to the first route as its primary indirect next-hop and the second route as its backup indirect next-hop. The primary static route corresponds to a prefix of remote DC Edge3/4 router and the backup static route to the prefix of a pair of edge routers in a different remote DC. The policy is applied to routers DC Edge1/2 in DC1.

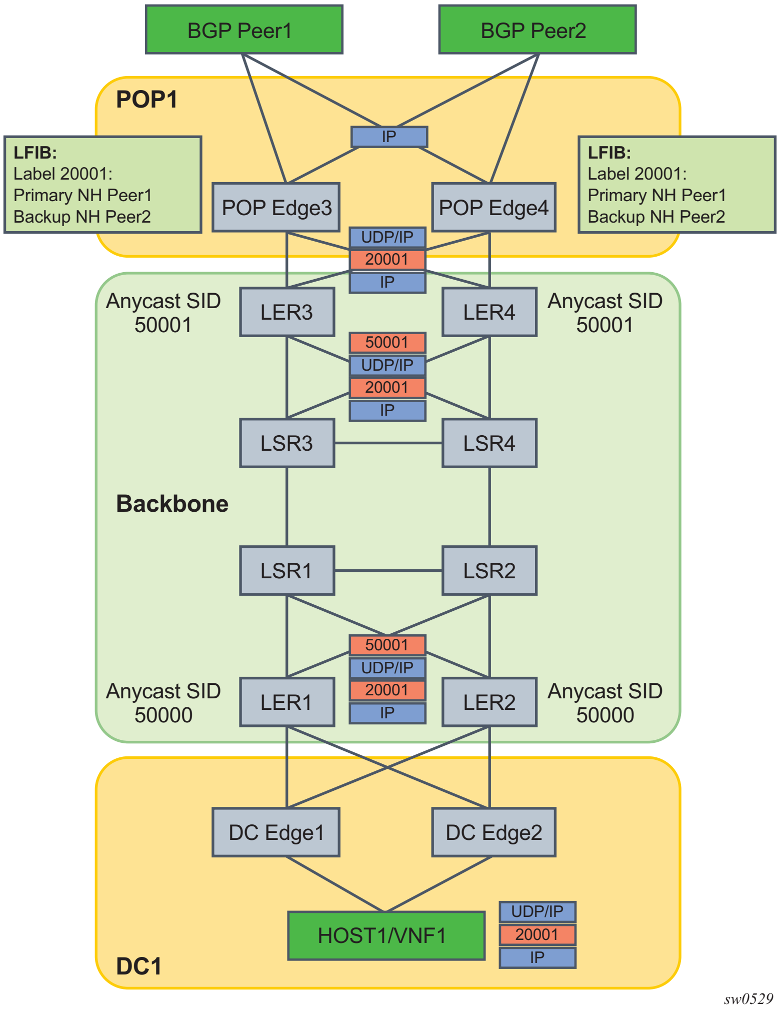
```
config>router
   static-route-entry fd84:a32e:1761:1888::1/128
      next-hop 3ffe::e0e:e05
         no shutdown
      next-hop 3ffe::f0f:f01
         no shutdown
   static-route-entry fd22:9501:806c:2387::2/128
      next-hop 3ffe::1010:1002
         no shutdown
      next-hop 3ffe::1010:1005
         no shutdown

config>router>mpls-labels
   reserved-label-block static-label-route-lbl-block
      start-label 20000 end-label 25000

config>router>mpls
   forwarding-policies
      reserved-label-block static-label-route-lbl-block
      forwarding-policy static-label-route-indirect
         binding-label 21000
         revert-timer 5
         next-hop-group 1 resolution-type indirect
            primary-next-hop
               next-hop fd84:a32e:1761:1888::1
            backup-next-hop
               next-hop fd22:9501:806c:2387::2
```

4.7.2 Steering Flows to a Direct Next-Hop

Figure 58 illustrates the traffic forwarding from a Virtual Network Function (VNF1) residing in a host in a Data Centre (DC1) to outside of the customer network via the remote peering Point Of Presence (POP1).

The traffic is forwarded over a segment routing capable backbone. DC1 and POP1 do not support segment routing and MPLS while the DC Edge routers do not support segment routing. Hence, MPLS packets of VNF1 flows are tunneled over a UDP/IP or GRE/IP tunnel and a static label route is configured on POP Edge3/4 to steer the decapsulated packets to the desired external BGP peer.

Figure 58 **Traffic Steering to a Direct Next-hop using a Static Label Route**



The intent is to override the BGP routing table at the peering routers (POP Edge3 and Edge4) and force packets of a flow originated in VNF1 to exit the network using a primary external BGP peer Peer1 and a backup external BGP peer Peer2, if Peer1 is down. This application is also referred to as Egress Peer Engineering (EPE).

The following are the data path manipulations of a packet across this network:

a. DC Edge1/2 receives a MPLS-over-UDP (or a MPLS-over-GRE) encapsulated packet from the host in the DC with the outer IP destination address set to the remote POP Edge3/4 routers in peering POP1 (anycast address). The host also pushes the static label 20001 for the remote external BGP Peer1 it wants to send to.
b. This label 20001 is configured on POP Edge3/4 using the MPLS forwarding policy feature with primary next-hop of Peer1 and backup next-hop of Peer2.
c. There is EBGP peering between DC Edge1/2 and LER1/2, and between POP Edge3/4 and LER3/4, and between POP Edge3/4 and Peer1/2.
d. LER1/LER2 pushes the anycast SID of remote LER3/4 as part of the BGP route resolution to a SR-ISIS tunnel or SR-TE policy.
e. LER3/4 removes the anycast SID and forwards the GRE packet to POP Edge3/4.
f. POP Edge3/4 removes UDP/IP (or GRE/IP) header and swaps the static label 20001 to implicit null and forwards to Peer1 (primary next-hop) or to Peer2 (backup next-hop).

The following CLI commands configure the static label route to achieve this use case. It creates a label-binding policy with a single NHG containing a primary and backup direct next-hops and is applied to peering routers POP Edge3/4.

```
config>router>mpls-labels
   reserved-label-block static-label-route-lbl-block
      start-label 20000 end-label 25000

config>router>mpls
   forwarding-policies
      forwarding-policy static-label-route-direct
         binding-label 20001
         revert-timer 10
         next-hop-group 1 resolution-type direct
            primary-next-hop
               next-hop fd84:a32e:1761:1888::1
            backup-next-hop
               next-hop fd22:9501:806c:2387::2
```

5 PCEP

5.1 Introduction to the Path Computation Element Protocol (PCEP)

The Path Computation Element Protocol (PCEP) is one of several protocols used for communication between a Wide-Area Network (WAN) Software-Define Networking (SDN) controller and network elements.

The Nokia WAN SDN Controller is known as the Network Services Platform (NSP). The NSP is a set of applications which are built on a common framework that hosts and integrates them by providing common functions. The applications are developed in a Java environment.

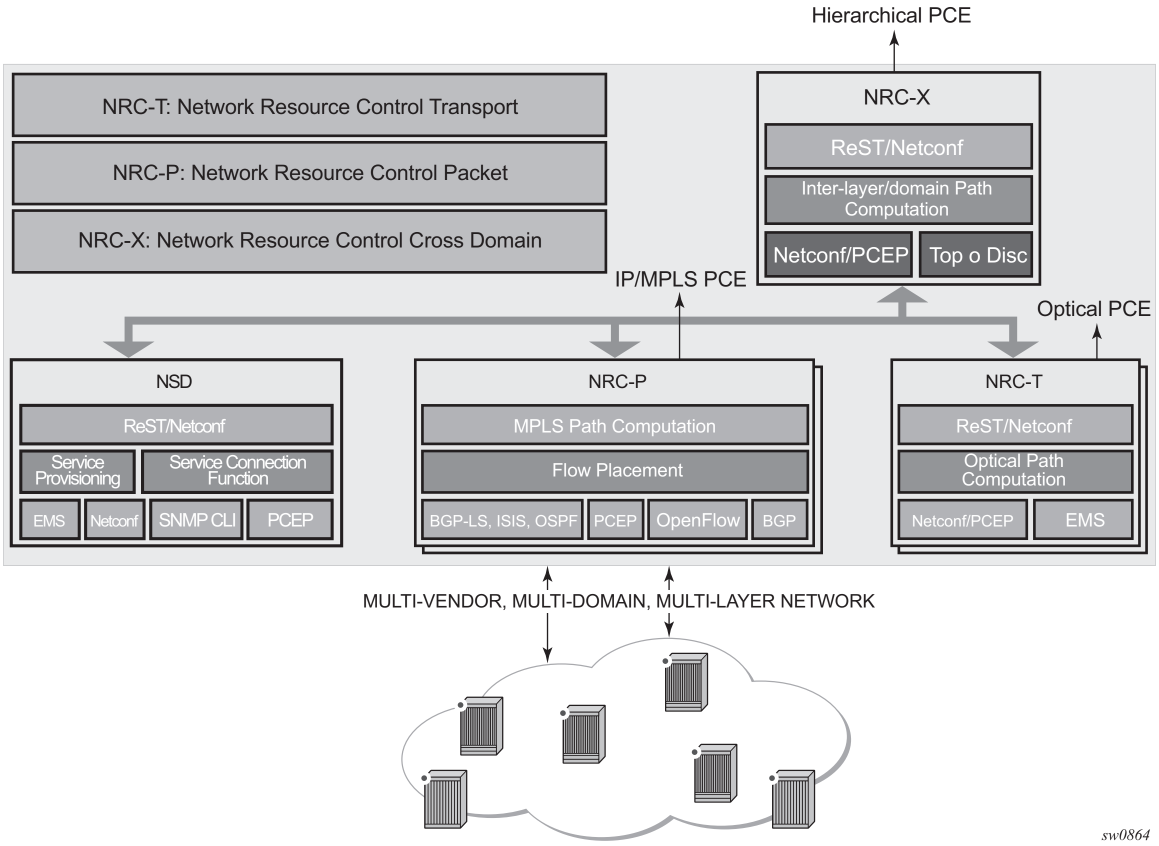
The NSP provides two major functions:

- programmable multi-vendor service provisioning
- network resource control, including resource management at Layer 0 (optical path), Layer 1 (ODU path), Layer 2 (MPLS tunnel), and at the IP flow level

The network discovery and control implements a common set of standards-based south-bound interfaces to the network elements for both topology discovery and tunnel and flow programming. It is a virtual SR OS (vSROS) image which applies the south-bound interfaces to the network elements and the adaptation layer to the applications. The south-bound interfaces include IGP and BGP-LS for topology discovery, PCEP for handling path computation requests and LSP state updates with the network elements, and forwarding plane programming protocols such as Openflow, BGP FlowSpec, and I2RS.

The above NSP functions are provided in a number of modules which can be used together or separately as illustrated in Figure 59.

Figure 59 **NSP Functional Modules**

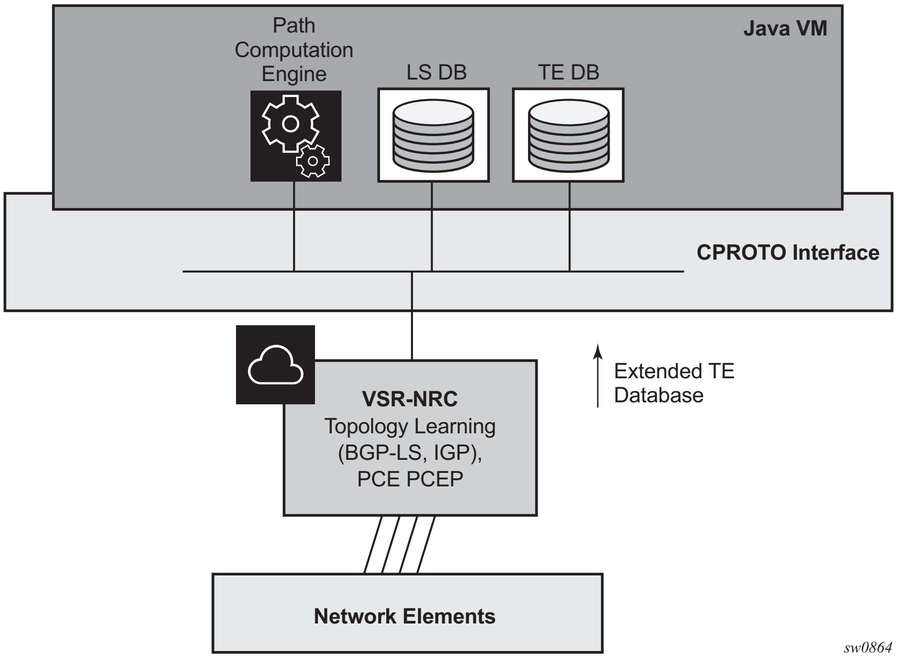


The two main features of the NSP are as follows:

- Network Services Director (NSD) — The NSD is a programmable and multi-vendor service provisioning tool exposing a single and simple API to the user and OSS. It implements service model abstraction and adapts to each vendor-specific service model. It supports provisioning services such as E-Line, E-LAN, E-Tree, L3 VPN, traffic steering, and service chaining.
- Network Resource Controller (NRC) — The NRC implements a separate module for computing and managing optimal paths for optical tunnels (NRC-T) and MPLS tunnels (NRC-P), and for computing optimal routing and placement of IP flows (NRC-P). In addition, a resource controller for inter-layer IP and optical path computation and more complex inter-domain MPLS path computation is provided as part of the NRC-X.

The NRC-P implements the stateful Path Computation Element (PCE) for packet networks. Figure 60 illustrates the NRC-P architecture and its main components.

Figure 60 **Packet Network Resource Controller (NRC-P) Architecture**



The NRC-P has the following architecture:

- a single Virtual Machine (VM) handling the Java implementation of an MPLS path computation engine, a TE graph database, and an LSP database
- a plug-in adapter with the Nokia CPROTO interface, providing reliable, TCP-based message delivery between vSROS and Java-VM. The plug-in adapter implements a compact encoding/decoding (codec) function for the message content using Google ProtoBuf. Google ProtoBuf also provides for automatic C++ (vSROS side) and Java (Java-VM side) code generation to process the exchanged message content.
- a single VM running a vSROS image handles the functions of topology discovery of multiple IGP instances and areas via IGP or BGP-LS and the PCE PCEP functions

The PCE module uses PCEP to communicate with its clients, such as the PCE Client (PCC). It also uses the PCEP to communicate with other PCEs to coordinate inter-domain path computation. Each router acting as a PCC initiates a PCEP session to the PCE in its domain.

When the user enables PCE control for one or more segment routing or RSVP LSPs, the PCE owns the path updating and periodic re-optimization of the LSP. In this case, the PCE acts in an active stateful role. The PCE can also act in a stateful passive role for other LSPs on the router by discovering them and taking into account their resource consumption when computing the path for the LSPs it has control ownership of.

The following is a high-level description of the PCE and PCC capabilities:

- base PCEP implementation, per RFC 5440
- active and passive stateful PCE LSP update, as per *draft-ietf-pce-stateful-pce*
- delegation of LSP control to PCE
- synchronization of the LSP database with network elements for PCE-controlled LSPs and network element-controlled LSPs
- support for the SR-TE P2P LSP type, as per *draft-ietf-pce-segment-routing*
- support for PCC-initiated LSPs, as per *draft-ietf-pce-stateful-pce*
- support for LSP path diversity across different LERs using extensions to the PCE path profile, as per *draft-alvarez-pce-path-profiles*
- support for LSP path bidirectionality constraints using extensions to the PCE path profile, as per *draft-alvarez-pce-path-profiles*

5.1.1 PCC and PCE Configuration

The following PCE parameters cannot be modified while the PCEP session is operational:

- **local-address**
- **keepalive**
- **dead-timer**

The **unknown-message-rate** PCE parameter can be modified while the PCEP session is operational.

The following PCC parameters cannot be modified while the PCEP session is operational:

- **local-address**
- **keepalive**
- **dead-timer**
- **peer** (regardless of **shutdown** state)

The following PCC parameters can be modified while the PCEP session is operational:

- **report-path-constraints**
- **unknown-message-rate**

5.1.2 Base Implementation of Path Computation Elements (PCE)

The base implementation of PCE uses the PCEP extensions defined in RFC 5440.

The main functions of the PCEP are:

- PCEP session establishment, maintenance, and closing
- path computation requests using the PCReq message
- path computation replies using the PCRep message
- notification messages (PCNtf) by which the PCEP speaker can inform its peer about events, such as path request cancellation by PCC or path computation cancellation by PCE
- error messages (PCErr) by which the PCEP speaker can inform its peer about errors related to processing requests, message objects, or TLVs

Table 26 lists the base PCEP messages and objects.

Table 26 **Base PCEP Message Objects and TLVs**

TLV, Object, or Message	Contained in Object	Contained in Message
OPEN Object	—	OPEN, PCErr
Request Parameter (RP) Object	—	PCReq, PCRep, PCErr, PCNtf
NO-PATH Object	—	PCRep
END-POINTS Object	—	PCReq
BANDWIDTH Object	—	PCReq, PCRep, PCRpt, PCInitiate
METRIC Object	—	PCReq, PCRep, PCRpt, PCInitiate
Explicit Route Object (ERO)	—	PCRep

Table 26 **Base PCEP Message Objects and TLVs (Continued)**

TLV, Object, or Message	Contained in Object	Contained in Message
Reported Route Object (RRO)	—	PCReq
LSPA Object	—	PCReq, PCRep, PCRpt, PCInitiate
Include Route Object (IRO)	—	PCReq, PCRep
SVEC Object	—	PCReq
NOTIFICATION Object	—	PCNtf
PCEP-ERROR Object	—	PCErr
LOAD-BALANCING Object	—	PCReq
CLOSE Object	—	CLOSE

The behavior and limitations of the implementation of the objects in Table 26 are as follows:

- PCE treats all supported objects received in a PCReq message as mandatory, regardless of whether the P-flag in the object’s common header is set (mandatory object) or not (optional object).
- The PCC implementation will always set the B-flag (B=1) in the METRIC object containing the hop metric value, which means that a bound value must be included in PCReq. PCE returns the computed value in PCRep with flags set identically to PCReq.
- The PCC implementation will always set flags B=0 and C=1 in the METRIC object for the IGP or TE metric values in the PCReq message. This means that the request is to optimize (minimize) the metric without providing a bound. PCE returns the computed value in PCRep with flags set identically to PCReq.
- The IRO and LOAD-BALANCING objects are not in the NSP PCE feature. If the PCE receives a PCReq message with one or more of these objects, it will ignore them regardless of the setting of the P-flag, and will process the path computations normally.
- LSP path setup and hold priorities will be configurable during SR-TE LSP configuration on the router, and PCC will pass the configurations on in an LSPA object. However, PCE does not implement LSP pre-emption.
- The LSPA, METRIC, and BANDWIDTH objects are also included in the PCRpt message.

The following features are not supported in the SR OS:

- PCE discovery using IS-IS, per RFC 5089, and OSPF, per RFC 5088, along with corresponding extensions for discovering stateful PCE, per *draft-sivabalan-pce-disco-stateful*
- security of the PCEP session using MD5 or TLS between PCEP peers
- PCEP synchronization optimization as per *draft-ietf-pce-stateful-sync-optimizations*
- support of end-to-end secondary backup paths for an LSP. PCE standards do not currently support an LSP container with multiple paths, and treats each request as a path with a unique PLSP ID. It is up to the router to tie the two paths together to create 1:1 protection, and to request path or SRLG diversity among them when it makes the request to PCE. This is not specific to PCE controlling an SR-TE LSP, but also to controlling an RSVP LSP.
- jitter, latency, and packet loss metrics support as per RFC 7471 and *draft-ietf-isis-te-metric-extensions*, and their use in the PCE METRIC object as per *draft-ietf-pce-pcep-service-aware*

5.1.3 PCEP Session Establishment and Maintenance

The PCEP protocol operates over TCP using destination TCP port 4189. The PCC always initiates the connection. Once the user configures the PCEP local address and the peer address on the PCC, the PCC initiates a TCP connection to the PCE. Once a connection is established, the PCC and PCE exchange OPEN messages, which initializes the PCEP session and exchanges the session parameters to be negotiated.

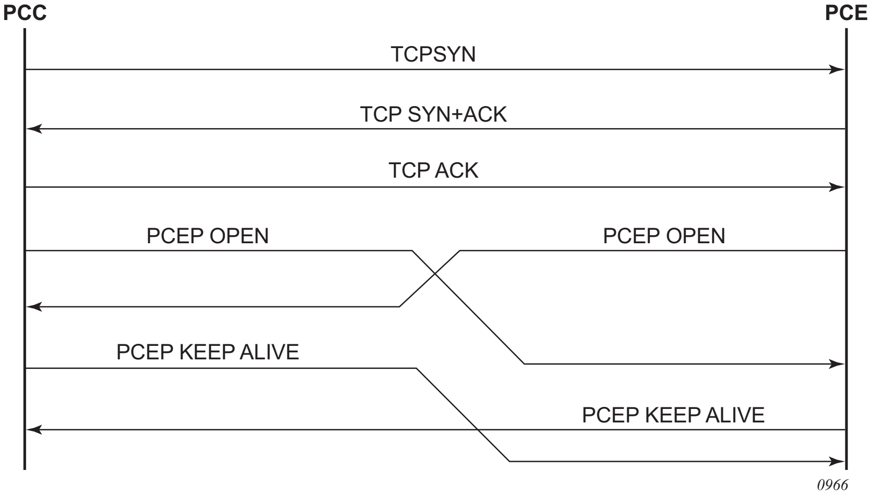
The PCC always checks first if the remote PCE address is reachable out-of-band via the management port. If not, it will try to reach the remote PCE address in-band. When the session comes up out-of-band, the management IP address is always used; the local address configured by the user is ignored and is only used for an in-band session.

A keep-alive mechanism is used as an acknowledgment of the acceptance of the session within the negotiated parameters. It is also used as a maintenance function to detect whether or not the PCEP peer is still alive.

The negotiated parameters include the Keepalive timer and the DeadTimer, and one or more PCEP capabilities such as support of Stateful PCE and the SR-TE LSP Path type.

The PCEP session initialization steps are illustrated in Figure 61.

Figure 61 **PCEP Session Initialization**



If the session to the PCE times out, the router acting as a PCC keeps the last successfully-programmed path provided by the PCE until the session to the PCE is re-established. Any subsequent change to the state of an LSP is synchronized at the time the session is re-established.

When a PCEP session to a peer times out or closes, the rate at which the PCEP speaker attempts the establishment of the session is subject to an exponential back-off mechanism.

5.1.4 PCEP Parameters

The following PCEP parameters are user-configurable on both the PCC and PCE. On the PCE, the configured parameter values are used on sessions to all PCCs.

- Keep-alive timer — A PCEP speaker (PCC or PCE) must send a keep-alive message if no other PCEP message is sent to the peer at the expiry of this timer. This timer is restarted every time a PCEP message is sent or the keep-alive message is sent.

  The keep-alive mechanism is asymmetric, meaning that each peer can use a different keep-alive timer value.

  The range of this parameter is 1 to 255 seconds, and the default value is 30 seconds. The no version returns to the default value.
- Dead timer — This timer tracks the amount of time a PCEP speaker (PCC or PCE) waits after the receipt of the last PCEP message before declaring its peer down.

The dead timer mechanism is asymmetric, meaning that each PCEP speaker can propose a different dead timer value to its peer to use to detect session timeouts.

The range of this parameter is 1 to 255 seconds, and the default value is 120 seconds. The no version returns to the default value.

- Maximum rate of unknown messages — When the rate of received unrecognized or unknown messages reaches this limit, the PCEP speaker closes the session to the peer.
- Session re-delegation and state timeout — If the PCEP session to the PCE goes down, all delegated PCC-initiated LSPs have their state maintained in the PCC and are not timed out. The PCC will continue to try re-establishing the PCEP session. When the PCEP session is re-established, the LSP database is synchronized with the PCE, and any LSP which went down since the last time the PCEP session was up will have its path updated by the PCE.

5.1.4.1 Stateful PCE

The main function introduced by stateful PCE over the base PCE implementation is the ability to synchronize the LSP state between the PCC and the PCE. This allows the PCE to have all the required LSP information to perform re-optimization and updating of the LSP paths.

Table 27 describes the messages and objects supported by stateful PCE in the SR OS.

Table 27 **PCEP Stateful PCE Extension Objects and TLVs**

TLV, Object, or Message	Contained in Object	Contained in Message
Path Computation State Report (PCRpt)	—	New message
Path Computation Update Request (PCUpd)	—	New message
Stateful PCE Capability TLV	OPEN	OPEN
Stateful Request Parameter (SRP) Object	—	PCRpt, PCErr, PCInitiate
LSP Object	ERO	PCRpt, PCReq, PCRep, PCInitiate
LSP Identifiers TLV	LSP	PCRpt
Symbolic Path Name TLV	LSP, SRP	PCRpt, PCInitiate

Table 27 **PCEP Stateful PCE Extension Objects and TLVs (Continued)**

TLV, Object, or Message	Contained in Object	Contained in Message
LSP Error Code TLV	LSP	PCRpt
RSVP Error Spec TLV	LSP	PCRpt

The behavior and limitations of the implementation of the objects in Table 27 are as follows:

- PCC and PCE support all PCEP capability TLVs defined in this document and will always advertise them. If the OPEN object received from a PCEP speaker does not contain one or more of the capabilities, the PCE or PCC will not use them during that specific PCEP session.
- The PCC always includes the LSP object in the PCReq message to make sure that the PCE can correlate the PLSP-ID for this LSP when a subsequent PCRpt message arrives with delegation bit set. The PCE will, however, still honor a PCReq message without the LSP Object.
- PCE path computation will only consider the bandwidth used by LSPs in its LSP-DB. As a result, there are two situations where PCE path computation will not accurately take into account the bandwidth used in the network:
  - When there are LSPs which are signaled by the routers but are not synchronized up with the PCE. The user can enable the reporting of the LSP to the PCE LSP database for each LSP.
  - When the stateful PCE is peering with a third party stateless PCC, implementing only the original RFC 5440. While PCE will be able to bring the PCEP session up, the LSP database will not be updated, since stateless PCC does not support the PCRpt message. As such, PCE path computation will not accurately take into account the bandwidth used by these LSPs in the network.
- PCE ignores the R-flag (re-optimize flag) in the PCReq message when acting in stateful-passive mode for a given LSP, and will always return the new computed path, regardless if it is link-by-link identical or has the same metric as the current path. The decision whether or not to initiate the new path in the network belongs to the PCC.
- The SVEC object is not supported in the SR OS and the NSP. If the PCE receives a PCReq message with the SVEC object, it will ignore the SVEC object and treat each path computation request in the PCReq message as independent, regardless of the setting of the P-flag in the SVEC object common header.

- When an LSP is delegated to the PCE, there can be no prior state in the NRC-P LSP database for the LSP. This could be due to the PCE not having received a PCReq message for the same PLSP-ID. In order for the PCE to become aware of the original constraints of the LSP, the following additional procedures are performed.
  - PCC appends a duplicate of each of the LSPA, METRIC, and BANDWIDTH objects in the PCRpt message. The only difference between the two objects of the same type is that the P-flag is set in the common header of the duplicate object to indicate a mandatory object for processing by the PCE.
  - The value of the metric or bandwidth in the duplicate object contains the original constraint value, while the first object contains the operational value. This is applicable to hop metrics in the METRIC object and BANDWIDTH object only. SR OS PCC does not support putting a bound on the IGP or TE metric in the path computation.
  - The path computation on the PCE uses the first set of objects when updating a path if the PCRpt contains a single set. If the PCRpt contains a duplicate set, PCE path computation must use the constraints in the duplicate set.
  - For interoperability, implementations compliant to PCEP standards should be able to accept the first metric object and ignore the second object without additional error handling. Since there are also BANDWIDTH and LSPA objects, the [**no**] **report-path-constraints** command is provided in the PCC on a per-PCEP session basis to disable the inclusion of the duplicate objects. Duplicate objects are included by default.

Stateful PCE uses the additional messages, TLVs, and objects described in Table 28 for PCE initiation of LSPs.

Table 28 **PCEP Stateful PCE Extension Objects and TLVs Locations**

TLV, Object, or Message	Contained in Object	Contained in Message
PCE LSP Initiate Message (PCInitiate)	—	New message
PCC LSP Create Flag (C-Flag)	LSP	PCRpt
PATH_PROFILE_ID TLV	Path Profile	N/A

5.1.4.2 PCEP Extensions in Support of SR-TE LSPs

In order for the PCE and PCC to manage the path of an SR-TE LSP, they both implement the following extensions to PCEP in support of segment routing.

- A new Segment Routing capability TLV in the OPEN object to indicate support of segment routing tunnels by the PCE and the PCC during PCEP session initialization. This TLV is referred to as the SR-PCE-CAPABILITY TLV.
- The PCC and PCE support all PCEP capability TLVs defined in this chapter, and will always advertise them. If the OPEN object received from a PCEP speaker does not contain one or more of the capabilities, the PCE or the PCC will not use them during that specific PCEP session.
- A new Path Setup Type TLV for SR-TE LSPs to be included in the Stateful PCE Request Parameters (SRP) Object during path report (PCRpt) messages by the PCC.

  A Path Setup Type TLV with a value of 1 identifies an SR-TE LSP.
- A new Segment Routing ERO and RRO with sub-objects, referred to as SR-ERO and SR-RRO sub-objects, which encode the SID information in PCRpt messages.
- The PCE implementation supports the Segment-ID (SID) Depth value in the METRIC object. This is always signaled by the PCC in the PCEP Open object as part of the as SR-PCE-CAPABILITY TLV. It is referred to as the Maximum Stack Depth (MSD). In addition, the per-LSP value for the **max-sr-labels** option, if configured, is signaled by the PCC to the PCE in the Segment-ID (SID) Depth value in a METRIC object for both a PCE-computed LSP and a PCE-controlled LSP. PCE will compute and provide the full explicit path with TE-links specified. If there is no path with the number of hops lower than the MSD value, or the Segment-ID (SID) Depth value if signaled, a reply with no path will be returned to the PCC.
- For a PCC controlled LSP, if the label stack returned by the TE-DB’s hop-to-label translation exceeds the per LSP maximum SR label stack size, the LSP is brought down.
- If the Path Setup Type (PST) TLV is not included in the PCReq message, the PCE or PCC must assume it is for an RSVP-TE LSP.

Table 29 describes the segment routing extension objects and TLVs supported in the SR OS.

Table 29 **PCEP Segment Routing Extension Objects and TLVs**

TLV, Object, or Message	Contained in Object	Contained in Message
SR PCE CAPABILITY TLV	OPEN	OPEN
Path Setup Type (PST) TLV	SRP	PCReq, PCRep, PCRpt
SR-ERO Sub-object	ERO	PCRep, PCRpt
SR-RRO Sub-object	RRO	PCReq, PCRpt

Table 29 **PCEP Segment Routing Extension Objects and TLVs**

TLV, Object, or Message	Contained in Object	Contained in Message
Segment-ID (SID) Depth Value in METRIC Object	METRIC	PCReq, PCRpt

5.1.5 LSP Initiation

An LSP that is configured on the router is referred to as a PCC-initiated LSP. An LSP that is not configured on the router, but is instead created by the PCE at the request of an application or a service instantiation, is referred to as a PCE-initiated LSP.

The SR OS support three different modes of operations for PCC-initiated LSPs which are configurable on a per-LSP basis.

- When the path of the LSP is computed and updated by the router acting as a PCE Client (PCC), the LSP is referred to as PCC-initiated and PCC-controlled.

  A PCC-initiated and PCC-controlled LSP has the following characteristics:
  - The LSP can contain strict or loose hops, or a combination of both.
  - CSPF is supported for RSVP-TE LSPs. Local path computation takes the form of hop-to-label translation for SR-TE LSPs.
  - LSPs can be reported to synchronize the LSP database of a stateful PCE server using the **pce-report** option. In this case, the PCE acts in passive stateful mode for this LSP. The LSP path can not be updated by the PCE. In other words, the control of the LSP is maintained by the PCC.
- When the path of the LSP is computed by the PCE at the request of the PCC, it is referred to as PCC-initiated and PCE-computed.

  A PCC-initiated and PCE-computed LSP has the following characteristics:
  - The user must enable the **path-computation-method pce** option for the LSP so that the PCE can perform path computation at the request of the PCC only. PCC retains control.
  - LSPs can be reported to synchronize the LSP database of a stateful PCE server using the **pce-report** option. In this case, the PCE acts in passive stateful mode for this LSP.
- When the path of the LSP is updated by the PCE following a delegation from the PCC, it is referred to as PCC-initiated and PCE-controlled.

  A PCC-initiated and PCE-controlled LSP has the following characteristics:

- The user must enable the **pce-control** option for the LSP so that the PCE can perform path updates following a network event without an explicit request from the PCC. PCC delegates full control.
- The user must enable the **pce-report** option for LSPs that cannot be delegated to the PCE. The PCE acts in active stateful mode for this LSP.

The SR OS also supports PCE-initiated LSPs. PCE-initiated LSP is a feature that allows a WAN SDN Controller, for example, the Nokia NSP, to automatically instantiate an LSP based on a service or application request. Only SR-TE PCE-initiated LSPs are supported.

The instantiated LSP does not have a configuration on the network routers and is therefore treated the same way as an auto-LSP. The parameters of the LSP are provided using policy lookup in the NSP and are passed to the PCC using PCEP as per RFC 8281. Missing LSP parameters are added using a default or specified LSP template on the PCC.

PCE-initiated LSPs have the following characteristics.

- The user must enable **pce-initiated-lsp sr-te** to enable the PCC to accept and process PCInitiate messages from the PCE.
- The user must configure one or more LSP templates of type **pce-init-p2p-srte** for SR-TE LSPs. A default template is supported that is used for LSPs for which no ID or an ID of 0 is included in the PCInitiate message. The user must configure at least one default PCE-initiated LSP template.

PCE-initiated LSPs are a form of SR-TE Auto-LSP and are available to the same forwarding contexts. See Forwarding Contexts Supported with SR-TE Auto-LSP. Similar to other auto-LSPs, they are installed in the TTM and are therefore available to advanced policy-based services using auto-bind such as VPRN and E-VPN. However, they cannot be used with provisioned SDPs.

5.1.5.1 PCC-Initiated and PCE-Computed/Controlled LSPs

The following is the procedure for configuring and programming a PCC-initiated SR-TE LSP when control is delegated to the PCE.

Step 1. The LSP configuration is created on the PE router via CLI or via the OSS/NSP NFM-P.

The configuration dictates which PCE control mode is desired: active (**pce-control** and **pce-report** options enabled) or passive (**path-computation-method pce** enabled and **pce-control** disabled).

Step 2. PCC assigns a unique PLSP-ID to the LSP. The PLSP-ID uniquely identifies the LSP on a PCEP session and must remain constant during its lifetime. PCC on the router must keep track of the association of the PLSP-ID to the Tunnel-ID and Path-ID, and use the latter to communicate with MPLS about a specific path of the LSP. PCC also uses the SRP-ID to correlate PCRpt messages for each new path of the LSP.

Step 3. The PE router does not validate the entered path. Note however that in the SR OS, the PCE supports the computation of a path for an LSP with empty-hops in its path definition. While PCC will include the IRO objects in the PCReq message to PCE, the PCE will ignore them and compute the path with the other constraints except the IRO.

Step 4. The PE router sends a PCReq message to the PCE to request a path for the LSP, and includes the LSP parameters in the METRIC object, the LSPA object, and the BANDWIDTH object. The PE router also includes the LSP object with the assigned PLSP-ID. At this point, the PCC does not delegate the control of the LSP to the PCE.

Step 5. The PCE computes a new path, reserves the bandwidth, and returns the path in a PCRep message with the computed ERO in the ERO object. It also includes the LSP object with the unique PLSP-ID, the METRIC object with any computed metric value, and the BANDWIDTH object.

Note: For the PCE to be able to use the SRLG path diversity and admin-group constraints in the path computation, the user must configure the SRLG and admin-group membership against the MPLS interface and make sure that the **traffic-engineering** option is enabled in IGP. This causes IGP to flood the link SRLG and admin-group membership in its participating area, and for PCE to learn it in its TE database.

Step 6. The PE router updates the CPM and the data path with the new path.

Up to this point, the PCC and PCE are using passive stateful PCE procedures. The next steps will synchronize the LSP database of the PCC and PCE for both PCE-computed and PCE-controlled LSPs. They will also initiate the active PCE stateful procedures for the PCE-controlled LSP only.

Step 7. The PE router sends a PCRpt message to update the PCE with an UP state, and also sends the RRO as confirmation. It now includes the LSP object with the unique PLSP-ID. For a PCE-controlled LSP, the PE router also sets the delegation control flag to delegate control to the PCE. The state of the LSP is now synchronized between the router and the PCE.

Step 8. Following a network event or a re-optimization, the PCE computes a new path for a PCE-controlled LSP and returns it in a PCUpd message with the new ERO. It will include the LSP object with the same unique PLSP-ID assigned by the PCC, as well as the Stateful Request Parameter (SRP) object with a unique SRP-ID-number to track error and state messages specific to this new path.

Step 9. The PE router updates the CPM and the data path with the new path.

Step 10. The PE router sends a PCRpt message to inform the PCE that the older path is deleted. It includes the unique PLSP-ID value in the LSP object and the R (Remove) bit set.

Step 11. The PE router sends a new PCRpt message to update PCE with an UP state, and also sends the RRO to confirm the new path. The state of the LSP is now synchronized between the router and the PCE.

Step 12. If PCE owns the delegation of the LSP and is making a path update, MPLS will initiate the LSP and update the operational value of the changed parameters while the configured administrative values will not change. Both the administrative and operational values are shown in the details of the LSP path in MPLS.

Step 13. If the user makes any configuration change to the PCE-computed or PCE-controlled LSP, MPLS requests that the PCC first revoke delegation in a PCRpt message (PCE-controlled only), and then MPLS and PCC follow the above steps to convey the changed constraint to PCE which will result in the programming of a new path into the data path, the synchronization of the PCC and PCE LSP databases, and the return of delegation to PCE.

The above procedure is followed when the user performs a **no shutdown** command on a PCE-controlled or PCE-computed LSP. The starting point is an LSP which is administratively down with no active path. For an LSP with an active path, the following items can apply:

a. If the user enabled the **path-computation-method pce** option on a PCC-controlled LSP with an active path, no action is performed until the next time the router needs a path for the LSP following a network event of a LSP parameter change. At that point, the prior procedure is followed.

b. If the user enabled the **pce-control** option on a PCC-controlled or PCE-computed LSP with an active path, the PCC will issue a PCRpt message to the PCE with an UP state, as well as the RRO of the active path. It will set the delegation control flag to delegate control to the PCE. The PCE will keep the active path of the LSP and make no updates to it until the next network event or re-optimization. At that point, the prior procedure is followed.

5.1.5.2 PCE-Initiated LSPs

The following is the procedure for configuring and programming a PCE-initiated SR-TE LSP.

Step 1. The user must enable **pce-initiated-lsp sr-te** using the CLI or using the OSS. The user can also optionally configure a limit to the number of PCE-Initiated LSPs that the PCE can instantiate on a node using the **max-srte-pce-init-lsps** command in the CLI or using the OSS.

Step 2. The user must configure at least one LSP template of type **pce-init-p2p-srte** to select the value of the LSP parameters that remain under the control of the PCC. At a minimum, a default template should be configured (type **pce-init-p2p-srte default**). In addition, LSP templates with a defined template ID can be configured. The template ID can be included in the path profile of the PCEInitiate message to indicate which non-default template to use for a particular LSP. If the PCInitiate message does not include the PCE path profile, MPLS uses the default PCE-initiated LSP template. Table 30 lists the applicable LSP template parameters. These are grouped into:

- parameters that are controlled by the PCE and that the PCC cannot change (invalid, implicit, and signaled in PCEP)
- parameters that are controlled by the PCC and are used for signaling the LSP in the control plane
- parameters that are controlled by the PCC and are related to the usability of the LSP by MPLS and other applications such as routing protocols, services, and OAM

The user can configure these parameters in the template.

Table 30 **LSP Template Parameters**

Controlled by PCE			Controlled by PCC	
Invalid	**Implicit**	**Signaled in PCEP**	**LSP Signaling Options**	**LSP Usability Options**
auto-bandwidth	pce-report	bandwidth	—	—
retry-limit	—	exclude	—	bgp-shortcut
retry-timer	pce-control	from	—	bgp-transport-tunnel
shutdown	pce-report	hop-limit	default-path (mandatory, must be empty)	—
least-fill	path-computation-method pce	include	—	—
metric-type		—	—	entropy-label
—	—	setup-priority	—	igp-shortcut

Table 30 **LSP Template Parameters (Continued)**

Controlled by PCE			Controlled by PCC	
Invalid	**Implicit**	**Signaled in PCEP**	**LSP Signaling Options**	**LSP Usability Options**
—	—	hold-priority	—	—
—	—	—	—	load-balancing-weight
—	—	—	—	max-sr-labels
—	—	—	—	additional-frr-labels
—	—	—	—	metric
—	—	—	—	vprn-auto-bind
—	—	—	—	admin-tag

All PCE-initiated LSPs using a particular LSP template are deleted if the user deletes the template. The default template can be created or deleted if the **pce-initiated-lsp>sr-te** context does not exist. However, the **pce-init-p2p-sr-te default lsp-template** cannot be deleted if the **pce-initiated-lsp>sr-te** context exists and is not shutdown. This context must be shutdown to delete the **pce-init-p2p-sr-te default** LSP template, which brings down all PCE Initiated LSPs. The **pce-initiated-lsp>sr-te** context cannot be administratively enabled if the **pce-init-p2p-sr-te default lsp-template** is not configured.

A shutdown of an LSP template does not bring down any already established LSPs. Parameters can only be changed once in the shutdown state and the changes do not take effect until a **no shutdown** is performed. This means that PCE updates use older parameters if the template is still shut down.

MPLS copies the lsp-template parameters into the lsp-entry when a PCE initiated LSP is created. MPLS handles lsp-updates based on the last copied parameters.

After the lsp-template parameter changes, when the lsp-template is **no shutdown**.

- MPLS copies the related TTM parameters (listed below) into the LSP entry, and updates TTM
- If there is a change in **max-sr-labels**, MPLS re-evaluates the related LSPs, and brings paths down if applicable (for example, if current hopCount is greater than the applicable **max-sr-labels** value).

The TTM LSP-related parameters include:

- Metric
- VprnAutoBind
- LoadBalWeight
- MaxSrLabels
- AdditionalFrrLabels
- MetricOffset
- IgpShortCut
- IgpShortcutLfaOnly
- IgpShortcutLfaProtect
- LspBgpShortCut
- LspBgpTransTunnel

A PCE-initiated LSP "update" request will be accepted regardless of the LSP template administrative state, as follows:

- If the LSP template is adminUp, the system copies the LSP template parameters to the LSP/path.
- If the LSP template is adminDown, the system uses the previously copied LSP template parameters and responds to the update with an LSP operUp report.

Step 3. The user can set the redelegation and state timers on the PCC. Redelegation timeout and state timeout timers are started when the PCEP session goes down or the PCE signals overload. The redelegation timer applies to both PCC-initiated and PCE-initiated LSPs, while the state timer applies only to PCE-initiated LSPs. The redelegation and state timers are configured in the CLI or through management, as follows:

config>**router**>**pcep**>**pcc**>

[**no**] **redelegation-timer** *seconds*

[**no**] **state-timer** *seconds* [**action** {**remove** | **none**}]

If the delegated PCE-initiated LSPs cannot be redelegated by the time these timers expire, a configurable action is performed by the PCC. The supported actions are **remove** or **none**, with a default of **remove**.

Step 4. The PCE can then initiate and remove LSPs on the PCC. These procedures are described in LSP Instantiation Using PCEP, LSP Deletion Using PCEP, and Dynamic State Handling for PCE Initiated LSPs.

5.1.5.2.1 LSP Instantiation Using PCEP

The following procedures are followed in the instantiation of a PCE-initiated LSP by both the NSP and SR OS router. Further protocol details can be found in RFC 8281.

NSP Generation of PCInitiate

1. When the PCEP session is established from the PCC to PCE, the PCC and PCE exchange the Open object and both set the new “I flag, LSP-INSTANTIATION CAPABILITY” flag, in the STATEFUL-PCE-CAPABILITY TLV flag field.
2. The operator, using the north-bound REST interface, the NSD or another interface, makes a request to the NSP to initiate an LSP, specifying the following parameters:
   a. source address
   b. destination address
   c. LSP type (SR-TE)
   d. bandwidth value
   e. include/exclude admin-group constraints
   f. optional PCE path profile ID for the path computation at the PCE
   g. optional PCE-initiated LSP template ID for use by the PCC to complete the instantiation of the LSP
3. The NSP crafts the PCInitiate message and sends it to the PCC using PCEP. The message contains the LSP object with PLSP-ID=0, the SRP object, the ENDPOINTS object, the computed SR-ERO (SR-TE) object, and the list of LSP attributes (bandwidth object, one or more metric objects, and the LSPA object). The LSP path name is inserted into the Symbolic Path Name TLV in the LSP object.
4. The PCE-initiated LSP template ID to be used at the PCC, if any, is included in the PATH-PROFILE-ID TLV of the Path Profile object. The Profile ID matches the PCE-initiated LSP template ID at the PCC and is not the same as
5. The Path Profile ID is used on the PCE to compute the path of this PCE-initiated LSP.

SR OS Router Procedures on Receiving a PCInitiate Message

1. If a PCInitiate message includes a name that is a duplicate of an existing LSP on the router, the system generates an error.
2. The router assigns a PLSP-ID and looks up the specified PCE-initiated LSP template ID, if any, or the default PCE-initiated LSP template, to retrieve the local parameters, and instantiates the SR-TE LSP.
3. The instantiated LSP is added to the TTM and is used by all applications that look up a tunnel in the TTM.
4. The router crafts a PCRpt message with the Tunnel-ID, LSP-ID, and the RRO and passes it along with the PLSP-ID set to the assigned value and the delegation bit set in the LSP object to the PCE.

NSP Procedures on Receiving a PCRpt Message for a PCE

1. The NSP confirms the bandwidth reservation and updates its LSP database. The PCC and PCE are synchronized at this point.
2. The NSP reports the PLSP-ID/Tunnel-ID to the application, for example NSD, or to the operator that uses it in the specific application that originated the request.
3. The PCE can perform updates to the path during the lifetime of the LSP by using the PCUpd message in the same way as with a delegated PCC-initiated LSP.

5.1.5.2.2 LSP Deletion Using PCEP

The following procedures apply in the deletion of a PCE-initiated LSP. More protocol level details are provided in RFC 8281. These procedures are applicable when the user manually deletes the PCE-initiated LSP or the NSP application, or when NSD requests the deletion of the PCE-initiated LSP. The procedures that apply when a network event occurs are described in SR OS Router Procedures.

- The NSP crafts a PCInitiate message for the corresponding PLSP-ID and sets the R-bit in the SRP object flags to indicate to the PCC that it must delete the LSP. The NSP sends the message to the PCC using PCEP.

SR OS Router Procedures on Receipt of a PCInitiate with the R-bit Set

1. The router deletes the state of the LSP.
2. The router crafts a PCRpt message with the R-bit set in the LSP object flags.

NSP Procedures Upon Issuance of pce-init delete Command

- The NSP deletes the LSP from its LSP database.

5.1.5.2.3 Dynamic State Handling for PCE Initiated LSPs

NSP Procedures

1. The NRC-P controls the creation and the deletion of the PCE-initiated LSP.
2. All LSP creation retries are performed by the NSP. If the PCC rejects an instantiation, the NSP can issue a new request for instantiation or give up and delete the LSP state locally after a configurable maximum number of retries.

3. The NSP can reject an instantiation request if it does not receive a PCRpt from the PCC message within a configured timeframe.
4. When the PCEP session comes up and the LSP DB synchronization from the PCC to PCE is complete, the NSP reinitiates the PCE-initiated LSPs that are missing from the PCC reports.
5. If a PCEP session goes down, the NSP stops sending any new or updated PCE-initiated LSP paths to that PCC; therefore, the LSP DB on the NSP and PCC can go out of synchronization during that time.
6. If the PCEP session to a PCC goes down, the NSP marks all PCE-initiated and PCC-initiated LSPs for that PCC as stale but keeps their reservation for an amount of time equal to the **state-timeout** timer. The **state-timeout** timer applies to both PCE-initiated and PCC-initiated LSPs on the PCE and is set to a fixed value of 10 minutes.

Note: The **state-timeout** timer must be considerably larger than the maximum state timer on the PCC to give the PCC time to clean up PCE-initiated LSPs and prevent PCInit requests for duplicate LSPs.

   a. If the PCEP session was re-established within that time, the NRC-P reinitiates all PCE-initiated LSPs toward the PCC from which a PCRpt remove with the special error code LSP_ERR_SYNC_DELETE was received during the LSP DB synchronization with the PCC.
   b. If the **state-timeout** timer expires, the NRC-P releases the resources but does not delete the LSPs from the LSP DB. If the PCEP session comes up subsequently, the NSP recomputes the path of each LSP from which a PCRpt remove with the special error code LSP_ERR_SYNC_DELETE was received during the LSP DB synchronization with the PCC and sends the PCC a PCInitiate message for each LSP.

7. If the NSP is informed by the VSR-NRC of a PCRpt with the remove flag in the LSP object and SRP object set for each of them, it follows the same procedures for these LSPs as when the PCEP session goes down.

SR OS Router Procedures

Table 31 summarizes the impact of various PCC operational events on the status of PCE-initiated LSPs.

Table 31 **Impact of PCC Operational Events**

Event	Impact on PCE-initiated LSPs	
	Oper-down	Deleted
MPLS shutdown	✓ [1]	
no mpls		✓[2]
no pce-initiated-lsp		✓ (all) [2]
no sr-te		✓ (sr-te) [2]
sr-te shutdown	✓ (sr-te) [1]	
pcc shutdown		✓ (all) [3]
pcc peer shutdown		✓[3]
Delete LSP template ID		✓ (LSPs using template) [2]
Delete default LSP template		✓ (all) [2]

Notes:

1. Also results in a PCRpt to the PCE with LSP error admin down.
2. Also results in a PCRpt to the PCE with LSP deleted.
3. A PCRpt with delete and a special error code, for example, LSP_ERR_SYNC_DELETE, is sent during the PCC rejoin synchronization that occurs when the PCC or PCC peer comes back up.

The following list describes in more detail the actions that the PCC takes on PCE-initiated LSPs as a result of PCC operational events:

1. If any event causes PCE-initiated LSPs to be deleted by the PCC, the PCC sends a PCRpt with remove the flag in both the SRP object and the LSP object set for each impacted LSP. If the event is a failure of the PCEP session to the PCE, or a shutdown of the PCC or PCC peer, the PCRpt is sent, with the special error code LSP_ERR_SYNC_DELETE, only after the PCEP session comes back up during the PCC resynchronization with the PCE.
2. If any event causes PCE-initiated LSPs to go operationally down, the PCC router sends a PCRpt with the operational bits in the LSP object set to DOWN for each impacted LSP.
3. If the user shuts down the PCC process on the router, all PCE-initiated LSPs are deleted. When the user performs a **no shutdown** of the PCC process, the PCC reports to the PCE so that the NSP is aware.

4. If a PCEP peer is shut down, the PCEP session goes down but the PCC keeps the state of all PCE-initiated LSPs, subject to the following rules regarding redelegation and the cleanup of state. See section 5.7.5 of RFC 8231 and section 6 of RFC 8281. These rules apply to all LSPs delegated to the PCE.

   Redelegation timeout and state timeout timers are started when the PCEP session goes down or the PCE signals overload. Configuration of these timers is described in PCE-Initiated LSPs. The system enforces that the **state-timer** be greater than the **redelegation-timer**, as specified in RFC 8231.

   The objectives of redelegation are described in Section 5.7.5 of RFC 8231. The redelegation process is as follows for both PCE-initiated and PCC-initiated LSPs.

   The existing LSP delegation state is maintained while the LSP redelegation timer is running. This gives the PCE time to recover. At the expiry of the redelegation timer, the PCC attempts to redelegate the LSPs to the PCE, as follows:

   - if the PCEP session to the existing PCE is still down or the PCE is still in overload, return delegation state to the PCC for all the delegated LSPs
   - wait until the PCEP session comes up and then attempt to redelegate the remaining LSPs back to the PCE. For each LSP, set a redelegation attempted flag once redelegation is attempted. If redelegation is accepted for all PCE-initiated LSPs delegated to the PCC before the state timeout timer expires, the system is behaving as expected.
   - if the state timeout timer expires, wait until all LSPs have been processed. The LSPs that are not redelegated but have the redelegation attempted flag set have the configured action applied to them. If this is **delete**, LSPs are deleted; otherwise, wait until the PCEP session comes up and then attempt to redelegate the remaining LSPs back to the PCE.

5.1.5.2.4 PCEP Support for RSVP-TE LSP

This section describes PCEP support of PCE Client-initiated (PCC-initiated) RSVP-TE LSP. The PCEP support of an RSVP-TE LSP provides the following modes of operation:

- PCC-initiated and PCC-controlled
- PCC-initiated and PCE-computed
- PCC-initiated and PCE-controlled

Each primary and secondary path is assigned its own unique Path LSP-ID (PLSP-ID). PCC indicates to PCE the state of each path (both UP and DOWN) and which path is currently active and carrying traffic (ACTIVE state).

The PCEP support of an RSVP-TE LSP differs from that of an SR-TE LSP in that PCE initiated RSVP-TE LSPs are not supported.

Feature Configuration

The following MPLS-level and LSP-level CLI commands, used to configure RSVP-TE LSP in a router acting as a PCEP Client (PCC).

- **config**>**router**>**mpls**>**pce-report rsvp-te** {**enable** | **disable**}
- **config**>**router**>**mpls**>**lsp**>**path-profile** *profile-id range* [**path-group** *group-id range*]
- **config**>**router**>**mpls**>**lsp**>**pce-report** {**enable** | **disable** | **inherit**}
- **config**>**router**>**mpls**>**lsp**>**path-computation-method pce**
- **config**>**router**>**mpls**>**lsp**>**pce-control**

Note: The PCE function implemented in the Nokia Network Services Platform (NSP) and referred to as the Network Resource Controller for Packet (NRC-P), supports only Shared Explicit (SE) style bandwidth management for TE LSPs. The PCEP protocol does not have means for the PCC to convey this value to the PCE, so, regardless of whether the LSP configuration option **rsvp-resv-style** is set to **se** or **ff**, the PCE will always use the SE style in the CSPF computation of the path for a PCE-computed or PCE-controlled RSVP-TE LSP.

A **one-hop-p2p** or a **mesh-p2p** RSVP-TE **auto-lsp** only supports the **pce-report** command in the LSP template:

config>**router**>**mpls**>**lsp-template**>**pce-report** {**enable** | **disable** | **inherit**}

The user must first shut down the LSP template before changing the value of the **pce-report** option.

A manual bypass LSP does not support any of the PCE-related commands. Reporting a bypass LSP to PCE is not required because it does not book bandwidth.

All other MPLS, LSP, and path-level commands are supported, with the exception of **backup-class-type**, **class-type**, **least-fill**, **main-ct-retry-limit**, **mbb-prefer-current-hops**, and **srlg** (on secondary standby path), which, if enabled, will result in a no operation.

The same instantiation modes are supported for RSVP-TE PCC-initiated LSPs as the SR-TE PCC-initiated LSPs. See LSP Initiation for more information.

Behavior of the LSP Path Update

When the **pce-control** option is enabled, the PCC delegates the control of the RSVP-TE LSP to the PCE.

The NRC-P sends a path update using the PCUpd message in the following cases:

- a failure event that impacts a link or a node in the path of a PCE-controlled LSP

  The operation is performed by the PCC as an MBB if the LSP remained in the UP state due to protection provided by FRR or a secondary path. If the LSP went down, then the update brings it into the UP state. A PCRpt message is sent by the PCC for each change to the state of the LSP during this process.
- a topology change that impacts a link in the path of a PCE-controlled LSP

  This topology change can be a change to the IGP metric, the TE metric, admin-group, or SRLG membership of an interface. This update is performed as an MBB by the PCC.
- the user performed a manual resignal of PCE-controlled RSVP-TE LSP path from the NRC-P

  This update is performed as an MBB by the PCC.
- the user performed a Global Concurrent Optimization (GCO) on a set of PCE-controlled RSVP-TE LSPs from the NRC-P

  This update is performed as an MBB by the PCC.

The procedures for the path update are the same as those for an SR-TE LSP. See LSP Initiation for more information. However, the PCUpd message from the PCE does not contain the label for each hop in the computed ERO. PCC then signals the path using the ERO returned by the PCE and, if successful, programs the data path, then sends the PCRpt message with the resulting RRO and hop labels provided by RSVP-TE signaling.

If the signaling of the ERO fails, then the ingress LER returns a PCErr message to PCE with the LSP Error code field of the LSP-ERROR-CODE TLV set to a value of 8 (RSVP signaling error).

If an RSVP-TE LSP has the **no adaptive** option set, the ingress LER cannot perform an MBB for such an LSP. A PCUpd message received from the PCE is then failed by the ingress LER, which returns a PCErr message to PCE with the LSP Error code field of the LSP-ERROR-CODE TLV set to a value of 8 (RSVP signaling error).

When the NRC-P reoptimizes the path of a PCE-controlled RSVP-TE LSP, it is possible that a path that satisfies the constraints of the LSP no longer exists. In this case, the NRC-P sends a PCUpd message with an empty ERO, which forces the PCC to bring down the path of the RSVP-TE LSP.

NRC-P sends a PCUpd message with an empty ERO if the following cases are true.

- The requested bandwidth is the same as current bandwidth, which avoids bringing down the path on a resignal during a MBB transition.
- Local protection is not currently in use, which avoids bringing down a path that activated an FRR backup path. The LSP can remain on the FRR backup path until a new primary path can be found by NRC-P.
- The links of the current path are all operationally up, which allows NRC-P to make sure that the RSVP control plane will report the path down when a link is down and not prematurely bring the path down with an empty ERO.

Behavior of LSP MBB

In addition to the Make-Before-Break (MBB) support when the PCC receives a path update, as described in Behavior of the LSP Path Update, an RSVP-TE LSP supports the MBB procedure for any parameter configuration change, including the PCEP-related commands when they result in a change to the path of the LSP.

If the user adds or modifies the **path-profile** command for an RSVP-TE LSP, a Config Change MBB is only performed if the **path-computation-method pce**, **pce-report**, or **pce-control** options are enabled on the LSP. Otherwise, no action occurs. When **path-computation-method pce**, **pce-report**, or **pce-control** are enabled on the LSP, the Path Update MBB (**tools perform router mpls update-path**) will be failed, resulting in a no operation.

MBB is also supported for the Manual Resignal and Auto-Bandwidth MBB types.

When the LSP goes into a MBB state at the ingress LER, the behavior is dependent on the LSP’s operating mode.

PCE-Controlled LSP

The LSP MBB procedures for a PCE-controlled LSP (**pce-control** enabled) are as follows.

Items 1 through 5 of the following procedures apply to the Config Change, Manual Resignal, and Auto-Bandwidth MBB types. The Delayed Retry MBB type used with the SRLG on secondary standby LSP feature is not supported with a PCE controlled LSP. See Behavior of Secondary LSP Paths for information about the SRLG on secondary standby LSP feature.

1. PCC temporarily removes delegation by sending a PCRpt message for the corresponding PLSP-ID with the delegation D-bit clear.

2. For an LSP with **path-computation-method** disabled, MPLS submits a path request to the local CSPF including the updated path constraints.
3. For an LSP with **path-computation-method pce** enabled, PCC issues a PCReq for the same PLSP-ID and includes the updated constraints in the metric, LSPA, or bandwidth objects. The bandwidth object contains the current operational bandwidth of the LSP in the case of the auto-bandwidth MBB.
   - If the PCE successfully finds a path, it replies with a PCRep message with the ERO.
   - If the PCE does not find a path, it replies with a PCRep message containing the No-Path object.
4. If the local CSPF or the PCE return a path, the PCC performs the following actions.
   - PCC signals the LSP with RSVP control plane and moves traffic to the new MBB path. It then sends a PCRpt message with the delegation D-bit set to return delegation and containing the RRO and LSP object, with the LSP identifiers TLV containing the LSP-ID of the new MBB path. The message includes the metric, LSPA, and bandwidth objects where the P-flag is clear, which indicates the operational values of these parameters. Unless the user disabled the **report-path-constraints** option under the **pcc** context, the PCC also includes a second set of metric, LSPA, or bandwidth objects with the P-flag set to convey to PCE the constraints of the path.
   - PCC sends a PathTear message to delete the state of the older path in the network. PCC then sends a PCRpt message to PCE with the older path PLSP-ID and the remove R-bit set to also have PCE remove the state of that LSP from its database.
5. If the local CSPF or the PCE returns no path or the RSVP-TE signaling of the returned path fails, the router makes no further requests. That is, there is no retry for the MBB.
   - The PCC sends a PCErr message to PCE with the LSP Error code field of the LSP-ERROR-CODE TLV set to a value of 8 (RSVP signaling error) if the MBB failed due to a RSVP-TE signaling error.
   - The PCC sends a PCRpt message with the delegation D-bit set to return delegation and containing the RRO and LSP objects with the LSP identifiers TLV containing the LSP-ID of the currently active path. The message includes the metric, LSPA, and bandwidth objects with the P-flag is clear to indicate the operational values of these parameters. Unless the user disabled the **report-path-constraints** option under the **pcc** context, the PCC also includes a second set of metric, LSPA, and bandwidth objects with the P-flag set to convey to PCE the constraints of the path.
6. The ingress LER takes no action in the case of a network event triggered MBB, such as FRR Global Revertive, TE Graceful Shutdown, or Soft Pre-Emption.

- The ingress PE keeps the information as required and sets the state of MBB to one of the FRR global Revertive, TE Graceful Shutdown, or Soft Pre-emption MBB values but does not perform the MBB action.
- The NRC-P computes a new path in the case of Global Revertive MBB due to a failure event. This computation uses the PCUpd message to update the path using the MBB procedure described in Behavior of the LSP Path Update. The activation of a bypass LSP by a PLR in the network causes the PCC to issue an updated PCRpt message with the new RRO reflecting the PLR and RRO hops. PCE should release the bandwidth on the links that are no longer used by the LSP path.
- The NRC-P computes a new path in the case of the TE graceful MBB if the RSVP-TE is using the TE metric, because the TE metric of the link in TE graceful shutdown is set to infinity. This computation uses the PCUpd message to update the path using the MBB procedure described in Behavior of the LSP Path Update.
- The NRC-P does not act on the TE graceful MBB if the RSVP-TE is using the IGP metric or is on the soft pre-emption MBB; however, the user can perform a manual resignal of the LSP path from the NRC-P to force a new path computation, which accounts for the newly available bandwidth on the link that caused the MBB event. This computation uses the PCUpd message to update the path using the MBB procedure described in Behavior of the LSP Path Update.
- The user can perform a manual resignal of the LSP path from the ingress LER, which forces an MBB for the path as per the remove-delegation/MBB/return-delegation procedures described in this section.
- If the user performs **no pce-control** while the LSP still has the state for any of the network event triggered MBBs, the MBB is performed immediately by the PCC as described in the procedures in PCE-Computed LSP for a PCE-computed LSP and as described in the procedures in PCC-Controlled LSP for a PCC-controlled LSP.

7. The timer-based resignal MBB behaves like the TE graceful or soft pre-emption MBB. The user can perform a manual resignal of the LSP path from the ingress LER or from PCE.
8. The Path Update MBB (**tools perform router mpls update-path**) is failed and will result in a no operation. This is true in all cases when the RSVP-TE LSP enables the **pce-report** option.

PCE-Computed LSP

All MBB types are supported for PCE-computed LSP. The LSP MBB procedures for a PCE-computed LSP (**path-computation-method pce** enabled and **pce-control** disabled) are as follows.

1. PCC issues a PCReq for the same PLSP-ID and includes the updated constraints in the metric, LSPA, and bandwidth objects.
   - If PCE successfully finds a path, it replies with a PCRep message with the ERO.
   - If PCE does not find a path, it replies with a PCRep message containing the No-Path object.
2. If the PCE returns a path, the PCC signals the LSP with RSVP control plane and moves traffic to the new MBB path. If **pce-report** is enabled for this LSP, the PCC sends a PCRpt message with the delegation D-bit clear to retain control and containing the RRO and LSP object with the LSP identifiers TLVs containing the LSP-ID of the new MBB path. The message includes the metric, LSPA, and bandwidth objects where the P-flag is clear, which indicates the operational values of these parameters. Unless the user disables the **report-path-constraints** option under the **pcc** context, PCC also includes a second set of metric, LSPA, and bandwidth objects with the P-flag set to convey to PCE the constraints of the path.
3. If the PCE returns no path or the RSVP-TE signaling of the returned path failed, MPLS puts the LSP into retry mode and sends a request to PCE every *retry-timer* seconds and up to the value of *retry-count*.
4. When the **pce-report** is enabled for the LSP and the FRR Global Revertive MBB is triggered following a bypass LSP activation by a PLR in the network, PCC issues an updated PCRpt message with the new RRO reflecting the PLR and RRO hops. PCE releases the bandwidth on the links that are no longer used by the LSP path.
5. If the user changes the RSVP-TE LSP configuration from **path-computation-method pce** to **no path-computation-method**, then MBB procedures are not supported. In this case, the LSP path is torn down and is put into retry mode to compute a new path from the local CSPF on the router to signal it.

PCC-Controlled LSP

All MBB types are supported for PCC-controlled LSP. The LSP MBB procedures for a PCC-controlled LSP (**path-computation-method pce** and **pce-control** disabled) are as follows.

1. MPLS submits a path request, including the updated path constraints, to the local CSPF.
2. If the local CSPF returns a path, PCC signals the LSP with RSVP control plane and moves traffic to the new MBB path. If **pce-report** is enabled for this LSP, the PCC sends a PCRpt message with the delegation bit clear to retain control and containing the RRO and LSP object with the LSP identifiers TLV containing the LSP-ID of the new MBB path. It includes the metric, LSPA, and bandwidth

objects where the P-flag is clear, which indicates the operational values of these parameters. Unless the user disables the **report-path-constraints** option under the **pcc** context, PCC also includes a second set of metric, LSPA, and bandwidth objects with the P-flag set to convey to PCE the constraints of the path.

3. If the CSPF returns no path or the RSVP-TE signaling of the returned path fails, MPLS puts the LSP into retry mode and sends a request to the local CSPF every *retry-timer* seconds and up to the value of *retry-count*.
4. When **pce-report** is enabled for the LSP and the FRR Global Revertive MBB is triggered following a bypass LSP activation by a PLR in the network, PCC issues an updated PCRpt message with the new RRO reflecting the PLR and RRO hops. PCE releases the bandwidth on the links that are no longer used by the LSP path.

5.1.5.2.5 Behavior of Secondary LSP Paths

Each of the primary, secondary standby, and secondary non-standby paths of the same LSP must use a separate PLSP-ID. In the PCE function of the NSP, the NRC-P, checks the LSP-IDENTIFIERS TLVs in the LSP object and can identify which PLSP-IDs are associated with the same LSP or the same RSVP session. The parameters are the IPv4 Tunnel Sender Address, the Tunnel ID, the Extended Tunnel ID, and the IPv4 Tunnel Endpoint Address. This approach allows the use of all the PCEP procedures for all three types of LSP paths.

PCC indicates to PCE the following states for the path in the LSP object: down, up (signaled but is not carrying traffic), or active (signaled and carrying traffic).

PCE tracks active paths and displays them in the NSP GUI. It also provides only the tunnel ID of an active PLSP-ID to a specific destination prefix when a request is made by a service or a steering application.

PCE recomputes the paths of all PLSP-IDs that are affected by a network event. The user can select each path separately on the NSP GUI and trigger a manual resignal of one or more paths of the LSP.

Note: Enabling the **srlg** option on a secondary standby path results in a **no** operation. The NRC-P supports link and SRLG disjointness using the PCE path profile, and the user can apply to the primary and secondary paths of the same LSP. See PCE Path Profile Support for more information.

5.1.5.2.6 PCE Path Profile Support

The PCE path profile ID and path group ID are configured at the LSP level.

The NRC-P can enforce path disjointness and bidirectionality among a pair of forward and a pair of reverse LSP paths. Both pairs of LSP paths must use a unique path group ID along with the same Path Profile ID, which is configured on the NRC-P to enforce path disjointness or path bidirectionality constraints.

When the user wants to apply path disjointness and path bidirectionality constraints to LSP paths, it is important to follow the following guidelines. The user can configure the following sets of LSP paths.

- Configure a set consisting of a pair of forward LSPs and a pair of reverse LSPs each with a single path, primary or secondary. The pair of forward LSPs can originate and terminate on different routers. The pair of reverse LSPs must mirror the forward pair. In this case, the path profile ID and the path group ID configured for each LSP must match. Because each LSP has a single path, the bidirectionality constraint applies automatically to the forward and reverse LSPs, which share the same originating node and the same terminating routers.
- Configure a pair consisting of a forward LSP and a reverse LSP, each with a primary path and a single secondary path, or each with a couple of secondary paths. Because the two paths of each LSP inherit the same LSP level path profile ID and path group ID configuration, the NRC-P path computation algorithm cannot guarantee that the primary paths in both directions meet the bidirectionality constraint. That is, it is possible that the primary path for the forward LSP shares the same links as the secondary path of the reverse LSP and vice-versa.

5.1.6 LSP Path Diversity and Bidirectionality Constraints

The PCE path profile defined in *draft-alvarez-pce-path-profiles* is used to request path diversity or a disjoint for two or more LSPs originating on the same or different PE routers. It is also used to request that paths of two unidirectional LSPs between the same two routers use the same TE links. This is referred to as the bidirectionality constraint.

Path profiles are defined by the user directly on the NRC-P Policy Manager with a number of LSP path constraints, which are metrics with upper bounds specified, and with an objective, which are metrics optimized with no bound specified. The NRC-P Policy Manager allows the following PCE constraints to be configured within each PCE Path Profile:

- path diversity, node-disjoint, link-disjoint

- path bidirectionality, symmetric reverse route preferred, symmetric reverse route required
- maximum path IGP metric (cost)
- maximum path TE metric
- maximum hop count

The user can also specify which PCE objective to use to optimize the path of the LSP in the PCE Path Profile:

- IGP metric (cost)
- TE metric
- hops (span)

The CSPF algorithm will optimize this objective. If a constraint is provided for the same metric, then the CSPF algorithm makes sure that the selected path achieves a lower or equal value to the bound specified in the constraint.

For hop-count metrics, if a constraint is sent in a METRIC object, and is also specified in a PCE profile referenced by the LSP, the constraint in the METRIC object is used.

For IGP and TE metrics, if an objective is sent in a METRIC object, and is also specified in a PCE profile referenced by the LSP, the objective in the Path Profile is used.

The constraints in the Bandwidth object and the LSPA object, specifically the include/exclude admin-group constraints and setup and hold priorities, are not supported in the PCE profile.

In order to indicate the path diversity and bidirectionality constraints to the PCE, the user must configure the profile ID and path group ID of the PCE path that the LSP belongs to. The CLI for this is described in the Configuring and Operating SR-TE section. The path group ID does not need to be defined in the PCE as part of the path profile configuration, and identifies implicitly the set of paths which must have the path diversity constraint applied.

The user can only associate a single path group ID with a specific PCE path profile ID for a given LSP. However, the same path group ID can be associated with multiple PCE profile IDs for the same LSP.

The path profiles are inferred using the path ID in the path request by the PCC. When the PE router acting as a PCC wants to request path diversity from a set of other LSPs belonging to a path group ID value, it adds a new path profile object into the PCReq message. The object contains the path profile ID and the path group ID as an extended ID field. In other words, the diversity metric is carried in an opaque way from PCC to PCE.

The bidirectionality constraint operates the same way as the diversity constraint. The user can configure a PCE profile with both the path diversity and bidirectionality constraints. PCE will check if there is an LSP in the reverse direction which belongs to the same path group ID as an originating LSP it is computing the path for, and will enforce the constraint.

In order for the PCE to be aware of the path diversity and bidirectionality constraints for an LSP that is delegated but for which there is no prior state in the NRC-P LSP database, the path profile object is included in the PCRpt message with the P-flag set in the common header to indicate that the object must be processed.

Table 32 describes the new objects introduced in the PCE path profile.

Table 32 **PCEP Path Profile Extension Objects and TLVs**

TLV, Object, or Message	Contained in Object	Contained in Message
PATH-PROFILE-CAPABILITY TLV	OPEN	OPEN
PATH-PROFILE Object	—	PCReq, PCRpt, PCInitiate

A path profile object can contain multiple TLVs containing each profile-id and extend-id, and should be processed properly. If multiple path profile objects are received, the first object is interpreted and the others are ignored. The PCC and the PCE support all PCEP capability TLVs defined in this chapter and will always advertise them. If the OPEN object received from a PCEP speaker does not contain one or more of the capabilities, the PCE or PCC will not use them during that PCEP session.

5.1.7 Path Computation Fallback for PCC-Initiated LSPs

For PCC-initiated RSVP-TE LSPs, the router supports fallback to a local path computation method in the case where the configured PCEP sessions are down or the PCE is unreachable, or when all configured PCEs are signaling overload and the redelegation timer expires while all configured LSPs are signaling overload so that the LSP cannot be redelegated. The fallback method can be configured to be the local CSPF or none. In the latter case, MPLS uses the explicit IGP path (RSVP-TE LSPs).

This capability is supported by both active and passive stateful LSPs. Active stateful LSPs are fully delegated to the PCE by being both PCE computed (**path-computation-method pce**) and PCE controlled. Passive stateful LSPs are PCE computed.

Note: For the passive stateful case, it is important that the **retry-timer** and **retry-limit** values exceed the **redelegation-timer** value, otherwise, the LSP may go operationally down before the fallback path computation has occurred.

A fallback path computation method is configured as follows:

```
configure>router>
   mpls
      lsp <xyz>
         pce-control
         path-computation-method {pce | local-cspf}
         fallback-path-computation-method {none | local-cspf}
```

If **none** is configured, MPLS uses the default method based on the configured path, which is hop-to-label path computation for SR-TE LSPs and IGP-based path computation for RSVP-TE LSPs.

The **fallback-path-computation-method** command is only valid for **path-computation-method pce**, irrespective of whether **pce-control** is configured. It is mutually exclusive with the **path-computation-method local-cspf** and **no path-computation-method** commands.

The fallback mechanism is only triggered if PCC informs MPLS that PCEP is down. It is not triggered while the PCC is administratively down or not yet configured.

Note: On the first local path computation following a fallback, MPLS is not aware of the list of SRLGs or administrative groups that are used by the original path computed by PCE. As a result, MPLS can only provide a list of hops or links to avoid on the first computation.

PCE reports are sent, where applicable, with the delegation bit cleared.

5.2 NSP and VSR-NRC PCE Redundancy

This feature introduces resilience support to the PCE and PCC capabilities.

5.2.1 Feature Configuration

In Release 16.0.R4, a CLI command parameter is introduced in the PCC for configuring the PCEP session to the standby backup PCE peer. A **preference** parameter value is used to decide the primary and the standby backup PCE peer:

# **configure router pcep pcc peer** *ip-address* [**preference** *preference*]

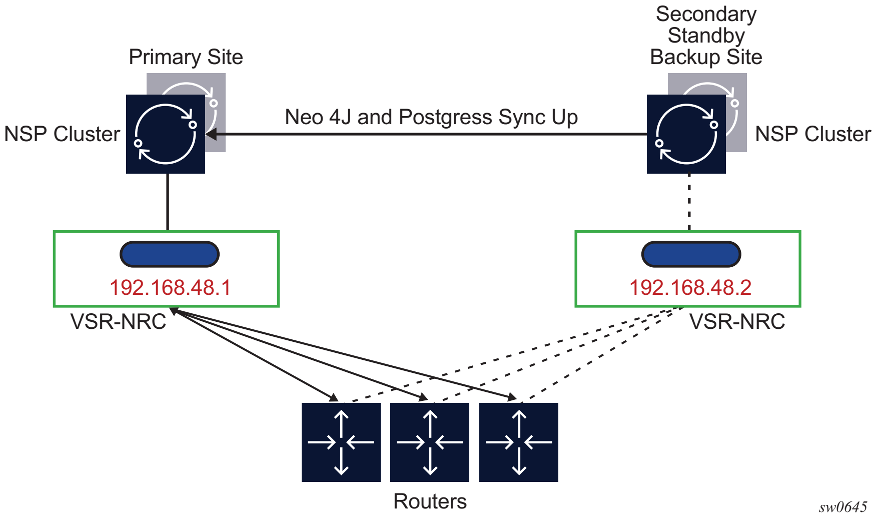
A maximum of two PCE peers are supported. The PCE peer that is not in overload is always selected by the PCC as the active PCE. However, if neither of the PCEs are signaling the overload state, the PCE with the higher numerical preference value is selected. In case of a tie, the PCE with the lower IP address is selected.

In order to change the value of the **preference** parameter, the peer must be deleted and recreated.

5.2.2 Feature Behavior

Figure 62 illustrates the NSP ecosystem and the provision of resilience across two separate sites. This is referred to as Disaster Recovery (DR) and is also sometimes referred to as geo-redundancy.

Figure 62 **NSP Ecosystem Resilience**



NSP ecosystem resilience consists of two mechanisms that can be deployed separately or together:

- High-Availability (HA) at a single site
  - NSP, where the applications reside, is protected by a cluster of three Virtual Machines (VMs)
  - the VSR-NRC module, which implements PCEP, OpenFlow, and BGP-LS/IGP, does not support HA and is deployed with a single VM which contains the combined CPM and IOM codes
- DR, which consists of a primary site and a secondary standby backup site. Each site consists of an NSP cluster and an VSR-NRC VM complex. A heartbeat protocol runs between the NSP clusters at the primary site and the standby backup sites.

  The VSR-NRC can be deployed as a standalone configuration; however, the NSP must be deployed in a cluster at each site. This configuration is also referred to as a 3+3 deployment.

Each parent NSP cluster establishes a reliable TCP session with a virtual IP to the local VSR-NRC. The TCP session runs an internal protocol, also known as cproto. This configuration is done prior to system startup and cannot be changed with an active NSP; the NSP must be shut down for any changes.

5.2.2.1 NSP Cluster Behavior

The following describes NSP cluster rules:

- At a single site, a master is elected among the cluster of three VMs. Between sites, a single cluster at one site, is the primary/active site and the other DR site is the secondary/standby site.
- The application processes at the standby site are shut down, but the neo4j and other databases are synchronized with the primary/active site.
- Switching to the standby site can be initiated manually or by using an automated approach stemming from the loss of heartbeat between the primary and standby sites.
- When the NSP cluster at the primary/active site is down (two out of three servers must be inactive, shut down, or failed), the heartbeat mechanism between the primary and standby NSP clusters fails after three timeouts. This initiates the activity at the inactive secondary/standby site.
- When the NSP cluster at the primary site is back up, the heartbeat mechanism between the primary/standby and secondary/active NSP clusters is restored. The primary site can be restored to the active site manually. Automatic reversion to the primary NSP cluster is not supported.

5.2.2.2 VSR-NRC Behavior

The following describes VSR-NRC rules:

- steady state behavior
  - The VSR-NRC at the secondary/standby site, in the same way as the VSR-NRC at the primary/active site, establishes PCEP sessions to the PCCs. However, the VSR-NRC at the standby site has its PCEP sessions to the PCCs in the overload state. The VSR-NRC enters this PCEP overload state when its upstream cproto session to the NSP cluster is down, resulting from either the NSP cluster going into the standby state or the cproto session failing.
  - The VSR-NRC acting as a PCE signals the overload state to the PCCs in a PCEP notification message. In the overload state, the VSR-NRC PCE accepts reports (PCRpt) without delegation but rejects requests (PCReq) and reject reports (PCRpt) with delegation. The VSR-NRC PCE also does not originate initiate messages (PCInitiate) and update messages (PCUpd).
  - The VSR-NRC at the secondary/standby site maintains its BGP and IGP peerings with the network and updates its TE database as a result of any network topology changes.

- primary/active NSP cluster failure

  When the NSP cluster at the primary/active site is down (two out of three servers must be inactive, shut down, or failed), the heartbeat mechanism between the primary/active and secondary/standby NSP clusters fails. This initiates the NSP cluster activity at the secondary/standby site.

  The following are the activities on the VSR-NRC:

  - The VSR-NRC at the primary site detects cproto session failure and puts all its PCEP sessions to the PCCs into the overload state.
  - The NSP cluster at the secondary site establishes the cproto session to the local VSR-NRC which then brings its PCEP sessions out of the overload state.
  - The VSR-NRC at the secondary site begins synchronizing the TE and LSP databases with the parent NSP cluster at the secondary site that is now the active site.
  - The VSR-NRC at the primary site must also return the delegation of all LSPs back to the PCCs by sending an empty LSP Update Request that has the Delegate flag set to 0 as per RFC 8231. This allows the PCCs to delegate all eligible LSPs, including PCE-initiated LSPs, to the PCE function in the VSR-NRC at the secondary site.

➡ **Note:** If the entire primary site fails, the above actions of the VSR-NRC at the primary site do not apply; however, the remaining actions do apply.

- VSR-NRC complex failure at the primary site (NSP server is still up)

  A VSR-NRC complex failure at the primary/active NSP site does not initiate an NSP switchover to the secondary/standby NSP site. If the VSR-NRC at the primary site does not recover, a manual switchover to the secondary NSP site is required. The VSR-NRC failure causes alarms to be raised on the NSP (cproto session failure alarm indicating that the NSP cannot communicate with the VSR-NRC). An operator can manually perform a switchover of the NSP activity to the secondary site.

5.2.2.3 PCC Behavior

The following describes PCC rules:

- PCCs can establish upstream PCEP sessions with at most two VSR-NRC PCEs.

- Each upstream session has a preference that takes effect when both upstream PCEP sessions are successfully established. The PCE peer that is not in overload is always selected by the PCC as the active PCE. However, if neither of the PCEs are signaling the overload state, the PCE with the higher numerical preference value is selected, and in case of a tie, the PCE with the lower IP address is selected.
- In the steady state, because one upstream VSR-NRC PCE is in overload, only one PCEP session is active. The PCCs delegate an LSP using a report message (PCRpt) with the Delegate flag set to the active VSR-NRC PCE only. Request messages (PCReq) are not sent to the secondary/standby VSR-NRC PCE in overload. PCRpt messages are sent with the Delegate flag clear to the secondary/standby VSR-NRC PCE in overload.
- If the current active PCEP session signals overload state, the PCC will select the other PCE as the active PCE as long as the corresponding PCEP session is not in overload. Any new path request message (PCReq) or path report message (PCRpt) with the Delegate flag set, is sent to the new PCE.

  The PCE in overload should return the delegation of all existing LSPs back to this PCC by sending an empty LSP Update Request that has the Delegate flag set as per RFC 8231. This PCC will then delegate these LSPs to the new active PCE by sending a path report (PCRpt) with the Delegate flag set.
- If the current active PCEP session goes operationally down, the PCC starts the redelegation timer (default 90 seconds) and state timeout timer (default 180 seconds).
  - If the PCEP session is restored before the redelegation timer expires, no delegation change is performed and the LSP state is maintained.
  - Upon expiration of the redelegation timer, the PCC looks for the other PCEP session and, if not in overload, it immediately delegates the LSPs to the newly active PCE. If the new PCE accepts the delegation, the LSP state is maintained.
  - If the PCEP session does not recover before the redelegation timer expires and the PCC fails to find another active PCEP session, then by default the PCC clears the LSP state of PCE-initiated LSPs after state timeout expiry; the PCC deletes the PCE-initiated LSPs and releases all their resources. A configuration option of the redelegation timer CLI command allows the user to keep the state of the pce-initiated LSPs instead. The PCC does not clear the state of PCC-initiated LSPs; however, the user can do this by deleting the configuration.

5.3 Configuring and Operating RSVP-TE LSP with PCEP

This section provides information about configuring and operating RSVP-TE LSP with PCEP using CLI.

The following describes the detailed configuration of an inter-area RSVP-TE LSP with both a primary path and a secondary path. The network uses IS-IS with the backbone area in Level 2 and the leaf areas in Level 1. Topology discovery is learned by NRC-P using BGP-LS.

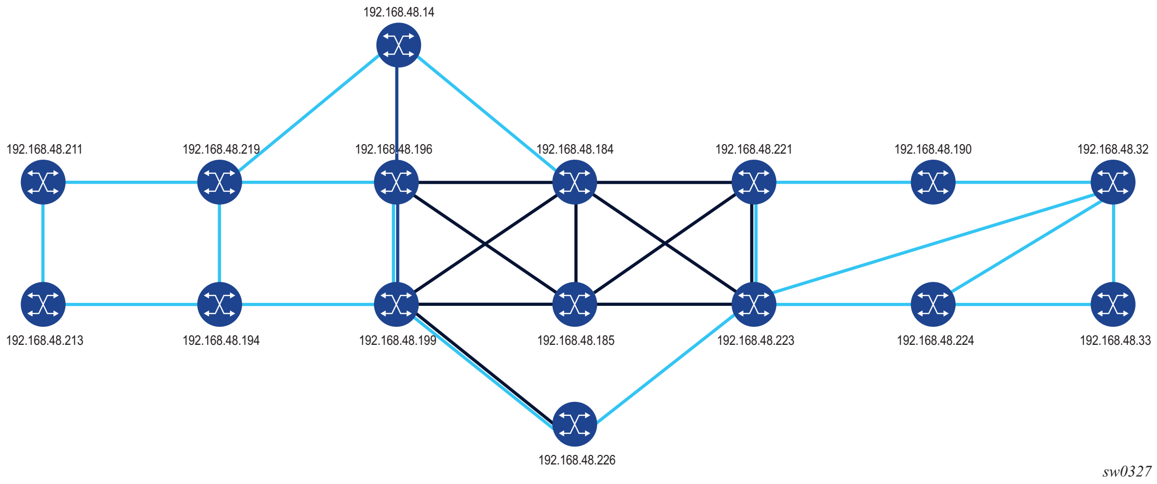
The LSP uses an admin-group constraint to keep the paths of the secondary and primary link disjoint in the backbone area. The LSP is PCE-controlled but also has **path-computation-method pce** enabled so the initial path, and any MBB path, is also computed by PCE.

The NSP and SR OS load versions used to produce this example are:

- NSP: NSP-2.0.3-rel.108
- PCE SR OS: TiMOS-B-0.0.W129
- PCC: TiMOS-B-0.0.I4902

Figure 63 shows a multi-level IS-IS topology in the NSP GUI:

Figure 63 **Multi-level IS-IS Topology in the NSP GUI**



The following example shows the configuration and **show** command output of the PCEP on the PCE node and the PCC node.

```
*A:PCE Server 226>config>router>pcep>pce# info
```

```
----------------------------------------------
            local-address 192.168.48.226
            no shutdown
----------------------------------------------
*A:Reno 194>config>router>pcep>pcc# info
----------------------------------------------
            peer 192.168.48.226
                no shutdown
            exit
            no shutdown
----------------------------------------------

*A:PCE Server 226>config>router>pcep>pce# show router pcep pce status
===============================================================================
Path Computation Element Protocol (PCEP) Path Computation Element (PCE) Info
===============================================================================
Admin Status           : Up                Oper Status          : Up
Unknown Msg Limit      : 10 msg/min
Keepalive Interval     : 30 seconds       DeadTimer Interval   : 120 seconds
Capabilities List      : stateful-delegate stateful-pce segment-rt-path
Local Address          : 192.168.48.226
PCE Overloaded         : false
-------------------------------------------------------------------------------
PCEP Path Computation Element (PCE) Peer Info
-------------------------------------------------------------------------------
Peer                       Sync State            Oper Keepalive/Oper DeadTimer
-------------------------------------------------------------------------------
192.168.48.190:4189        done                  30/120
192.168.48.194:4189        done                  30/120
192.168.48.198:4189        done                  30/120
192.168.48.199:4189        done                  30/120
192.168.48.219:4189        done                  30/120
192.168.48.221:4189        done                  30/120
192.168.48.224:4189        done                  30/120
-------------------------------------------------------------------------------
===============================================================================

*A:Reno 194# show router pcep pcc status
=======================================================================
Path Computation Element Protocol (PCEP) Path Computation Client (PCC) Info
=======================================================================
Admin Status           : Up                Oper Status          : Up
Unknown Msg Limit      : 10 msg/min
Keepalive Interval     : 30 seconds       DeadTimer Interval   : 120 seconds
Capabilities List      : stateful-delegate stateful-pce segment-rt-path
Address                : 192.168.48.194
Report Path Constraints: True
-----------------------------------------------------------------------
PCEP Path Computation Client (PCC) Peer Info
-----------------------------------------------------------------------
Peer                       Admin State/Oper State Oper Keepalive/Oper DeadTimer
-----------------------------------------------------------------------
192.168.48.226             Up/Up                 30/120
-----------------------------------------------------------------------
=======================================================================

*A:Reno 194# show router pcep pcc lsp-db
=======================================================================
PCEP Path Computation Client (PCC) LSP Update Info
```

```
=======================================================================
PCEP-specific LSP ID: 11
LSP ID              : 14378             LSP Type            : rsvp-p2p
Tunnel ID           : 1                 Extended Tunnel Id  : 192.168.48.194
LSP Name            : From Reno to Atlanta RSVP-TE::primary_empty
Source Address      : 192.168.48.194    Destination Address : 192.168.48.224
LSP Delegated       : True              Delegate PCE Address: 192.168.48.226
Oper Status         : active
-----------------------------------------------------------------------
PCEP-specific LSP ID: 12
LSP ID              : 14380             LSP Type            : rsvp-p2p
Tunnel ID           : 1                 Extended Tunnel Id  : 192.168.48.194
LSP Name            : From Reno to Atlanta RSVP-TE::secondary_empty
Source Address      : 192.168.48.194    Destination Address : 192.168.48.224
LSP Delegated       : True              Delegate PCE Address: 192.168.48.226
Oper Status         : up
=======================================================================
```

The following examples shows the configuration and **show** command output of BGP on the PCE node and the ABR node-to-learn topology using the BGP-LS NLRI family.

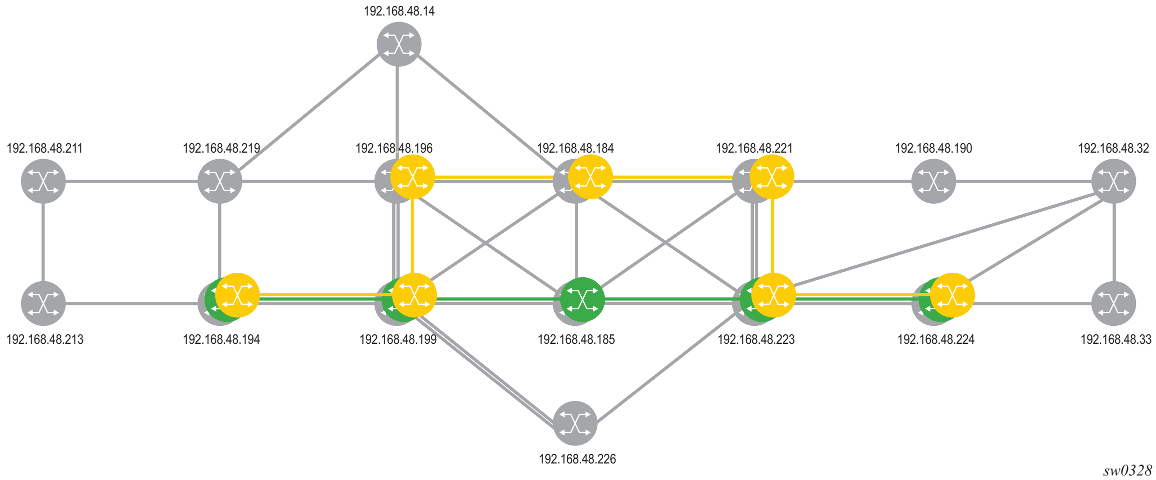
```
*A:PCE Server 226>config>router>bgp# info
----------------------------------------------
            family bgp-ls
            min-route-advertisement 1
            link-state-export-enable
            group "IBGP_L2"
                family bgp-ls
                peer-as 65000
                neighbor 192.168.48.198
                exit
                neighbor 192.168.48.199
                exit
                neighbor 192.168.48.221
                exit
            exit
            no shutdown
----------------------------------------------

*A:Chicago 221>config>router>bgp# info
----------------------------------------------
            min-route-advertisement 1
            advertise-inactive
            link-state-import-enable
            group "IBGP_L2"
                family bgp-ls
                peer-as 65000
                neighbor 192.168.48.226
                exit
            exit
            no shutdown
----------------------------------------------

*A:PCE Server 226# show router bgp summary
===============================================================================
 BGP Router ID:192.168.48.226    AS:65000       Local AS:65000
===============================================================================
BGP Admin State         : Up          BGP Oper State              : Up
Total Peer Groups       : 1           Total Peers                 : 3
Total BGP Paths         : 182         Total Path Memory           : 44896
Total IPv4 Remote Rts   : 0           Total IPv4 Rem. Active Rts  : 0
Total McIPv4 Remote Rts : 0           Total McIPv4 Rem. Active Rts: 0
Total McIPv6 Remote Rts : 0           Total McIPv6 Rem. Active Rts: 0
Total IPv6 Remote Rts   : 0           Total IPv6 Rem. Active Rts  : 0
Total IPv4 Backup Rts   : 0           Total IPv6 Backup Rts       : 0
Total Supressed Rts     : 0           Total Hist. Rts             : 0
Total Decay Rts         : 0
Total VPN Peer Groups   : 0           Total VPN Peers             : 0
Total VPN Local Rts     : 0
Total VPN-IPv4 Rem. Rts : 0           Total VPN-IPv4 Rem. Act. Rts: 0
Total VPN-IPv6 Rem. Rts : 0           Total VPN-IPv6 Rem. Act. Rts: 0
Total VPN-IPv4 Bkup Rts : 0           Total VPN-IPv6 Bkup Rts     : 0
Total VPN Supp. Rts     : 0           Total VPN Hist. Rts         : 0
Total VPN Decay Rts     : 0
Total L2-VPN Rem. Rts   : 0           Total L2VPN Rem. Act. Rts   : 0
Total MVPN-IPv4 Rem Rts : 0           Total MVPN-IPv4 Rem Act Rts : 0
Total MDT-SAFI Rem Rts  : 0           Total MDT-SAFI Rem Act Rts  : 0
Total MSPW Rem Rts      : 0           Total MSPW Rem Act Rts      : 0
Total RouteTgt Rem Rts  : 0           Total RouteTgt Rem Act Rts  : 0
Total McVpnIPv4 Rem Rts : 0           Total McVpnIPv4 Rem Act Rts : 0
Total McVpnIPv6 Rem Rts : 0           Total McVpnIPv6 Rem Act Rts : 0
Total MVPN-IPv6 Rem Rts : 0           Total MVPN-IPv6 Rem Act Rts : 0
Total EVPN Rem Rts      : 0           Total EVPN Rem Act Rts      : 0
Total FlowIpv4 Rem Rts  : 0           Total FlowIpv4 Rem Act Rts  : 0
Total FlowIpv6 Rem Rts  : 0           Total FlowIpv6 Rem Act Rts  : 0
Total LblIpv4 Rem Rts   : 0           Total LblIpv4 Rem. Act Rts  : 0
Total LblIpv6 Rem Rts   : 0           Total LblIpv6 Rem. Act Rts  : 0
Total LblIpv4 Bkp Rts   : 0           Total LblIpv6 Bkp Rts       : 0
Total Link State Rem Rts: 271         Total Link State Rem. Act Rts: 0
===============================================================================
BGP Summary
===============================================================================
Legend : D - Dynamic Neighbor
===============================================================================
Neighbor
Description
                   AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                      PktSent OutQ
-------------------------------------------------------------------------------
192.168.48.198
               65000        0    0 02h42m56s Active
                            0    0
192.168.48.199
               65000      503    0 02h42m56s 76/0/0 (LinkState)
                          328    0
192.168.48.221
               65000      519    0 02h42m56s 195/0/0 (LinkState)
                          328    0
-------------------------------------------------------------------------------

*A:PCE Server 226# show router bgp routes bgp-ls hunt link
===============================================================================
 BGP Router ID:192.168.48.226    AS:65000       Local AS:65000
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
```

```
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete
===============================================================================
BGP-LS Link NLRIs
===============================================================================
-------------------------------------------------------------------------------
RIB In Entries
-------------------------------------------------------------------------------
Network:
 Type          : LINK-NLRI
 Protocol      : ISIS Level-2           Identifier     : 0xa
 Local Node descriptor:
  Autonomous System  : 0.0.253.232
  Link State Id      : 10
  IGP Router Id      : 0x38120048184
 Remote Node descriptor:
  Autonomous System  : 0.0.253.232
  Link State Id      : 10
  IGP Router Id      : 0x38120048223
 Link descriptor:
  IPV4 Interface Addr: 10.0.14.184
  IPV4 Neighbor Addr : 10.0.14.223
Nexthop        : 192.168.48.199
From           : 192.168.48.199
Res. Nexthop   : 0.0.0.0
Local Pref.    : 100                    Interface Name : NotAvailable
Aggregator AS  : None                   Aggregator     : None
Atomic Aggr.   : Not Atomic             MED            : None
AIGP Metric    : None
Connector      : None
Community      : No Community Members
Cluster        : No Cluster Members
Originator Id  : None                   Peer Router Id : 192.168.48.199
Flags          : Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                      Dest Class     : 0
Add Paths Send : Default
Last Modified  : 02h27m50s
-------------------------------------------------------------------------------
Link State Attribute TLVs :
 Administrative group (color) : 0x0
 Maximum link bandwidth : 100000 Kbps
 Max. reservable link bandwidth : 100000 Kbps
 Unreserved bandwidth0 : 100000 Kbps
 Unreserved bandwidth1 : 100000 Kbps
 Unreserved bandwidth2 : 100000 Kbps
 Unreserved bandwidth3 : 100000 Kbps
 Unreserved bandwidth4 : 100000 Kbps
 Unreserved bandwidth5 : 100000 Kbps
 Unreserved bandwidth6 : 100000 Kbps
 Unreserved bandwidth7 : 100000 Kbps
 TE Default Metric : 100
 IGP Metric : 100
 Adjacency Segment Identifier (Adj-SID) :       flags 0x30 weight 0 sid 262136
-------------------------------------------------------------------------------
```

```
Network:
 Type           : LINK-NLRI
 Protocol       : ISIS Level-2           Identifier     : 0xa
 Local Node descriptor:
  Autonomous System  : 0.0.253.232
  Link State Id      : 10
  IGP Router Id      : 0x38120048184
 Remote Node descriptor:
  Autonomous System  : 0.0.253.232
  Link State Id      : 10
  IGP Router Id      : 0x38120048223
 Link descriptor:
  IPV4 Interface Addr: 10.0.14.184
  IPV4 Neighbor Addr : 10.0.14.223
Nexthop        : 192.168.48.221
From           : 192.168.48.221
Res. Nexthop   : 0.0.0.0
Local Pref.    : 100                     Interface Name : NotAvailable
Aggregator AS  : None                    Aggregator     : None
Atomic Aggr.   : Not Atomic              MED            : None
AIGP Metric    : None
Connector      : None
Community      : No Community Members
Cluster        : No Cluster Members
Originator Id  : None                    Peer Router Id : 192.168.48.221
Flags          : Valid  IGP
TieBreakReason : OriginatorID
Route Source   : Internal
AS-Path        : No As-Path
Route Tag      : 0
Neighbor-AS    : N/A
Orig Validation: N/A
Source Class   : 0                       Dest Class     : 0
Add Paths Send : Default
Last Modified  : 02h27m54s
-------------------------------------------------------------------------------
Link State Attribute TLVs :
 Administrative group (color) : 0x0
 Maximum link bandwidth : 100000 Kbps
 Max. reservable link bandwidth : 100000 Kbps
 Unreserved bandwidth0 : 100000 Kbps
 Unreserved bandwidth1 : 100000 Kbps
 Unreserved bandwidth2 : 100000 Kbps
 Unreserved bandwidth3 : 100000 Kbps
 Unreserved bandwidth4 : 100000 Kbps
 Unreserved bandwidth5 : 100000 Kbps
 Unreserved bandwidth6 : 100000 Kbps
 Unreserved bandwidth7 : 100000 Kbps
 TE Default Metric : 100
 IGP Metric : 100
 Adjacency Segment Identifier (Adj-SID) :       flags 0x30 weight 0 sid 262136
-------------------------------------------------------------------------------
```

Figure 64 shows primary and secondary RSVP-TE LSP paths in the NSP GUI.

Figure 64 **Primary and Secondary RSVP-TE LSP Paths in the NSP GUI**



The following example shows the configuration and **show** command output of the MPLS on the PCC node.

```
*A:Reno 194>config>router>mpls>lsp# info
----------------------------------------------
                to 192.168.48.224
                egress-statistics
                    shutdown
                exit
                fast-reroute facility
                    no node-protect
                exit
                path-computation-method pce
                pce-report enable
                pce-control
                revert-timer 1
                primary "primary_empty"
                    exclude "top"
                    bandwidth 10
                exit
                secondary "secondary_empty"
                    standby
                    exclude "bottom"
                    bandwidth 5
                exit
                no shutdown
----------------------------------------------

*A:Reno 194# show router mpls lsp "From Reno to Atlanta RSVP-TE" path detail
===============================================================================
MPLS LSP From Reno to Atlanta RSVP-TE Path  (Detail)
===============================================================================
Legend :
    @ - Detour Available                  # - Detour In Use
    b - Bandwidth Protected               n - Node Protected
    s - Soft Preemption
    S - Strict                            L - Loose
    A - ABR
===============================================================================
-------------------------------------------------------------------------------
LSP From Reno to Atlanta RSVP-TE Path primary_empty
-------------------------------------------------------------------------------
LSP Name         : From Reno to Atlanta RSVP-TE
Path LSP ID      : 14382
From             : 192.168.48.194       To                   : 192.168.48.224
Admin State      : Up                   Oper State           : Up
Path Name        : primary_empty        Path Type            : Primary
Path Admin       : Up                   Path Oper            : Up
Out Interface    : 1/1/1                Out Label            : 262094
Path Up Time     : 0d 00:00:22          Path Down Time       : 0d 00:00:00
Retry Limit      : 0                    Retry Timer          : 30 sec
Retry Attempt    : 0                    Next Retry In        : 0 sec
BFD Template     : None                 BFD Ping Interval    : 60
BFD Enable       : False
Adspec           : Disabled             Oper Adspec          : Disabled
CSPF             : Enabled              Oper CSPF            : Enabled
Least Fill       : Disabled             Oper LeastFill       : Disabled
FRR              : Enabled              Oper FRR             : Enabled
FRR NodeProtect  : Disabled             Oper FRR NP          : Disabled
FR Hop Limit     : 16                   Oper FRHopLimit      : 16
FR Prop Admin Gr*: Disabled             Oper FRPropAdmGrp    : Disabled
Propogate Adm Grp: Disabled             Oper Prop Adm Grp    : Disabled
Inter-area       : False
PCE Updt ID      : 0
PCE Report       : Enabled              Oper PCE Report      : Enabled
PCE Control      : Enabled              Oper PCE Control     : Enabled
PCE Compute      : Enabled
Neg MTU          : 1496                 Oper MTU             : 1496
Bandwidth        : 10 Mbps              Oper Bandwidth       : 10 Mbps
Hop Limit        : 255                  Oper HopLimit        : 255
Record Route     : Record               Oper Record Route    : Record
Record Label     : Record               Oper Record Label    : Record
Setup Priority   : 7                    Oper Setup Priority  : 7
Hold Priority    : 0                    Oper Hold Priority   : 0
Class Type       : 0                    Oper CT              : 0
Backup CT        : None
MainCT Retry     : n/a
    Rem          :
MainCT Retry     : 0
    Limit        :
Include Groups   :                      Oper Include Groups  :
None                                           None
Exclude Groups   :                      Oper Exclude Groups  :
top                                            top
Adaptive         : Enabled              Oper Metric          : 40
Preference       : n/a
Path Trans       : 7                    CSPF Queries         : 7172
Failure Code     : noError
Failure Node     : n/a
Explicit Hops    :
    No Hops Specified
Actual Hops      :
    10.202.5.194 (192.168.48.194) @               Record Label        : N/A
 -> 10.202.5.199 (192.168.48.199) @               Record Label        : 262094
 -> 192.168.48.185 (192.168.48.185)               Record Label        : 262111
 -> 10.0.5.185                                    Record Label        : 262111
 -> 192.168.48.223 (192.168.48.223)               Record Label        : 262121
 -> 10.0.7.223                                    Record Label        : 262121
 -> 192.168.48.224 (192.168.48.224)               Record Label        : 262116
 -> 10.101.4.224                                  Record Label        : 262116
Computed Hops    :
    10.202.5.199(S)
 -> 10.0.5.185(S)
 -> 10.0.7.223(S)
 -> 10.101.4.224(S)
Resignal Eligible: False
Last Resignal    : n/a                  CSPF Metric          : 40
-------------------------------------------------------------------------------
LSP From Reno to Atlanta RSVP-TE Path secondary_empty
-------------------------------------------------------------------------------
LSP Name         : From Reno to Atlanta RSVP-TE
Path LSP ID      : 14384
From             : 192.168.48.194       To                   : 192.168.48.224
Admin State      : Up                   Oper State           : Up
Path Name        : secondary_empty      Path Type            : Standby
Path Admin       : Up                   Path Oper            : Up
Out Interface    : 1/1/1                Out Label            : 262091
Path Up Time     : 0d 00:00:25          Path Down Time       : 0d 00:00:00
Retry Limit      : 0                    Retry Timer          : 30 sec
Retry Attempt    : 0                    Next Retry In        : 0 sec
BFD Template     : None                 BFD Ping Interval    : 60
BFD Enable       : False
Adspec           : Disabled             Oper Adspec          : Disabled
CSPF             : Enabled              Oper CSPF            : Enabled
Least Fill       : Disabled             Oper LeastFill       : Disabled
Propogate Adm Grp: Disabled             Oper Prop Adm Grp    : Disabled
Inter-area       : False
PCE Updt ID      : 0
PCE Report       : Enabled              Oper PCE Report      : Enabled
PCE Control      : Enabled              Oper PCE Control     : Enabled
PCE Compute      : Enabled
Neg MTU          : 1496                 Oper MTU             : 1496
Bandwidth        : 5 Mbps               Oper Bandwidth       : 5 Mbps
Hop Limit        : 255                  Oper HopLimit        : 255
Record Route     : Record               Oper Record Route    : Record
Record Label     : Record               Oper Record Label    : Record
Setup Priority   : 7                    Oper Setup Priority  : 7
Hold Priority    : 0                    Oper Hold Priority   : 0
Class Type       : 0                    Oper CT              : 0
Include Groups   :                      Oper Include Groups  :
None                                           None
Exclude Groups   :                      Oper Exclude Groups  :
bottom                                         bottom
Adaptive         : Enabled              Oper Metric          : 60
Preference       : 255
Path Trans       : 28                   CSPF Queries         : 10
Failure Code     : noError
Failure Node     : n/a
Explicit Hops    :
    No Hops Specified
Actual Hops      :
    10.202.5.194 (192.168.48.194)                 Record Label        : N/A
 -> 10.202.5.199 (192.168.48.199)                 Record Label        : 262091
 -> 10.0.9.198 (192.168.48.198)                   Record Label        : 262096
 -> 192.168.48.184 (192.168.48.184)               Record Label        : 262102
 -> 10.0.2.184                                    Record Label        : 262102
 -> 192.168.48.221 (192.168.48.221)               Record Label        : 262119
 -> 10.0.4.221                                    Record Label        : 262119
 -> 192.168.48.223 (192.168.48.223)               Record Label        : 262088
 -> 10.0.10.223                                   Record Label        : 262088
 -> 192.168.48.224 (192.168.48.224)               Record Label        : 262115
 -> 10.101.4.224                                  Record Label        : 262115
Computed Hops    :
    10.202.5.199(S)
 -> 10.0.9.198(S)
 -> 10.0.2.184(S)
 -> 10.0.4.221(S)
 -> 10.0.10.223(S)
 -> 10.101.4.224(S)
Srlg             : Disabled
Srlg Disjoint    : False
Resignal Eligible: False
Last Resignal    : n/a                  CSPF Metric          : 60
===============================================================================
```

6 Segment Routing Policies

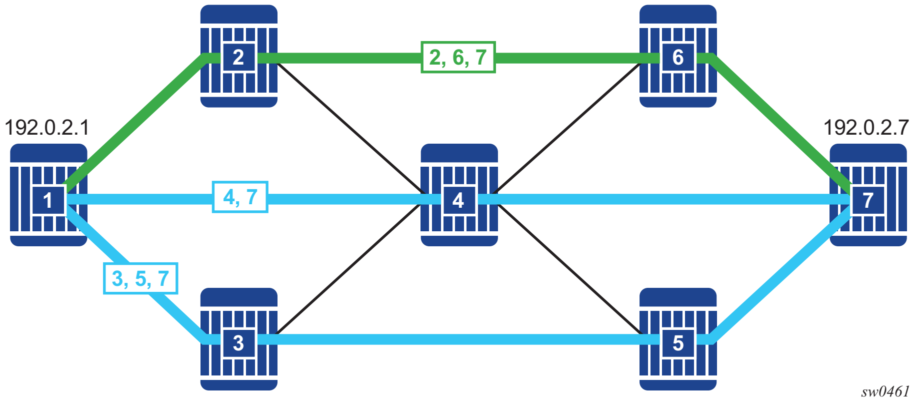
The concept of a Segment Routing (SR) policy is described by the IETF draft *draft-filsfils-spring-segment-routing-policy*. A segment-routing policy specifies a source-routed path from a head-end router to a network endpoint, and the traffic flows that are steered to that source-routed path. A segment-routing policy intended for use by a particular head-end router can be statically configured on that router or advertised to it in the form of a BGP route.

The following terms are important to understanding the structure of a segment routing policy and the relationship between one policy and another.

- Segment-routing policy — a policy identified by the tuple of (head-end router, endpoint and color). Each segment routing policy is associated with a set of one or more candidate paths, one of which is selected to implement the segment routing policy and installed in the dataplane. Certain properties of the segment routing policy come from the currently selected path - for example, binding SID, segment list(s), and so on.
- Endpoint — the far-end router that is the destination of the source-routed path. The endpoint may be null (all-zero IP address) if no specific far-end router is targeted by the policy.
- Color — a property of a segment routing policy that determines the sets of traffic flows that are steered by the policy.
- Path — a set of one or more segment lists that are explicitly or statically configured or dynamically signaled. If a path becomes active then traffic matching the segment routing policy is load-balanced across the segment lists of the path in an equal, unequal, or weighted distribution. Each path is associated with:
  - a protocol origin (BGP or static)
  - a preference value
  - a binding SID value
  - a validation state (valid or invalid)
- Binding SID — a SID value that opaquely represents a segment routing policy (or more specifically, its selected path) to upstream routers. BSIDs provide isolation or decoupling between different source-routed domains and improve overall network scalability. Usually, all candidate paths of a segment routing policy are assigned the same BSID.

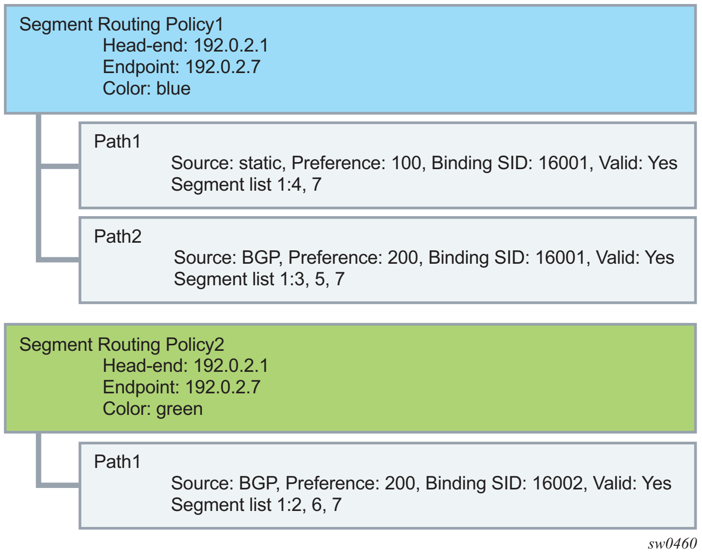
These concepts are illustrated by the following example. Suppose there is a network of 7 nodes as shown in Figure 65 and there are two classes of traffic (blue and green) to be transported between node1 and node 7. There is a segment routing policy for the blue traffic between node1 and node7 and another segment routing policy for the green traffic between these same two nodes.

Figure 65 Network Example with 2 Segment Routing Policies



The two segment routing policies that are involved in this example and the associated relationships are depicted in Figure 66.

Figure 66 Relationship Between Segment Routing Policies and Paths



Segment Routing Policy1
Head-end: 192.0.2.1
Endpoint: 192.0.2.7
Color: blue

Path1
Source: static, Preference: 100, Binding SID: 16001, Valid: Yes
Segment list 1:4, 7

Path2
Source: BGP, Preference: 200, Binding SID: 16001, Valid: Yes
Segment list 1:3, 5, 7

Segment Routing Policy2
Head-end: 192.0.2.1
Endpoint: 192.0.2.7
Color: green

Path1
Source: BGP, Preference: 200, Binding SID: 16002, Valid: Yes
Segment list 1:2, 6, 7

6.1 Statically-Configured Segment Routing Policies

A segment routing policy is statically configured on the router using one of the supported management interfaces. In the Nokia data model, static policies are configured under **config**>**router**>**segment-routing**>**sr-policies**.

There are two types of static policies: local and non-local. A static policy is local when its **head-end** parameter is configured with the value **local**. This means that the policy is intended for use by the router where the static policy is configured. Local static policies are imported into the local segment routing database for further processing. If the local segment routing database chooses a local static policy as the best path for a particular (color, endpoint) then the associated path and its segment lists will be installed into the tunnel table (for next-hop resolution) and as a BSID-indexed MPLS label entry.

A static policy is non-local when its **head-end** parameter is set to any IPv4 address (even an IPv4 address that is associated with the local router, which is a configuration that should generally be avoided). A non-local policy is intended for use by a different router than the one where the policy is configured. Non-local policies are not installed in the local segment routing database and do not affect the forwarding state of the router where they are configured. In order to advertise non-local policies to the target router, either directly (over a single BGP session) or indirectly (using other intermediate routers, such as BGP route reflectors), the static non-local policies must be imported into the BGP RIB and then re-advertised as BGP routes. In order to import static non-local policies into BGP, you must configure the **sr-policy-import** command under **config**>**router**>**bgp**. In order to advertise BGP routes containing segment routing policies, you must add the **sr-policy-ipv4** or the **sr-policy-ipv6** family to the configuration of a BGP neighbor or group (or the entire base router BGP instance) so that the capability is negotiated with other routers.

Local and non-local static policies have the same configurable attributes. The function and rules associated with each attribute are:

- **shutdown** — used to administratively enable or disable the static policy
- **binding-sid** — used to associate a binding SID with the static policy in the form of an MPLS label in the range 32 to 1048575. This is a mandatory parameter. The binding SID must be an available label in the **reserved-label-block** associated with segment routing policies, otherwise the policy cannot be activated.
- **color** — used to associate a color with the static policy. This is a mandatory parameter.

- **distinguisher** — used to uniquely identify a non-local static policy when it is a re-advertised as a BGP route. The value is copied into the BGP NLRI field. A unique distinguisher ensures that BGP does not suppress BGP routes for the same (color, endpoint) but targeted to different head-end routers. This is mandatory for non-local policies but optional in local policies.
- **endpoint** — used to identify the endpoint IPv4 or IPv6 address associated with the static policy. A value of 0.0.0.0 or 0::0 is permitted and interpreted as a null endpoint. This is a mandatory parameter.

Note: When a non-local SR policy with either an IPv4 or IPv6 endpoint is selected for advertisement, the **head-end** parameter supports an IPv4 address only. This is converted into an IPv4-address-specific RT extended community (0x4102) in the advertised route in the BGP Update message.

- **head-end** — used to identify the router that is the targeted node for installing the policy. This is a mandatory parameter. The value **local** must be used when the target is the local router itself. Otherwise, any valid IPv4 address is allowed, and the policy is considered non-local. When a non-local static policy is re-advertised as a BGP route, the configured head-end address is embedded in an IPv4-address-specific route-target extended community that is automatically added to the BGP route.
- **preference** — used to indicate the degree of preference of the policy if the local segment routing database has other policies (static or BGP) for the same (color, endpoint). In order for a path to be selected as the active path for a (color, endpoint), it must have the highest preference value amongst all the candidate paths.

The following are configuration rules related to the previously described attributes:

1. Every static local policy must have a unique combination of **color**, **endpoint**, and **preference**.
2. Every static non-local policy must have a unique distinguisher.

Each static policy (local and non-local) must include, in its configuration, at least one segment-list containing at least one segment. Each static-policy can have up to 32 segment-lists, each containing up to 11 segments. Each segment-list can be assigned a weight to influence the share of traffic that it carries compared to other segment-lists of the same policy. The default weight is 1.

The segment routing policy draft standard allows a segment-list to be configured (and signaled) with a mix of different segment types. When the head-end router attempts to install such a segment routing policy, it must resolve all of the segments into a stack of MPLS labels. In the current SR OS implementation this complexity is avoided by requiring that all (configured and signaled) segments must already be provided in the form of MPLS label values. In terms of the draft standard, this means that only type-1 segments are supported.

6.2 BGP Signaled Segment Routing Policies

The base router BGP instance is configured to send and receive BGP routes containing segment routing policies. In order to exchange routes belonging to the (AFI=1, SAFI=73) or (AFI=2, SAFI=73) address family with a particular base router BGP neighbor, the family configuration that applies to that neighbor must include the **sr-policy-ipv4** or the **sr-policy-ipv6** keyword respectively.

When BGP receives an **sr-policy-ipv4** route (AFI=1, SAFI=73) or a **sr-policy-ipv6 route** (AFI=2, SAFI=73) from a peer, it runs its standard BGP best path selection algorithm to choose the best path for each NLRI combination of distinguisher, endpoint, and color. If the best path is targeted to this router as head-end, BGP extracts the segment routing policy details into the local segment routing database. A BGP segment routing policy route is deemed to be targeted to this router as the head-end if either:

- it has no route-target extended community and a NO-ADVERTISE standard community
- it has an IPv4 address-specific route-target extended community with an IPv4 address matching the system IPv4 address of this router

An **sr-policy-ipv4** or a **sr-policy-ipv6** route can be received from either an IBGP or EBGP peer but it is never propagated to an EBGP peer. An **sr-policy-ipv4** or a **sr-policy-ipv6** route can be reflected to route reflector clients if this is allowed (a NO_ADVERTISE community is not attached) and the router does not consider itself the head-end of the policy.

Note: A BGP segment routing policy route is considered malformed, and triggers error-handling procedures such as session reset or treat-as-withdraw, if it does not have at least one segment-list TLV with at least one segment TLV.

6.3 Segment Routing Policy Path Selection and Tie-Breaking

Segment Routing policies (static and BGP) for which the local router is head-end are processed by the local segment routing database. For each (color, endpoint) combination, the database must validate each candidate path and choose one to be the active path. The steps of this process are outlined in Table 33.

Table 33 **Segment Routing Policy Validation and Selection Process**

Step	Logic
1	Is the path missing a binding SID in the form of an MPLS label? Yes: This path is invalid and cannot be used. No: Go to next step
2	Does the path have any segment-list containing a segment type not equal to 1 (an MPLS label)? Yes: This path is invalid and cannot be used. No: Go to next step
3	Are all segment-lists of the path invalid? A segment-list is invalid if it is empty, if the first SID cannot be resolved to a set of one or more next-hops, or if the weight is 0. Yes: This path is invalid and cannot be used. No: Go to next step
4	Is the binding-SID an available label in the reserved-label-block range? Yes: Go to next step. No: This path is invalid and cannot be used.
5	Is there another path that has reached this step that has a higher preference value? Yes: This path loses the tie-break and cannot be used. No: Go to next step.
6	Is there a static path? Yes: Select the static path as the active path because the protocol-origin value associated with static paths (30) is higher than the protocol-origin value associated with BGP learned paths (20). No: Go to next step.

Table 33 **Segment Routing Policy Validation and Selection Process**

Step	Logic
7	Is there a BGP path with a lower originator value? The originator is a 160-bit numerical value formed by the concatenation of a 32-bit ASN and a 128-bit peer address (with IPv4 addresses encoded in the lowest 32 bits.) Yes: This path loses the tie-break and cannot be used.
8	Is there another BGP path with a higher distinguisher value? Yes: Select the BGP path with the highest distinguisher value.

At step 3 of Table 33, the router attempts to resolve the first segment of each segment-list to a set of one or more next-hops and outgoing labels. It does so by looking for a matching SID in the segment routing module, which must correspond to one of the following:

- SR-ISIS or SR-OSPF node SID
- SR-IS or SR-OSPF adjacency SID
- SR-IS or SR-OSPF adjacency-set SID (parallel or non-parallel set)

Note: The label value in the first segment of the segment-list is matched against ILM label values that the local router has assigned to node-SIDs, adjacency-SIDs, and adjacency-set SIDs. The matched ILM entry may not program a swap to the same label value encoded in the segment routing policy - for example, in the case of an adjacency SID, or a node-SID reachable through a next-hop using a different SRGB base.

6.4 Resolving BGP Routes to Segment Routing Policy Tunnels

When a statically configured or BGP signaled segment routing policy is selected to be the active path for a (color, endpoint) combination, the corresponding path and its segment lists are programmed into the tunnel table of the router. An IPv4 tunnel of type **sr-policy** (**endpoint** parameter is an IPv4 address) is programmed into the IPv4 tunnel table (TTMv4). Similarly, an IPv6 tunnel of type **sr-policy** (**endpoint** parameter is an IPv6 address) is programmed into the IPv6 tunnel table (TTMv6). The resulting tunnel entries can be used to resolve the following types of BGP routes:

- Unlabeled IPv4 routes
- Unlabeled IPv6 routes
- Label-unicast IPv4 routes
- Label-unicast IPv6 (6PE) routes
- VPN IPv4 and IPv6 routes
- EVPN routes

Specifically, an IPv4 tunnel of type **sr-policy** can be used to resolve:

- an IPv4 or the IPv4-mapped IPv6 next hop of the following route families:
  ipv4, **ipv6**, **vpn-ipv4**, **vpn-ipv6**, **label-ipv4**, **label-ipv6**, **evpn**
- the IPv6 next hop of the following route families:
  ipv6, **label-ipv4** and **label-ipv6** (SR policy with **endpoint**=0.0.0.0 only).

An IPv6 tunnel of type **sr-policy** can be used to resolve:

- the IPv6 next hop of the following route families:
  ipv4, **ipv6**, **vpn-ipv4**, **vpn-ipv6**, **label-ipv4**, **label-ipv6**, **evpn**
- the IPv4 next hop of the following route families:
  ipv4 and **label-ipv4** (SR policy with **endpoint**=0::0 only).
- the IPv4-mapped IPv6 next hop of the following route families:
  label-ipv6 (SR policy with **endpoint**=0::0 only).

6.4.1 Resolving Unlabeled IPv4 BGP Routes to Segment Routing Policy Tunnels

For an unlabeled IPv4 BGP route to be resolved by an SR policy:

- A color-extended community must be attached to the IPv4 route
- The base instance BGP next-hop-resolution configuration of **shortcut-tunnel**>**family ipv4** must allow SR policy tunnels

As an example, if under these conditions, there is an IPv4 route with a color-extended community (value C) and BGP next-hop address N. The order of resolution is as follows:

1. If there is an SR policy in TTMv4 for which end-point = BGP next-hop address and color = Cn, then use this tunnel to resolve the BGP next hop.
2. If no SR policy is found in the previous step and the Cn color-extended community has its color-only (CO) bits set to '01' or '10', then try to find in TTMv4 an SR policy for which endpoint = null (0.0.0.0) and color = Cn. If there is such a policy, use it to resolve the BGP next hop.
3. If no SR policy is found in the previous steps and the Cn color-extended community has its CO bits set to '01' or '10', then try to find in TTMv6 an SR policy for which endpoint = null (0::0) and color = Cn. If there is such a policy, use it to resolve the BGP next hop.
4. If no SR policy is found in the previous steps but there is a non-SR policy tunnel in TTMv4 that is allowed by the resolution options and for which endpoint = BGP next-hop address (and for which the admin-tag meets the admin-tag-policy requirements applied to the BGP route, if applicable) then use this tunnel to resolve the BGP next hop if it has the highest TTM preference.
5. Otherwise, fall back to IGP, unless the **disallow-igp** option is configured.

Note: Contrary to section 8.8.2 of draft-filsfils-segment-routing-05, BGP only resolves a route with multiple color-extended communities to an SR policy using the color-extended community with the highest value.

6.4.2 Resolving Unlabeled IPv6 BGP Routes to Segment Routing Policy Tunnels

For an unlabeled IPv6 BGP route to be resolved by an SR policy:

- A color-extended community must be attached to the IPv6 route.
- The base instance BGP next-hop-resolution configuration of **shortcut-tunnel**>**family ipv6** must allow SR policy tunnels.

As an example, if under these conditions, there is an IPv6 route with a color-extended community (value C) and BGP next-hop address N. The order of resolution is as follows:

1. If there is an SR policy in TTMv6 for which endpoint = the BGP next-hop address and color = Cn, then use this tunnel to resolve the BGP next hop.
2. If no SR policy is found in the previous step and the Cn color-extended community has its CO bits set to '01' or '10', then try to find a SR policy in TTMv6 for which endpoint = null (0::0) and color = Cn. If there is such a policy, use it to resolve the BGP next hop.
3. If no SR policy is found in the previous steps and the Cn color-extended community has its CO bits set to '01' or '10' and there is an SR policy in TTMv4 for which endpoint = null (0.0.0.0) and color = Cn, then use this tunnel to resolve the BGP next hop.
4. If no SR policy is found in the previous steps but there is a non-SR policy tunnel in TTMv6 that is allowed by the resolution options and for which endpoint = BGP next-hop address (and for which the admin-tag meets the admin-tag-policy requirements applied to the BGP route, if applicable), then use this tunnel to resolve the BGP next hop if it has the highest TTM preference.
5. Otherwise, fall back to IGP, unless the **disallow-igp** option is configured.

Note:

- Contrary to section 8.8.2 of draft-filsfils-segment-routing-05, BGP only resolves a route with multiple color-extended communities to an SR policy using the color-extended community with the highest value.
- For AFI2/SAFI1 routes, an IPv6 explicit null label should be always be pushed at the bottom of the stack if the policy endpoint is IPv4.

6.4.3 Resolving Label-IPv4 BGP Routes to Segment Routing Policy Tunnels

For a label-unicast IPv4 BGP route to be resolved by an SR policy:

- A color-extended community must be attached to the label-IPv4 route.
- The base instance BGP next-hop-resolution configuration of **labeled-routes**>**transport-tunnel**>**family label-ipv4** must allow SR policy tunnels.

For example, if under these conditions, there is a label-IPv4 route with a color-extended community (value C) and BGP next-hop address N. The order of resolution is as follows:

1. If there is an interface route that can resolve the BGP next hop, then use the direct route.

2. If **allow-static** is configured and there is a static route that can resolve the BGP next hop, then use the static route.
3. If there is no interface route or static route available or allowed to resolve the BGP next hop and the next hop is IPv4 then:
    - Look for an SR policy in TTMv4 for which end-point = BGP next-hop address and color = Cn. If there is such an SR policy then try to use it to resolve the BGP next hop. If the selected SR policy has any segment-list with more than {11- **max-sr-frr-labels** under the IGPs} labels or segments, then the label-IPv4 route is unresolved.
    - If no SR policy is found in the previous steps and the Cn color-extended community has its CO bits set to '01' or '10' then try to find an SR policy in TTMv4 for which endpoint = null (0.0.0.0) and color = Cn. If there is such a policy, use it to resolve the BGP next hop. If the selected SR policy has any segment-list with more than {11- **max-sr-frr-labels** under the IGPs} labels or segments, then the label-IPv4 route will be unresolved.
    - If no SR policy is found in the previous steps and the Cn color-extended community has its CO bits set to '01' or '10' then try to find an SR policy in TTMv6 for which endpoint = null (0::0) and color = Cn. If there is such a policy, use it to resolve the BGP next hop. If the selected SR policy has any segment-list with more than {11- **max-sr-frr-labels** under the IGPs} labels or segments, then the label-IPv4 route is unresolved.
4. If there is no interface route or static route that is available or allowed to resolve the BGP next hop and the next hop is IPv6 then:
    - Look for an SR policy in TTMv6 for which end-point = BGP next-hop address and color = Cn. If there is such an SR policy then try to use it to resolve the BGP next hop. If the selected SR policy has any segment-list with more than {11- **max-sr-frr-labels** under the IGPs} labels or segments, then the label-IPv4 route is unresolved.
    - If no SR policy is found in the previous steps and the Cn color-extended community has its CO bits set to '01' or '10' then try to find an SR policy in TTMv6 for which endpoint = null (0::0) and color = Cn. If there is such a policy, use it to resolve the BGP next hop. If the selected SR policy has any segment-list with more than {11- **max-sr-frr-labels** under the IGPs} labels or segments, then the label-IPv4 route is unresolved.
    - If no SR policy is found in the previous steps and the Cn color-extended community has its CO bits set to '01' or '10' then try to find an SR policy in TTMv4 for which endpoint = null (0.0.0.0) and color = Cn. If there is such a policy, use it to resolve the BGP next hop. If the selected SR policy has any segment-list with more than {11- **max-sr-frr-labels** under the IGPs} labels or segments, then the label-IPv4 route is unresolved.

5. If no SR policy is found in the previous steps but there is a non-SR policy tunnel in TTMv4 (next hop is IPv4) or TTMv6 (next hop is IPv6) that is allowed by the resolution options and for which endpoint = BGP next-hop address (and for which the admin-tag meets the admin-tag-policy requirements applied to the BGP route, if applicable), then use this tunnel to resolve the BGP next hop if it has the highest TTM preference.

➡ **Note:** Contrary to section 8.8.2 of draft-filsfils-segment-routing-05, BGP only resolves a route with multiple color-extended communities to an SR policy using the color-extended community with the highest value.

6.4.4 Resolving Label-IPv6 BGP Routes to Segment Routing Policy Tunnels

For a label-unicast IPv6 BGP route to be resolved by an SR policy:

- A color-extended community must be attached to the label-IPv6 route.
- The base instance BGP next-hop-resolution configuration of **labeled-routes**>**transport-tunnel**>**family label-ipv6** must allow SR policy tunnels.

For example, if under these conditions, there is a label-IPv6 route with a color-extended community (value C) and BGP next-hop address N. The order of resolution is as follows:

1. If there is an interface route that can resolve the BGP next hop, then use the direct route.
2. If **allow-static** is configured and there is a static route that can resolve the BGP next hop, then use the static route.
3. If there is no interface route or static route available or allowed to resolve the BGP next hop and the next hop is IPv6 then:
   - Look for an SR policy in TTMv6 for which end-point = BGP next-hop address and color = Cn. If there is such an SR policy then try to use it to resolve the BGP next hop.
   - If no SR policy is found in the previous steps and the Cn color-extended community has its CO bits set to '01' or '10' then try to find an SR policy in TTMv6 for which endpoint = null (0::0) and color = Cn. If there is such a policy, use it to resolve the BGP next hop.
   - If no SR policy is found in the previous steps and the Cn color-extended community has its CO bits set to '01' or '10' then try to find an SR policy in TTMv4 for which endpoint = null (0.0.0.0) and color = Cn. If there is such a policy, use it to resolve the BGP next hop.

4. If there is no interface route or static route that is available or allowed to resolve the BGP next hop and the next hop is IPv4-mapped-IPv6 then:
   - Look for an SR policy in TTMv4 for which end-point = BGP next-hop address and color = Cn. If there is such an SR policy then try to use it to resolve the BGP next hop.
   - If no SR policy is found in the previous steps and the Cn color-extended community has its CO bits set to '01' or '10' then try to find an SR policy in TTMv4 for which endpoint = null (0.0.0.0) and color = Cn. If there is such a policy, use it to resolve the BGP next hop.
   - If no SR policy is found in the previous steps and the Cn color-extended community has its CO bits set to '01' or '10' then try to find an SR policy in TTMv6 for which endpoint = null (0::0) and color = Cn. If there is such a policy, use it to resolve the BGP next hop.
5. If no SR policy is found in the previous steps but there is a non-SR-policy tunnel in TTMv6 (next hop is IPv6) or in TTMv4 (next hop is IPv4-mapped-IPv6) that is allowed by the resolution options and for which endpoint = BGP next-hop address (and for which the admin-tag meets the admin-tag-policy requirements applied to the BGP route, if applicable) then use this tunnel to resolve the BGP next hop if it has the highest TTM preference.

Note: Contrary to section 8.8.2 of draft-filsfils-segment-routing-05, BGP only resolves a route with multiple color-extended communities to an SR policy using the color-extended community with the highest value.

6.4.5 Resolving EVPN-MPLS Routes to Segment Routing Policy Tunnels

The next-hop resolution for all EVPN-VXLAN routes and for EVPN-MPLS routes without a color-extended community is unchanged by this feature.

When the resolution options associated with the **auto-bind-tunnel** configuration of an EVPN-MPLS service (vpls, b-vpls, r-vpls or E-pipe) allow **sr-policy** tunnels from TTM, then the next-hop resolution of EVPN-MPLS routes (RT-1 per-EVI, RT-2, RT-3 and RT-5) with one or more color-extended communities C1, C2, .. Cn (Cn = highest value) is based on the following rules.

1. If the next hop is IPv6 and there is an SR policy in TTMv6 for which end-point = BGP next-hop address and color = Cn, then use this tunnel to resolve the BGP next hop.

2. Otherwise, if the next hop is IPv4 or IPv4-mapped-IPv6 and there is an SR policy in TTMv4 for which end-point = BGP next-hop address (or the IPv4 address extracted from the IPv4-mapped IPv6 BGP next-hop address) and color = Cn, then use this tunnel to resolve the BGP next hop.
3. If no SR policy is found in the previous steps but there is a non-SR policy tunnel in TTMv4 (next hop is IPv4 or IPv4-mapped-IPv6) or TTMv6 (next hop is IPv6) that is allowed by the resolution options and for which endpoint = BGP next-hop address, then use this tunnel to resolve the BGP next hop if it has the highest TTM preference.

Note: Contrary to section 8.8.2 of draft-filsfils-segment-routing-05, BGP only resolves a route with multiple color-extended communities to an SR policy using the color-extended community with the highest value.

6.4.6 VPRN Auto-Bind-Tunnel Using Segment Routing Policy Tunnels

When the resolution options associated with the **auto-bind-tunnel** configuration of VPRN service allow **sr-policy** tunnels from TTM, next-hop resolution of VPN-IPv4 and VPN-IPv6 routes that are imported into the VPRN and have one or more color-extended communities C1, C2, .. Cn (Cn = highest value) is based on the following rules.

1. If the next hop is IPv6 and there is an SR policy in TTMv6 for which end-point = BGP next-hop address and color = Cn, then use this tunnel to resolve the BGP next hop.
2. Otherwise, if the next hop is IPv4 or IPv4-mapped-IPv6 and there is an SR policy in TTMv4 for which end-point = BGP next-hop address (or the IPv4 address extracted from the IPv4-mapped IPv6 BGP next-hop address in the case of VPN-IPv6 routes) and color = Cn, then use this tunnel to resolve the BGP next hop.
3. If no SR policy is found in the previous step but there is a non-SR policy tunnel in TTMv4 (next hop is IPv4 or IPv4-mapped-IPv6) or TTMv6 (next hop is IPv6) that is allowed by the resolution options and for which endpoint = BGP next-hop address, then use this tunnel to resolve the BGP next hop if it has the highest TTM preference.

Note: Contrary to section 8.8.2 of draft-filsfils-segment-routing-05, BGP only resolves a route with multiple color-extended communities to an SR policy using the color-extended community with the highest value.

6.5 Seamless BFD and End-to-End Protection for SR Policies

6.5.1 Introduction

This feature reuses of the capabilities of SR-TE LSPs to SR policy, so that operators wishing to use SR policies to enable more flexible and dynamic policy-based routing can benefit from network-based data path monitoring and fast protection switching in case a path failures.

Seamless BFD (S-BFD) is a form of BFD that requires significantly less state and reduces the need for session bootstrapping as compared to LSP BFD. Refer to *Seamless Bidirectional Forwarding Detection (S-BFD)* in the *7450 ESS, 7750 SR, 7950 XRS, and VSR OAM and Diagnostics Guide*. S-BFD requires centralized configuration of a reflector function, as well as a mapping at the head-end node between the remote session discriminator and the IP address for the reflector by each session. This configuration and the mapping are described in the *7450 ESS, 7750 SR, 7950 XRS, and VSR OAM and Diagnostics Guide*.

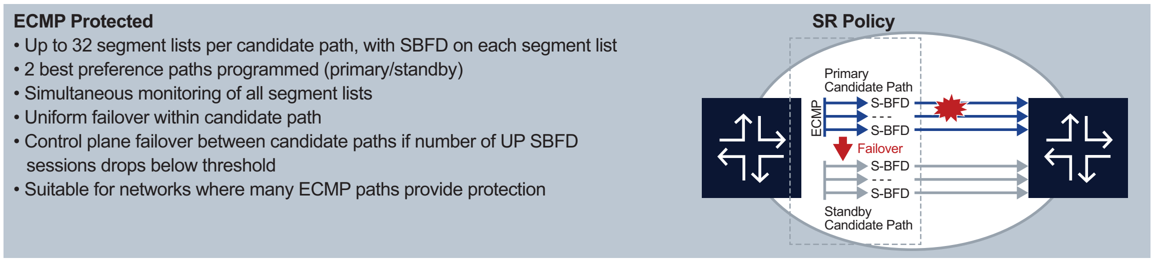
This section describes the application of S-BFD to SR-Polices and the configuration required for this feature. See Seamless BFD for SR-TE LSPs for details of the application of S-BFD to SR-TE LSPs.

S-BFD provides a connectivity check for the data path of a segment list in an SR policy, and can determine whether the segment list is up. In addition, the router also supports two protection modes for an SR policy that are distinguished by the data path programming characteristics and whether uniform failover is required between segment lists in the same SR policy candidate path (ECMP protected mode), or between the programmed candidate paths (linear mode). These protection modes are driven by the S-BFD session state on the programmed segment lists of an SR policy.

6.5.1.1 ECMP Protected Mode

ECMP protected mode programs all segment lists of the top-two candidate paths of an SR policy in the IOM. ECMP protected mode allows establishment of S-BFD on all of those segment lists. All of the segment lists of a specified candidate path are in the same protection group, but different candidate paths are not in the same protection group. Switchover between candidate paths is triggered by the control plane. A segment list is only included in the ECMP set of segment lists if its S-BFD session is up (user traffic is forwarded on a segment list whose S-BFD session is down). See Figure 67.

Figure 67 **ECMP Protected SR Policy with S-BFD**



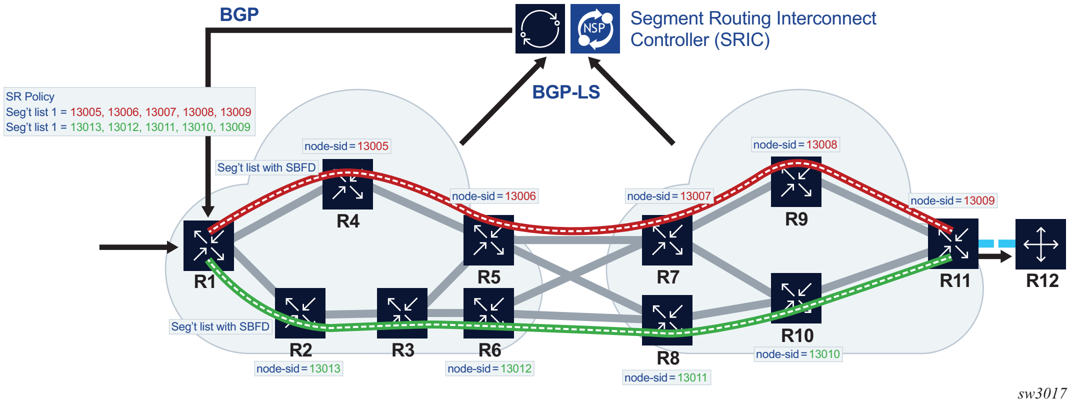
ECMP Protected

- Up to 32 segment lists per candidate path, with SBFD on each segment list
- 2 best preference paths programmed (primary/standby)
- Simultaneous monitoring of all segment lists
- Uniform failover within candidate path
- Control plane failover between candidate paths if number of UP SBFD sessions drops below threshold
- Suitable for networks where many ECMP paths provide protection

sw3015

Figure 68 depicts an application for S-BFD on SR policies with ECMP protected mode. Here, an SR policy is programmed at R1 by the Nokia NSP with two segment lists from R1 to R11. One segment list is using R4/R5/R7R9, and the other segment list is using R2/R3/R6/R8 and R10. These segment lists are using diverse paths and traffic that is sprayed across both of them according to the configured hashing algorithm. Separate S-BFD sessions are run on each segment list and allow the rapid detection of data path failures along the whole segment list path. R1 is able to rapidly remove a segment list from the ECMP set if S-BFD goes down, and is also able to failover to a backup SR policy (not shown) (or fall back to a less preferred LSP) if more than a certain number of the S-BFD sessions go down.

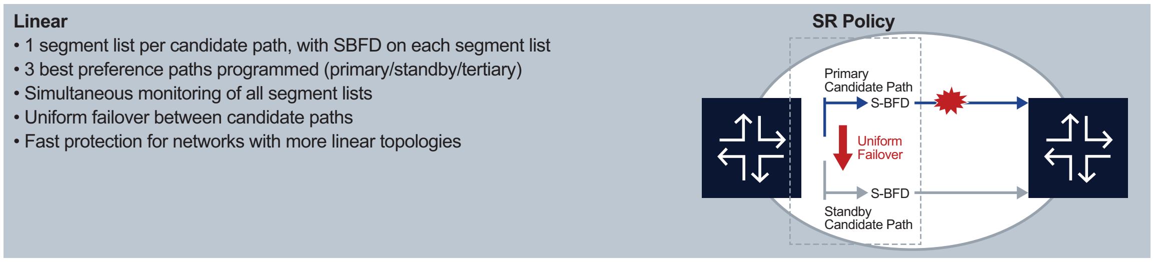
Figure 68 Example Application of ECMP Protected Mode with S-BFD



6.5.1.2 Linear Mode

This mode is termed linear because it is similar in operation to traditional 1-for-1 linear protection switching. It is intended to allow one or more backup paths to protect a primary path, with fast failover between candidate paths. Uniform failover is supported between candidate paths of the same SR policy. Only one segment list from each of the top-three preference candidate paths is programmed in the IOM. All of the programmed candidate paths of a specified SR policy are in the same protection group. See Figure 69.

Figure 69 Linear Protected SR Policy with S-BFD



Linear

- 1 segment list per candidate path, with SBFD on each segment list
- 3 best preference paths programmed (primary/standby/tertiary)
- Simultaneous monitoring of all segment lists
- Uniform failover between candidate paths
- Fast protection for networks with more linear topologies

6.5.2 Detailed Description

This section describes the S-BFD for SR policies, support for primary and backup candidate paths, and configuration steps for S-BFD and protection for SR policies.

6.5.2.1 S-BFD for SR Policies

S-BFD is supported on segment lists for both static SR policies and BGP SR policies by binding a maintenance policy containing an S-BFD configuration to an imported SR policy route or a static SR policy. S-BFD packets are encapsulated on the SR policy segment lists in the same way as for SR-TE LSP paths. As in the case of SR-TE LSPs, the discriminator of the local node as well as a mapping to the far-end reflector node discriminators is first required. BFD sets the remote discriminator at the initiator of the S-BFD session based on a lookup in the S-BFD reflector discriminator using the endpoint address of the SR policy candidate path. A candidate path of an SR policy is only treated as available if the number of up S-BFD sessions equals or exceeds a configurable threshold.

Note: When an SR policy candidate path is first programmed, a 3 second initialization hold timer is triggered. This allows the establishment of all the S-BFD sessions for all programmed paths before it decides which candidate path to activate among the eligible ones (eligible means number of segment lists with S-BFD sessions in up-state that is higher or equal to a configured threshold).

Since this is set to 3 seconds, it is recommended that the transmit and receive control packet timers are set to no more than 1 second with a maximum multiplier of 3 for S-BFD sessions.

S-BFD control packet timers, that are configurable down to 10ms, are supported for specific SR OS platforms with CPM network processor support.

The router supports an uncontrolled return path for S-BFD packets on SR policies. By default, the BFD reply packet from the reflector node is routed out-of-band to the head end of the SR policy.

6.5.2.2 Support for Primary and Backup Candidate Paths

End-to-end protection of static and BGP SR policies is supported using ECMP-protected or linear mode.

If an SR policy for a specified {headend, color, endpoint} is imported (by BGP) or configured (in the static case) and is selected for use, then the best (highest) preference candidate path is treated as the primary path while the next preference candidate preference policy is treated as the standby path. In linear mode, if a third path is present, then this is treated as a tertiary standby path. All of the valid segment lists for these are programmed in the IOM and made available for forwarding S-BFD packets, subject to a limitation in linear mode of one segment list per candidate path. In ECMP protected mode, the two best preference candidate paths are programmed in the IOM (up to 32 segment lists per path), while in linear mode, the three best preference candidate paths are programmed in the IOM (one segment list per candidate path).

In each case, segment lists of the best preference path are initially programmed as forwarding NHLEs while the others are programmed as non-forwarding. If the maximum number of programmed paths for a specified mode has been reached (for example, two for ECMP protected mode, and three for linear mode), and a consistent new path is received with a better preference than the existing active path, then this new path is only considered if or when the route for one of the current programmed paths is withdrawn or deleted. However, if the maximum number of programmed paths for the mode has not been reached, then the new path is programmed and any configured revert timer is started. The system switches to that better preference path immediately or when the revert timer expires (as applicable).

Failover is supported between the currently active path and the next best preference path if the currently active path is down due to S-BFD. Similar to the case of SR-TE LSPs, by default, if ECMP protected or linear mode is configured, the system switches back to the primary (best preference) SR policy path as soon as it recovers. This can happen when the number of up S-BFD sessions equals or exceeds a threshold and a hold-down timer has expired. However, it is possible to configure a revert timer to control reversion to the primary path.

All candidate paths of an SR policy must have the same binding SID when one of these two modes is applied.

6.5.2.3 Configuration of S-BFD and Protection for SR Policies

S-BFD and protection for SR Policies is configured using the following steps.

Step 1. Configure an S-BFD reflector and mapping parameters for the remote reflector under the **configure**>**router**>**bfd**>**seamless-bfd** context.

Step 2. Configure one or more BFD templates defining the BFD session parameters under the **configure**>**router**>**bfd** context.

Step 3. Configure protection and BFD parameters that are applied to SR policies in a named maintenance policy under the **configure**>**router**>**segment-routing** context.

Step 4. For static SR policies, apply a named maintenance policy to the static SR policy under the **configure**>**router**>**segment-routing**>**sr-policies**>**static-policy** context.

Step 5. For dynamic BGP SR policies, configure a policy statement entry to match on a specific route or a set of routes of type **sr-policy-ipv4** with an **action** of **accept** and applying a named SR maintenance policy to them.

Refer to the *7450 ESS, 7750 SR, 7950 XRS, and VSR OAM and Diagnostics Guide* for detailed information on Steps 1 and 2. See Configuration of SR Policy S-BFD and Mode Parameters, Application of S-BFD and Protection Parameters to Static SR-Policies, and Application of S-BFD and Protection Parameters to BGP SR-Policies for details on Steps 3 through 5.

6.5.2.3.1 Configuration of SR Policy S-BFD and Mode Parameters

S-BFD and protection mode parameters are configured in a named maintenance policy. This is applied to SR policy paths that are imported by BGP as a policy statement action or by binding to a static SR policy configuration.

Maintenance policies are configured as follows:

```
configure>router>segment-routing>
   maintenance-policy <name>
      bfd-enable
      bfd-template <name>
      mode {linear | ecmp-protected}
      threshold <number>
      revert-timer <timer-value>
      hold-down-timer <timer-value>
      no shutdown
```

The **bfd-enable** command enables or disables BFD on all of the segment lists of the candidate path that are installed in the data path.

The **bfd-template** command refers to a named BFD template that must exist on the system.

The **mode** command specifies how to program the data path and how to behave if the number of BFD sessions that are up is less than the threshold and the hold-down timer has expired. All of the paths in the set must have the same mode (see SR Policy Route and Candidate Path Parameter Consistency). All of the allowed segment lists of the SR policy path are programmed in the IOM. The default mode is none.

In both the **linear** mode and **ecmp-protected** modes, if two or more SR policy paths with the same {headend, color, endpoint} have the same mode, then the highest preference path is treated as an effective primary path while the next highest path preference is treated as the standby path. If a third path is present in the **linear** mode, then this is treated as a tertiary path and also programmed in the IOM.

In the **ecmp-protected** mode, all the segment lists of the top two best preference paths are programmed it the IOM. However, in **linear** mode, the lowest index segment list of each of the top three preference paths is programmed in the IOM and linear protection is supported between that set. All of the segment lists of the programmed paths are made available for forwarding S-BFD packets.

If the currently active path becomes unavailable due to S-BFD, the system fails over to the next best preference candidate path that is available. If all programmed candidate paths are unavailable, then the SR policy is marked as down in TTM.

The **linear** mode supports uniform failover between candidate paths (policy routes) of the same SR policy. If **linear** mode is configured then the following rules apply.

- Only one segment list is allowed per SR policy path. If more than one is configured, then only the lowest index segment list is programmed in the data path.
- The top-3 best preference valid SR policy paths belonging to the same SR policy are programmed in the IOM and are assigned to the same protection group. Uniform failover is supported between these paths.

The **threshold** command configures the minimum number of S-BFD sessions that must be up to consider the SR policy candidate path to be up. If it is below this number, the SR policy candidate path is marked as BFD degraded by the system. The threshold parameter is only valid in the **ecmp-protected** mode (a threshold of 1 is implicit in the **linear** mode).

If the **revert-timer** command is configured, then the router starts a revert timer when the primary path recovers (for example, after the number of S-BFD sessions that are up is ≥ threshold and the hold-down timer has expired) and switches back when the timer expires. If **no revert-timer** is configured, then the system reverts to the primary path for the policy when it is restored.

If a secondary or tertiary path is currently active, and the revert timer is started (due to recovery of the primary path), but the secondary path subsequently goes down due to the number of up S-BFD sessions being less than the threshold, and no other better preference standby path is available, then the router reverts immediately to the primary path. However, if a better preference standby path is available and up, the revert timer is not canceled and the system switches to the better preference standby path and switches back to the primary path when the revert timer expires. If the hold-down timer is currently active on a better-preference path, then the system immediately switches to the primary path. If the system needs to switch to the primary path but the hold-down timer is still active on the primary path, the system cancels the timer and switches immediately.

The **hold-down-timer** command is intended to prevent bouncing of the SR policy path state if one or more BFD sessions associated with segment lists flap and cause the threshold to be repeatedly crossed in a short period of time. The hold-down timer is started when the number of S-BFD sessions that are up drops below the threshold. The SR policy path is not considered to be up again until the hold-down timer has expired and the number of S-BFD sessions that are up equals or exceeds the threshold.

A maintenance policy can only be deleted or a value changed if the maintenance policy is administratively disabled (shutdown). A maintenance policy can only be enabled if the **bfd-enable**, **bfd-template** and **mode** commands are configured. All associated SR policy paths are deleted from the IOM if a maintenance template is shutdown.

6.5.2.3.2 Application of S-BFD and Protection Parameters to Static SR-Policies

A named maintenance policy is applied to a static SR policy using the **maintenance-policy** command as follows:

```
config router segment-routing sr-policies
            static-policy <name>
                head-end local
                binding-sid <number>
                maintenance-policy <name>
                ...
```

A maintenance policy can only be configured if the static SR policy **head-end** is set to **local**. Policies with an IP address that is not local to the node are not programmed in the SR database and cannot have S-BFD sessions established on them by this node because they are not the head end for the SR policy path.

S-BFD needs an endpoint address for the session so that the S-BFD reflector discriminator can be looked-up as a part of the session addressing. A maintenance policy cannot be configured on an SR policy with a null endpoint.

6.5.2.3.3 Application of S-BFD and Protection Parameters to BGP SR-Policies

S-BFD and protection parameters can be applied to matching imported SR policy routes. Match criteria in the route import policy for the color, endpoint and route distinguisher of a policy enable matching on a specific SR policy route for **family sr-policy-v4** and **sr-policy-v6** types.

➡ **Note:** For routes with the same matching distinguisher, only those with the best criteria are pushed to the SR database.

For example, matching a unique SR policy requires the following fully qualified set of match criteria:

```
configure router policy-options
   policy-statement <name>
      entry <id>
         from family sr-policy-ipv4
         from distinguisher <rd-value>
         from color <color>
         from endpoint <ip-address>
```

However, users may only require more general match criteria (for example, to apply the same maintenance template to all imported SR policy IPv4 routes, irrespective of color or endpoint).

An SR policy maintenance template is applied to matching SR policy routes using the **sr-maintenance-policy action** commands.

```
configure policy-options
   policy-statement <name>
      entry <id>
         from family sr-policy-ipv4
         ...
         action accept
           sr-maintenance-policy <name>
```

Maintenance policy statements are applicable as actions on a specific entry or as the default action.

The named SR maintenance policy must exist on the system when the commit is executed for the routing policy. If parameterization of actions is used and the named SR maintenance policy exists, the router still validates.

A change in policy options action deletes all programmed paths for that route and based on the new action, re-downloads applicable routes to the IOM.

6.5.2.3.4 SR Policy Route and Candidate Path Parameter Consistency

An SR policy consists of a set of one or more candidate paths. Each candidate path may be described by an SR policy route, that may be a static SR policy that is configured under the **config**>**router**>**segment-routing**>**sr-policies** context, or a dynamic route imported by BGP. The router checks the consistency of the following BFD and protection parameters across all of the SR policy routes for a specified SR policy.

{Maintenance-policy existence}

```
bfd-enable
bfd-template <name>
mode {linear | ecmp-protected}
revert-timer <timer-value>
```

Maintenance-policy existence covers the case where the existing programmed route is an SR policy with no maintenance policy, and the new route has a maintenance policy, and vice-versa.

Consistency is enforced across all of the static SR policy candidate paths and dynamic SR policy routes that make up a segment routing policy. Since SR policy routes or paths are imported sequentially and cannot be considered together, inconsistencies are handled as follows:

```
First policy route imported/configured:
Check: valid set of parameters
Action: If OK, program in data path and activate

Second policy route imported/configured:
Check: valid set of parameters, consistency with existing activated policy route
Action If OK, program in data path and activate, else hold in CPM but do not program

Third policy route imported/configured:
Check: valid set of parameters, consistency with existing activated policy route (s)
Action If OK, program in data path and activate, else hold in CPM but do not program
```

Inconsistent policy routes (paths) are only programmed if their parameters are valid and any programmed routes for that SR policy are deleted.

By using the same maintenance policy for all of the SR policy's routes, inconsistencies between the BFD and protection parameters of SR policy routes belonging to a specified SR policy can be avoided.

6.6 Traffic Statistics

SR policies provide the ability to collect statistics for ingress and egress traffic. In both cases, traffic statistics are collected without any forwarding class or QoS distinction.

Traffic statistics collection is enabled as follows:

- **config**>**router**>**segment-routing**>**sr-policies**>**ingress-statistics**

  Ingress — Ingress traffic collection only applies to **binding-sid** SR policies as the statistic index is attached to the ILM entry for that label. The traffic statistics provide traffic for all the instances that share the binding SID. The statistic index is released and statistics are lost when ingress traffic statistics are disabled for that binding SID, or the last instance of a policy using that label is removed from the database.
- **config**>**router**>**segment-routing**>**sr-policies**>**egress-statistics**

  Egress — Egress traffic statistics are collected globally, for all policies at the same time. Both static and signaled policies are subject to traffic statistics collection. Statistic indexes are allocated per segment list, which allows for a fine grain monitoring of traffic evolution. Also, statistic indexes are only allocated at the time the segment list is effectively programmed. However, the system allocates at most 32 statistic indexes across all the instances of a given policy. Therefore, in the case where an instance of a policy is deprogrammed and a more preferred instance is programmed, the system behaves as follows:
  - If the segment list IDs of the preferred instance are different from any of the segment list IDs of any previously programmed instance, the system allocates new statistic indexes. While that condition holds, the statistics associated with a segment list of an instance strictly reflect the traffic that used that segment list in that instance.
  - If some of the segment list IDs of the preferred instance are equal to any of the segment list IDs of any previously programmed instance, the system reuses the indexes of the preferred instance and keeps the associated counter value and increment. In this case, the traffic statistics provided per segment list not only reflect the traffic that used that segment list in that instance. It incorporates counter values of at least another segment-list in another instance of that policy.

In all cases, the aggregate values provided across all instances truly reflect traffic over the various instances of the policy.

Statistic indexes are not released at deprogramming time. They are, however, released when all the instances of a policy are removed from the database, or when the **egress-statistics** command is disabled.

7 Label Distribution Protocol

7.1 Label Distribution Protocol

Label Distribution Protocol (LDP) is a protocol used to distribute labels in non-traffic-engineered applications. LDP allows routers to establish label switched paths (LSPs) through a network by mapping network-layer routing information directly to data link layer-switched paths.

An LSP is defined by the set of labels from the ingress Label Switching Router (LSR) to the egress LSR. LDP associates a Forwarding Equivalence Class (FEC) with each LSP it creates. A FEC is a collection of common actions associated with a class of packets. When an LSR assigns a label to a FEC, it must let other LSRs in the path know about the label. LDP helps to establish the LSP by providing a set of procedures that LSRs can use to distribute labels.

The FEC associated with an LSP specifies which packets are mapped to that LSP. LSPs are extended through a network as each LSR splices incoming labels for a FEC to the outgoing label assigned to the next hop for the given FEC. The next-hop for a FEC prefix is resolved in the routing table. LDP can only resolve FECs for IGP and static prefixes. LDP does not support resolving FECs of a BGP prefix.

LDP allows an LSR to request a label from a downstream LSR so it can bind the label to a specific FEC. The downstream LSR responds to the request from the upstream LSR by sending the requested label.

LSRs can distribute a FEC label binding in response to an explicit request from another LSR. This is known as Downstream On Demand (DOD) label distribution. LSRs can also distribute label bindings to LSRs that have not explicitly requested them. This is called Downstream Unsolicited (DU).

7.1.1 LDP and MPLS

LDP performs the label distribution only in MPLS environments. The LDP operation begins with a hello discovery process to find LDP peers in the network. LDP peers are two LSRs that use LDP to exchange label/FEC mapping information. An LDP session is created between LDP peers. A single LDP session allows each peer to learn the other's label mappings (LDP is bi-directional) and to exchange label binding information.

LDP signaling works with the MPLS label manager to manage the relationships between labels and the corresponding FEC. For service-based FECs, LDP works in tandem with the Service Manager to identify the virtual leased lines (VLLs) and Virtual Private LAN Services (VPLSs) to signal.

An MPLS label identifies a set of actions that the forwarding plane performs on an incoming packet before discarding it. The FEC is identified through the signaling protocol (in this case, LDP) and allocated a label. The mapping between the label and the FEC is communicated to the forwarding plane. In order for this processing on the packet to occur at high speeds, optimized tables are maintained in the forwarding plane that enable fast access and packet identification.

When an unlabeled packet ingresses the router, classification policies associate it with a FEC. The appropriate label is imposed on the packet, and the packet is forwarded. Other actions that can take place before a packet is forwarded are imposing additional labels, other encapsulations, learning actions, and so on When all actions associated with the packet are completed, the packet is forwarded.

When a labeled packet ingresses the router, the label or stack of labels indicates the set of actions associated with the FEC for that label or label stack. The actions are performed on the packet and then the packet is forwarded.

The LDP implementation provides DOD, DU, ordered control, liberal label retention mode support.

7.1.2 LDP Architecture

LDP comprises a few processes that handle the protocol PDU transmission, timer-related issues, and protocol state machine. The number of processes is kept to a minimum to simplify the architecture and to allow for scalability. Scheduling within each process prevents starvation of any particular LDP session, while buffering alleviates TCP-related congestion issues.

The LDP subsystems and their relationships to other subsystems are illustrated in Figure 70. This illustration shows the interaction of the LDP subsystem with other subsystems, including memory management, label management, service management, SNMP, interface management, and RTM.   In addition, debugging capabilities are provided through the logger.

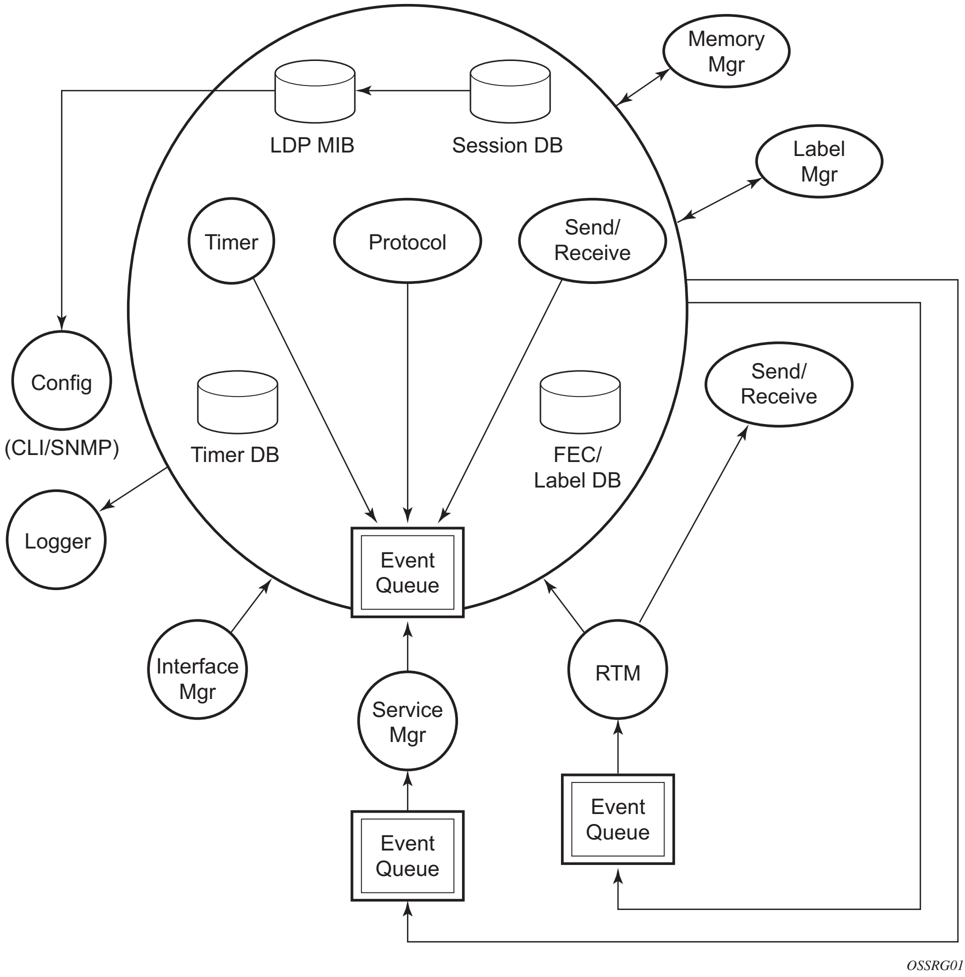
Communication within LDP tasks is typically done by inter-process communication through the event queue, as well as through updates to the various data structures. The primary data structures that LDP maintains are:

- FEC/label database — Contains all FEC to label mappings that include both sent and received. It also contains both address FECs (prefixes and host addresses) and service FECs (L2 VLLs and VPLS)
- Timer database — Contains all timers for maintaining sessions and adjacencies
- Session database — Contains all session and adjacency records, and serves as a repository for the LDP MIB objects

7.1.3 Subsystem Interrelationships

The sections below describe how LDP and the other subsystems work to provide services. Figure 70 shows the interrelationships among the subsystems.

Figure 70 **Subsystem Interrelationships**



7.1.3.1 Memory Manager and LDP

LDP does not use any memory until it is instantiated. It pre-allocates some amount of fixed memory so that initial startup actions can be performed. Memory allocation for LDP comes out of a pool reserved for LDP that can grow dynamically as needed. Fragmentation is minimized by allocating memory in larger chunks and managing the memory internally to LDP. When LDP is shut down, it releases all memory allocated to it.

7.1.3.2 Label Manager

LDP assumes that the label manager is up and running. LDP will abort initialization if the label manager is not running. The label manager is initialized at system boot up; hence, anything that causes it to fail will likely imply that the system is not functional. The router uses the dynamic label range to allocate all dynamic labels, including RSVP and BGP allocated labels and VC labels.

7.1.3.3 LDP Configuration

The router uses a single consistent interface to configure all protocols and services. CLI commands are translated to SNMP requests and are handled through an agent-LDP interface. LDP can be instantiated or deleted through SNMP. Also, LDP targeted sessions can be set up to specific endpoints. Targeted-session parameters are configurable.

7.1.3.4 Logger

LDP uses the logger interface to generate debug information relating to session setup and teardown, LDP events, label exchanges, and packet dumps. Per-session tracing can be performed.

7.1.3.5 Service Manager

All interaction occurs between LDP and the service manager, since LDP is used primarily to exchange labels for Layer 2 services. In this context, the service manager informs LDP when an LDP session is to be set up or torn down, and when labels are to be exchanged or withdrawn. In turn, LDP informs service manager of relevant LDP events, such as connection setups and failures, timeouts, labels signaled/withdrawn.

7.1.4 Execution Flow

LDP activity in the operating system is limited to service-related signaling. Therefore, the configurable parameters are restricted to system-wide parameters, such as hello and keepalive timeouts.

7.1.4.1 Initialization

LDP makes sure that the various prerequisites, such as ensuring the system IP interface is operational, the label manager is operational, and there is memory available, are met. It then allocates itself a pool of memory and initializes its databases.

7.1.4.2 Session Lifetime

In order for a targeted LDP (T-LDP) session to be established, an adjacency must be created. The LDP extended discovery mechanism requires hello messages to be exchanged between two peers for session establishment. After the adjacency establishment, session setup is attempted.

7.1.4.2.1 Adjacency Establishment

In the router, the adjacency management is done through the establishment of a Service Distribution Path (SDP) object, which is a service entity in the Nokia service model.

The Nokia service model uses logical entities that interact to provide a service. The service model requires the service provider to create configurations for four main entities:

- Customers
- Services
- Service Access Paths (SAPs) on the local routers
- Service Distribution Points (SDPs) that connect to one or more remote routers.

An SDP is the network-side termination point for a tunnel to a remote router. An SDP defines a local entity that includes the system IP address of the remote routers and a path type. Each SDP comprises:

- The SDP ID
- The transport encapsulation type, either MPLS or GRE
- The far-end system IP address

If the SDP is identified as using LDP signaling, then an LDP extended hello adjacency is attempted.

If another SDP is created to the same remote destination, and if LDP signaling is enabled, no further action is taken, since only one adjacency and one LDP session exists between the pair of nodes.

An SDP is a uni-directional object, so a pair of SDPs pointing at each other must be configured in order for an LDP adjacency to be established. Once an adjacency is established, it is maintained through periodic hello messages.

7.1.4.2.2 Session Establishment

When the LDP adjacency is established, the session setup follows as per the LDP specification. Initialization and keepalive messages complete the session setup, followed by address messages to exchange all interface IP addresses. Periodic keepalives or other session messages maintain the session liveliness.

Since TCP is back-pressured by the receiver, it is necessary to be able to push that back-pressure all the way into the protocol. Packets that cannot be sent are buffered on the session object and re-attempted as the back-pressure eases.

7.1.5 Label Exchange

Label exchange is initiated by the service manager. When an SDP is attached to a service (for example, the service gets a transport tunnel), a message is sent from the service manager to LDP. This causes a label mapping message to be sent. Additionally, when the SDP binding is removed from the service, the VC label is withdrawn. The peer must send a label release to confirm that the label is not in use.

7.1.5.1 Other Reasons for Label Actions

Other reasons for label actions include:

- MTU changes: LDP withdraws the previously assigned label, and re-signals the FEC with the new MTU in the interface parameter.
- Clear labels: When a service manager command is issued to clear the labels, the labels are withdrawn, and new label mappings are issued.
- SDP down: When an SDP goes administratively down, the VC label associated with that SDP for each service is withdrawn.
- Memory allocation failure: If there is no memory to store a received label, it is released.
- VC type unsupported: When an unsupported VC type is received, the received label is released.

7.1.5.2 Cleanup

LDP closes all sockets, frees all memory, and shuts down all its tasks when it is deleted, so its memory usage is 0 when it is not running.

7.1.5.3 Configuring Implicit Null Label

The implicit null label option allows an egress LER to receive MPLS packets from the previous hop without the outer LSP label. The user can configure to signal the implicit operation of the previous hop is referred to as penultimate hop popping (PHP). This option is signaled by the egress LER to the previous hop during the FEC signaling by the LDP control protocol.

Enable the use of the implicit null option, for all LDP FECs for which this node is the egress LER, using the following command:

config>**router**>**ldp**>**implicit-null-label**

When the user changes the implicit null configuration option, LDP withdraws all the FECs and re-advertises them using the new label value.

7.1.6 Global LDP Filters

Both inbound and outbound LDP label binding filtering are supported.

Inbound filtering is performed by way of the configuration of an import policy to control the label bindings an LSR accepts from its peers. Label bindings can be filtered based on:

- Prefix-list: Match on bindings with the specified prefix/prefixes.
- Neighbor: Match on bindings received from the specified peer.

The default import policy is to accept all FECs received from peers.

Outbound filtering is performed by way of the configuration of an export policy. The Global LDP export policy can be used to explicitly originate label bindings for local interfaces. The Global LDP export policy does not filter out or stop propagation of any FEC received from neighbors. Use the LDP peer export prefix policy for this purpose.

By default, the system does not interpret the presence or absence of the system IP in global policies, and as a result always exports a FEC for that system IP. The **consider-system-ip-in-gep** command causes the system to interpret the presence or absence of the system IP in global export policies in the same way as it does for the IP addresses of other interfaces.

Export policy enables configuration of a policy to advertise label bindings based on:

- Direct: All local subnets.
- Prefix-list: Match on bindings with the specified prefix or prefixes.

The default export policy is to originate label bindings for system address only and to propagate all FECs received from other LDP peers.

Finally, the 'neighbor interface' statement inside a global import policy is not considered by LDP.

7.1.6.1 Per LDP Peer FEC Import and Export Policies

The FEC prefix export policy provides a way to control which FEC prefixes received from prefixes received from other LDP and T-LDP peers are re-distributed to this LDP peer.

The user configures the FEC prefix export policy using the following command:

config>router>ldp>session-params>peer>export-prefixes policy-name

By default, all FEC prefixes are exported to this peer.

The FEC prefix import policy provides a mean of controlling which FEC prefixes received from this LDP peer are imported and installed by LDP on this node. If resolved these FEC prefixes are then re-distributed to other LDP and T-LDP peers.

The user configures the FEC prefix export policy using the following command:

config>router>ldp>session-params>peer>import-prefixes policy-name

By default, all FEC prefixes are imported from this peer.

7.1.7 Configuring Multiple LDP LSR ID

The multiple LDP LSR-ID feature provides the ability to configure and initiate multiple Targeted LDP (T-LDP) sessions on the same system using different LDP LSR-IDs. In the current implementation, all T-LDP sessions must have the LSR-ID match the system interface address. This feature continues to allow the use of the system interface by default, but also any other network interface, including a loopback, address on a per T-LDP session basis. The LDP control plane will not allow more than a single T-LDP session with different local LSR ID values to the same LSR-ID in a remote node.

An SDP of type LDP can use a provisioned targeted session with the local LSR-ID set to any network IP for the T-LDP session to the peer matching the SDP far-end address. If, however, no targeted session has been explicitly pre-provisioned to the far-end node under LDP, then the SDP will auto-establish one but will use the system interface address as the local LSR ID.

An SDP of type RSVP must use an RSVP LSP with the destination address matching the remote node LDP LSR-ID. An SDP of type GRE can only use a T-LDP session with a local LSR-ID set to the system interface.

The multiple LDP LSR-ID feature also provides the ability to use the address of the local LDP interface, or any other network IP interface configured on the system, as the LSR-ID to establish link LDP Hello adjacency and LDP session with directly connected LDP peers. The network interface can be a loopback or not.

Link LDP sessions to all peers discovered over a given LDP interface share the same local LSR-ID. However, LDP sessions on different LDP interfaces can use different network interface addresses as their local LSR-ID.

By default, the link and targeted LDP sessions to a peer use the system interface address as the LSR-ID unless explicitly configured using this feature. The system interface must always be configured on the router or else the LDP protocol will not come up on the node. There is no requirement to include it in any routing protocol.

When an interface other than system is used as the LSR-ID, the transport connection (TCP) for the link or targeted LDP session will also use the address of that interface as the transport address.

7.1.7.1 Advertisement of FEC for Local LSR ID

The FEC for a Local LSR ID is not advertised by default by the system, unless it is explicitly configured to do so. The advertisement of the local-lsr-id is configured using the **adv-local-lsr-id** commands in the session parameters for a given peer or the targeted-session peer-template.

7.1.8 Extend LDP policies to mLDP

In addition to link LDP, a policy can be assigned to mLDP as an import policy. For example, if the following policy was assigned as an import policy to mLDP, any FEC arriving with an IP address of 100.0.1.21 will be dropped.

```
*A:SwSim2>config>router>policy-options# info
----------------------------------------------
            prefix-list "100.0.1.21/32"
                prefix 100.0.1.21/32 exact
            exit
            policy-statement "policy1"
                entry 10
                    from
                        prefix-list "100.0.1.21/32"
                    exit
                    action drop
                    exit
                exit
                entry 20
                exit
                default-action accept
                exit
            exit
```

The policy can be assigned to mLDP using the **configure router ldp import-mcast-policy** *policy1* command. Based on this configuration, the prefix list will match the mLDP outer FEC and the action will be executed.

Note: mLDP import policies are only supported for IPv4 FECs.

The mLDP import policy is useful for enforcing root only functionality on a network. For a PE to be a root only, enable the mLDP import policy to drop any arriving FEC on the P router.

7.1.8.1 Recursive FEC behavior

In the case of recursive FEC, the prefix list will match the outer root. For example, for recursive FEC <outerROOT, opaque <ActualRoot, opaque<lspID>> the import policy will work on the outerROOT of the FEC.

The policy only matches to the outer root address of the FEC and no other field in the FEC.

7.1.8.2 Import Policy

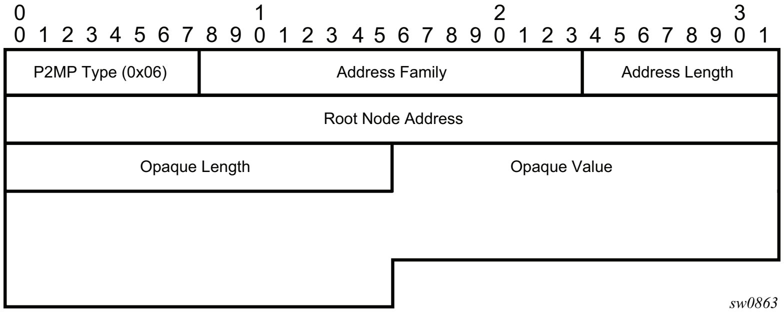
For mLDP, a policy can be assigned as an import policy only. Import policies only affect FECs arriving to the node, and will not affect the self-generated FECs on the node. The import policy will cause the multicast FECs received from the peer to be rejected and stored in the LDP database but not resolved. Therefore, the **show router ldp binding** command will display the FEC but the FEC will not be shown by the **show router ldp binding active** command. The FEC is not resolved if it is not allowed by the policy.

Only global import policies are supported for mLDP FEC. Per-peer import policies are not supported.

As defined in RFC 6388 for P2MP FEC, SR OS will only match the prefix against the root node address field of the FEC, and no other fields. This means that the policy will work on all P2MP Opaque types.

The P2MP FEC Element is encoded as shown in Figure 71.

Figure 71 **P2MP FEC Element Encoding**



7.1.9 LDP FEC Resolution Per Specified Community

LDP communities provide separation between groups of FECs at the LDP session level. LDP sessions are assigned a community value and any FECs received or advertised over them are implicitly associated with that community.

Note: The community value only has local significance to a node. The user must therefore ensure that communities are assigned consistently to sessions across the network.

SR OS supports multiple targeted LDP sessions over a specified network IP interface between LDP peer systems, each with its own local LSR ID. This makes it especially suitable for building multiple LDP overlay topologies over a common IP infrastructure, each with their own community.

LDP FEC resolution per specified community is supported in combination with stitching to SR or BGP tunnels as follows.

- Although a FEC is only advertised within a given LDP community, FEC can resolve to SR or BGP tunnels if those are the only available tunnels.
- If LDP has received a label from an LDP peer with an assigned community, that FEC is assigned the community of that session.
- If no LDP peer has advertised the label, LDP leaves the FEC with no community.
- The FEC may be resolvable over an SR or BGP tunnel, but the community it is assigned at the stitching node depends on whether LDP has also advertised that FEC to that node, and the community assigned to the LDP session over which the FEC was advertised.

7.1.9.1 Configuration

A community is assigned to an LDP session by configuring a community string in the corresponding session parameters for the peer or the targeted session peer template. A community only applies to a **local-lsr-id** for a session. It is never applied to a system FEC or local static FEC. The **no local-lsr-id** or **local-lsr-id system** commands are synonymous and mean that there is no local LSR ID for a session. A system FEC or static FEC cannot have a community associated with it and is therefore not advertised over an LDP session with a configured community. Only a single community string can be configured for a session towards a specified peer or within a specified targeted peer template. The FEC advertised by the **adv-local-lsr-id** command is automatically put in the community configured on the session.

The specified community is only associated to IPv4 and IPv6 Address FECs incoming or outgoing on the relevant session, and not to IPv4/IPv6 P2MP FECs, or service FECs incoming/outgoing on the session.

Static FECs are treated as having no community associated with them, even if they are also received over another session with an assigned community. A mismatch is declared if this situation arises.

7.1.9.2 Operation

If a FEC is received over a session of a specified community, it is assumed to be associated with that community and is only broadcast to peers using sessions of that community. Likewise, a FEC received over a session with no community is only broadcast over other sessions with no community.

If a FEC is received over a session that does not have an assigned community, the FEC is treated as if it was received from a session with a differing assigned community. In other words, any particular FEC must only be received from sessions with a single, assigned community or no community. In any other case (from sessions with differing communities, or from a combination of sessions with a community and sessions without a community), the FEC is considered to have a community mismatch.

The following procedures apply.

1. The system remembers the first community (including no community) of the session that a FEC is received on.
2. If the same FEC is subsequently received over a session with a differing community, the FEC is marked as mismatched and the system raises a trap indicating community mismatch.

Note: Subsequent traps due to a mismatch for a FEC arriving over a session of the same community (or no community) are squelched for a period of 60 seconds after the first trap. The trap indicates the session and the community of the session, but does not need to indicate the FEC itself.

3. After a FEC has been marked as mismatched, the FEC is no longer advertised over sessions (or resolved to sessions) that differ either from the original community or in whether a community has been assigned. This can result in asymmetrical leaking of traffic between communities in certain cases, as illustrated by the following scenario. It is therefore recommended that FEC mismatches be resolved as soon as possible after they occur.

   Consider a triangle topology of Nodes A-B-C with iLDP sessions between them, using community=RED. At bootstrap, all the adv-local-lsrId FECs are exchanged, and the FECs are activated correctly as per routing. On each node, for each FEC there will be a [local push] and a [local swap] as there is more than one peer advertising such a FEC. At this point all FECs are marked as being RED.

   - Focusing on Node C, consider:
     - Node A-owned RED FEC=X/32
     - Node B-owned RED FEC=Y/32
   - On Node C, the community of the session to node B is changed to BLUE. The consequence of this on Node C follows:
     - The [swap] operation for the remote Node A RED FEC=X/32 is de-programmed, as the Node B peer now BLUE, and therefore are not receiving Node A FEC=X/32 from B. Only the push is left programmed.
     - The [swap] operation for the remote Node B RED FEC=Y/32, is still programmed, even though this RED FEC is in mismatch, as it is received from both the BLUE peer Node B and the RED peer Node C.
4. When a session community changes, the session is flapped and the FEC community audited. If the original session is flapped, the FEC community changes as well. The following scenarios illustrate the operation of FEC community auditing.
   - Scenario A
     - The FEC comes in on blue session A. The FEC is marked blue.
     - The FEC comes in on red session B. The FEC is marked “mismatched” and stays blue.
     - Session B is changed to green. Session B is bounced. The FEC community is audited, stays blue, and stays mismatched.
   - Scenario B
     - The FEC comes in on blue session A. The FEC is marked blue.

    - The FEC comes in on red session B. The FEC is marked “mismatched” and stays blue.
    - Session A is changed to red. The FEC community audit occurs. The “mismatch” indication is cleared and the FEC is marked as red. The FEC remains red when session A comes back up.
  - Scenario C
    - The FEC comes in on blue session A. The FEC is marked blue.
    - The FEC comes in on red session B. The FEC is marked “mismatched” and stays blue.
    - Session A goes down. The FEC community audit occurs. The FEC is marked as red and the “mismatch” indication is cleared. The FEC is advertised over red session B.
    - Session A subsequently comes back up and it is still blue. The FEC remains red but is marked “mismatched”. The FEC is no longer advertised over blue session A.

The community mismatch state for a prefix FEC is visible through the **show**>**router**>**ldp**>**bindings**>**prefixes** command output, while the community mismatch state is visible via a MIB flag (in the vRtrLdpNgAddrFecFlags object).

The fact that a FEC is marked “mismatched” has no bearing on its accounting with respect to the limit of the number of FECs that may be received over a session.

The ability of a policy to reject a FEC is independent of the FEC mismatch. A policy prevents the system from using the label for resolution, but if the corresponding session is sending community-mismatched FECs, there is a problem and it should be flagged. For example, the policy and community mismatch checks are independent, and a FEC should still be marked with a community mismatch, if needed, per the rules above

7.1.10 T-LDP hello reduction

This feature implements a new mechanism to suppress the transmission of the Hello messages following the establishment of a Targeted LDP session between two LDP peers. The Hello adjacency of the targeted session does not require periodic transmission of Hello messages as in the case of a link LDP session. In link LDP, one or more peers can be discovered over a given network IP interface and as such, the periodic transmission of Hello messages is required to discover new peers in addition to the periodic Keep-Alive message transmission to maintain the existing LDP sessions. A Targeted LDP session is established to a single peer. Thus, once the Hello Adjacency is established and the LDP session is brought up over a TCP connection, Keep-Alive messages are sufficient to maintain the LDP session.

When this feature is enabled, the targeted Hello adjacency is brought up by advertising the Hold-Time value the user configured in the Hello timeout parameter for the targeted session. The LSR node will then start advertising an exponentially increasing Hold-Time value in the Hello message as soon as the targeted LDP session to the peer is up. Each new incremented Hold-Time value is sent in a number of Hello messages equal to the value of the Hello reduction factor before the next exponential value is advertised. This provides time for the two peers to settle on the new value. When the Hold-Time reaches the maximum value of 0xffff (binary 65535), the two peers will send Hello messages at a frequency of every [(65535-1)/local helloFactor] seconds for the lifetime of the targeted-LDP session (for example, if the local Hello Factor is three (3), then Hello messages will be sent every 21844 seconds).

Both LDP peers must be configured with this feature to bring gradually their advertised Hold-Time up to the maximum value. If one of the LDP peers does not, the frequency of the Hello messages of the targeted Hello adjacency will continue to be governed by the smaller of the two Hold-Time values. This feature complies to *draft-pdutta-mpls-tldp-hello-reduce*.

7.1.11 Tracking a T-LDP Peer with BFD

BFD tracking of an LDP session associated with a T-LDP adjacency allows for faster detection of the liveliness of the session by registering the peer transport address of a LDP session with a BFD session. The source or destination address of the BFD session is the local or remote transport address of the targeted or link (if peers are directly connected) Hello adjacency which triggered the LDP session.

By enabling BFD for a selected targeted session, the state of that session is tied to the state of the underneath BFD session between the two nodes. The parameters used for the BFD are set with the BFD command under the IP interface which has the source address of the TCP connection.

7.1.12 Link LDP Hello Adjacency Tracking with BFD

LDP can only track an LDP peer using the Hello and Keep-Alive timers. If an IGP protocol registered with BFD on an IP interface to track a neighbor, and the BFD session times out, the next-hop for prefixes advertised by the neighbor are no longer resolved. This however does not bring down the link LDP session to the peer since the LDP peer is not directly tracked by BFD.

In order to properly track the link LDP peer, LDP needs to track the Hello adjacency to its peer by registering with BFD.

The user effects Hello adjacency tracking with BFD by enabling BFD on an LDP interface:

config>**router**>**ldp**>**if-params**>**if**>**enable-bfd** [**ipv4**][**ipv6**]

The parameters used for the BFD session, that is, transmit-interval, receive-interval, and multiplier, are those configured under the IP interface:

config>**router**>**if**>**bfd**

The source or destination address of the BFD session is the local or remote address of link Hello adjacency. When multiple links exist to the same LDP peer, a Hello adjacency is established over each link. However, a single LDP session will exist to the peer and will use a TCP connection over one of the link interfaces. Also, a separate BFD session should be enabled on each LDP interface. If a BFD session times out on a specific link, LDP will immediately bring down the Hello adjacency on that link. In addition, if there are FECs that have their primary NHLFE over this link, LDP triggers the LDP FRR procedures by sending to IOM and line cards the neighbor/next-hop down message. This will result in moving the traffic of the impacted FECs to an LFA next-hop on a different link to the same LDP peer or to an LFA backup next-hop on a different LDP peer depending on the lowest backup cost path selected by the IGP SPF.

As soon as the last Hello adjacency goes down as a result of the BFD timing out, the LDP session goes down and the LDP FRR procedures will be triggered. This will result in moving the traffic to an LFA backup next-hop on a different LDP peer.

7.1.13 LDP LSP Statistics

RSVP-TE LSP statistics is extended to LDP to provide the following counters:

- Per-forwarding-class forwarded in-profile packet count

- Per-forwarding-class forwarded in-profile byte count
- Per-forwarding-class forwarded out-of-profile packet count
- Per-forwarding-class forwarded out-of-profile byte count

The counters are available for the egress data path of an LDP FEC at ingress LER and at LSR. Because an ingress LER is also potentially an LSR for an LDP FEC, combined egress data path statistics will be provided whenever applicable.

7.1.14 MPLS Entropy Label

The router supports the MPLS entropy label (RFC 6790) on LDP LSPs used for IGP and BGP shortcuts. This allows LSR nodes in a network to load-balance labeled packets in a much more granular fashion than allowed by simply hashing on the standard label stack.

To configure insertion of the entropy label on IGP or BGP shortcuts, use using the **entropy-label** command under the **configure router** context.

7.1.15 Importing LDP Tunnels to Non-Host Prefixes to TTM

When an LDP LSP is established, TTM is automatically populated with the corresponding tunnel. This automatic behavior does not apply to non-host prefixes. The **config**>**router**>**ldp**>**import-tunnel-table** command allows for TTM to be populated with LDP tunnels to such prefixes in a controlled manner for both IPv4 and IPv6.

7.2 TTL Security for BGP and LDP

The BGP TTL Security Hack (BTSH) was originally designed to protect the BGP infrastructure from CPU utilization-based attacks. It is derived from the fact that the vast majority of ISP EBGP peerings are established between adjacent routers. Since TTL spoofing is considered nearly impossible, a mechanism based on an expected TTL value can provide a simple and reasonably robust defense from infrastructure attacks based on forged BGP packets.

While TTL Security Hack (TSH) is most effective in protecting directly connected peers, it can also provide a lower level of protection to multi-hop sessions. When a multi-hop BGP session is required, the expected TTL value can be set to 255 minus the configured range-of-hops. This approach can provide a qualitatively lower degree of security for BGP (such as a DoS attack could, theoretically, be launched by compromising a box in the path). However, BTSH will catch a vast majority of observed distributed DoS (DDoS) attacks against EBGP.

TSH can be used to protect LDP peering sessions as well. For details, see draft-chen-ldp-ttl-xx.txt, *TTL-Based Security Option for LDP Hello Message*.

The TSH implementation supports the ability to configure TTL security per BGP/LDP peer and evaluate (in hardware) the incoming TTL value against the configured TTL value. If the incoming TTL value is less than the configured TTL value, the packets are discarded and a log is generated.

7.3 ECMP Support for LDP

ECMP support for LDP performs load balancing for LDP based LSPs by having multiple outgoing next-hops for a given IP prefix on ingress and transit LSRs.

An LSR that has multiple equal cost paths to a given IP prefix can receive an LDP label mapping for this prefix from each of the downstream next-hop peers. As the LDP implementation uses the liberal label retention mode, it retains all the labels for an IP prefix received from multiple next-hop peers.

Without ECMP support for LDP, only one of these next-hop peers is selected and installed in the forwarding plane. The algorithm used to determine the next-hop peer to be selected involves looking up the route information obtained from the RTM for this prefix and finding the first valid LDP next-hop peer (for example, the first neighbor in the RTM entry from which a label mapping was received). If, for some reason, the outgoing label to the installed next-hop is no longer valid, say the session to the peer is lost or the peer withdraws the label, a new valid LDP next-hop peer is selected out of the existing next-hop peers and LDP reprograms the forwarding plane to use the label sent by this peer.

With ECMP support, all the valid LDP next-hop peers, those that sent a label mapping for a given IP prefix, are installed in the forwarding plane. In both cases, ingress LER and transit LSR, an ingress label are mapped to the next hops that are in the RTM and from which a valid mapping label has been received. The forwarding plane then uses an internal hashing algorithm to determine how the traffic is distributed amongst these multiple next-hops, assigning each “flow” to a particular next-hop.

The hash algorithm at LER and transit LSR are described in the "Traffic Load Balancing Options" in the *7450 ESS, 7750 SR, 7950 XRS, and VSR Interface Configuration Guide*.

LDP supports up to 64 ECMP next hops. LDP takes its maximum limit from the lower of **config**>**router**>**ecmp** and **config**>**router**>**ldp**>**max-ecmp-routes**.

7.3.1 Label Operations

If an LSR is the ingress for a given IP prefix, LDP programs a push operation for the prefix in the forwarding engine. This creates an LSP ID to the Next Hop Label Forwarding Entry (NHLFE) (LTN) mapping and an LDP tunnel entry in the forwarding plane. LDP will also inform the Tunnel Table Manager (TTM) of this tunnel. Both the LTN entry and the tunnel entry will have a NHLFE for the label mapping that the LSR received from each of its next-hop peers.

If the LSR is to behave as a transit for a given IP prefix, LDP will program a swap operation for the prefix in the forwarding engine. This involves creating an Incoming Label Map (ILM) entry in the forwarding plane. The ILM entry will have to map an incoming label to possibly multiple NHLFEs. If an LSR is an egress for a given IP prefix, LDP will program a POP entry in the forwarding engine. This too will result in an ILM entry being created in the forwarding plane but with no NHLFEs.

When unlabeled packets arrive at the ingress LER, the forwarding plane will consult the LTN entry and will use a hashing algorithm to map the packet to one of the NHLFEs (push label) and forward the packet to the corresponding next-hop peer. For labeled packets arriving at a transit or egress LSR, the forwarding plane will consult the ILM entry and either use a hashing algorithm to map it to one of the NHLFEs if they exist (swap label) or simply route the packet if there are no NHLFEs (pop label).

Static FEC swap will not be activated unless there is a matching route in system route table that also matches the user configured static FEC next-hop.

7.3.2 Weighted ECMP Support for LDP

The router supports weighted ECMP in cases where LDP resolves a FEC over an ECMP set of direct next hops corresponding to IP network interfaces, and where it resolves the FEC over an ECMP set of RSVP-TE tunnels. See Weighted Load Balancing for LDP over RSVP for information about LDP over RSVP.

Weighted ECMP for direct IP network interfaces uses a **load-balancing-weight** configured under the **config**>**router**>**ldp**>**interface-parameters**>**interface** context. Similar to LDP over RSVP, Weighted ECMP for LDP is enabled using the **weighted-ecmp** command under the **config**>**router**>**ldp** context. If the interface becomes an ECMP next hop for an LDP FEC, and all the other ECMP next hops are interfaces with configured (non-zero) load-balancing weights, then the traffic distribution over the ECMP interfaces is proportional to the normalized weight. Then, LDP performs the normalization with a granularity of 64.

If one or more of the LDP interfaces in the ECMP set does not have a configured-load-balancing weight, then the system falls back to ECMP.

If both an IGP shortcut tunnel and a direct next hop exist to resolve a FEC, LDP prefers the tunneled resolution. Therefore, if an ECMP set consists of both IGP shortcuts and direct next hops, LDP only load balances across the IGP shortcuts.

➡ **Note:** LDP only uses configured LDP interface load balancing weights with non-LDP over RSVP resolutions.

7.4 Unnumbered Interface Support in LDP

This feature allows LDP to establish Hello adjacency and to resolve unicast and multicast FECs over unnumbered LDP interfaces.

This feature also extends the support of **lsp-ping**, **p2mp-lsp-ping**, and **ldp-treetrace** to test an LDP unicast or multicast FEC which is resolved over an unnumbered LDP interface.

7.4.1 Feature Configuration

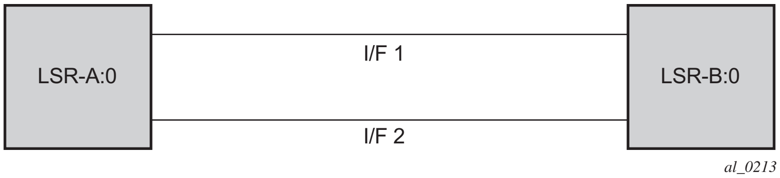
This feature does not introduce a new CLI command for adding an unnumbered interface into LDP. Rather, the **fec-originate** command is extended to specify the interface name because an unnumbered interface does not have an IP address of its own. The user can, however, specify the interface name for numbered interfaces.

See the CLI section for the changes to the **fec-originate** command.

7.4.2 Operation of LDP over an Unnumbered IP Interface

Consider the setup shown in Figure 72.

Figure 72 **LDP Adjacency and Session over Unnumbered Interface**



LSR A and LSR B have the following LDP identifiers respectively:

<LSR Id=A> : <label space id=0>

<LSR Id=B> : <label space id=0>

There are two P2P unnumbered interfaces between LSR A and LSR B. These interfaces are identified on each system with their unique local link identifier. In other words, the combination of {Router-ID, Local Link Identifier} uniquely identifies the interface in OSPF or IS-IS throughout the network.

A borrowed IP address is also assigned to the interface to be used as the source address of IP packets which need to be originated from the interface. The borrowed IP address defaults to the system loopback interface address, A and B respectively in this setup. The user can change the borrowed IP interface to any configured IP interface, loopback or not, by applying the following command:

config>router>if>unnumbered [<ip-int-name | ip-address>]

When the unnumbered interface is added into LDP, it will have the following behavior.

7.4.2.1 Link LDP

Hello adjacency will be brought up using link Hello packet with source IP address set to the interface borrowed IP address and a destination IP address set to 224.0.0.2.

As a consequence of (1), Hello packets with the same source IP address should be accepted when received over parallel unnumbered interfaces from the same peer LSR-ID. The corresponding Hello adjacencies would be associated with a single LDP session.

The transport address for the TCP connection, which is encoded in the Hello packet, will always be set to the LSR-ID of the node regardless if the user enabled the interface option under **config>router>ldp>if-params>if>ipv4>transport-address**.

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 to some other interface name, loopback or not, numbered or not. If the local interface is selected or the provided interface name corresponds to an unnumbered IP interface, the unnumbered interface borrowed IP address will be used as the LSR-ID. In all cases, the transport address for the LDP session will be updated to the new LSR-ID value but the link Hello packets will continue to use the interface borrowed IP address as the source IP address.

The LSR with the highest transport address, that is, LSR-ID in this case, will bootstrap the TCP connection and LDP session.

Source and destination IP addresses of LDP packets are the transport addresses, that is, LDP LSR-IDs of systems A and B in this case.

7.4.2.2 Targeted LDP

Source and destination addresses of targeted Hello packet are the LDP LSR-IDs of systems A and B. The user can configure the **local-lsr-id** option on the targeted session and change the value of the LSR-ID to either the local interface or to some other interface name, loopback or not, numbered or not. If the local interface is selected or the provided interface name corresponds to an unnumbered IP interface, the unnumbered interface borrowed IP address will be used as the LSR-ID. In all cases, the transport address for the LDP session and the source IP address of targeted Hello message will be updated to the new LSR-ID value.

The LSR with the highest transport address, that is, LSR-ID in this case, will bootstrap the TCP connection and LDP session. Source and destination IP addresses of LDP messages are the transport addresses, that is, LDP LSR-IDs of systems A and B in this case.

7.4.2.3 FEC Resolution

LDP will advertise/withdraw unnumbered interfaces using the Address/Address-Withdraw message. 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 next-hop are looked up in RTM cache. The next-hop consists of the router-id and link identifier of the interface at the peer LSR.

LDP FEC ECMP next-hops over a mix of unnumbered and numbered interfaces is supported.

All LDP FEC types are supported.

The **fec-originate** command is supported when the next-hop is over an unnumbered interface.

All LDP features are supported except for the following:

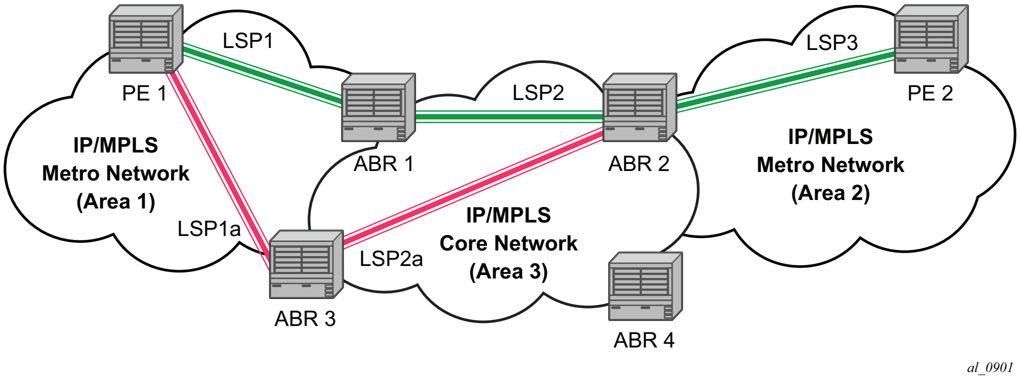
- BFD cannot be enabled on an unnumbered LDP interface. This is a consequence of the fact that BFD is not supported on unnumbered IP interface on the system.
- As a consequence of (1), LDP FRR procedures will not be triggered via a BFD session timeout but only by physical failures and local interface down events.
- Unnumbered IP interfaces cannot be added into LDP global and peer prefix policies.

7.5 LDP over RSVP Tunnels

LDP over RSVP-TE provides end-to-end tunnels that have two important properties, fast reroute and traffic engineering which are not available in LDP. LDP over RSVP-TE is focused at large networks (over 100 nodes in the network). Simply using end-to-end RSVP-TE tunnels will not scale. While an LER may not have that many tunnels, any transit node will potentially have thousands of LSPs, and if each transit node also has to deal with detours or bypass tunnels, this number can make the LSR overly burdened.

LDP over RSVP-TE allows tunneling of user packets using an LDP LSP inside an RSVP LSP. The main application of this feature is for deployment of MPLS based services, for example, VPRN, VLL, and VPLS services, in large scale networks across multiple IGP areas without requiring full mesh of RSVP LSPs between PE routers.

Figure 73 **LDP over RSVP Application**



The network displayed in Figure 73 consists of two metro areas, Area 1 and 2 respectively, and a core area, Area 3. Each area makes use of TE LSPs to provide connectivity between the edge routers. In order to enable services between PE1 and PE2 across the three areas, LSP1, LSP2, and LSP3 are set up using RSVP-TE. There are in fact 6 LSPs required for bidirectional operation but we will refer to each bi-directional LSP with a single name, for example, LSP1. A targeted LDP (T-LDP) session is associated with each of these bidirectional LSP tunnels. That is, a T-LDP adjacency is created between PE1 and ABR1 and is associated with LSP1 at each end. The same is done for the LSP tunnel between ABR1 and ABR2, and finally between ABR2 and PE2. The loopback address of each of these routers is advertised using T-LDP. Similarly, backup bidirectional LDP over RSVP tunnels, LSP1a and LSP2a, are configured by way of ABR3.

This setup effectively creates an end-to-end LDP connectivity which can be used by all PEs to provision services. The RSVP LSPs are used as a transport vehicle to carry the LDP packets from one area to another. Only the user packets are tunneled over the RSVP LSPs. The T-LDP control messages are still sent unlabeled using the IGP shortest path.

In this application, the bi-directional RSVP LSP tunnels are not treated as IP interfaces and are not advertised back into the IGP. A PE must always rely on the IGP to look up the next hop for a service packet. LDP-over-RSVP introduces a new tunnel type, tunnel-in-tunnel, in addition to the existing LDP tunnel and RSVP tunnel types. If multiple tunnels types match the destination PE FEC lookup, LDP will prefer an LDP tunnel over an LDP-over-RSVP tunnel by default.

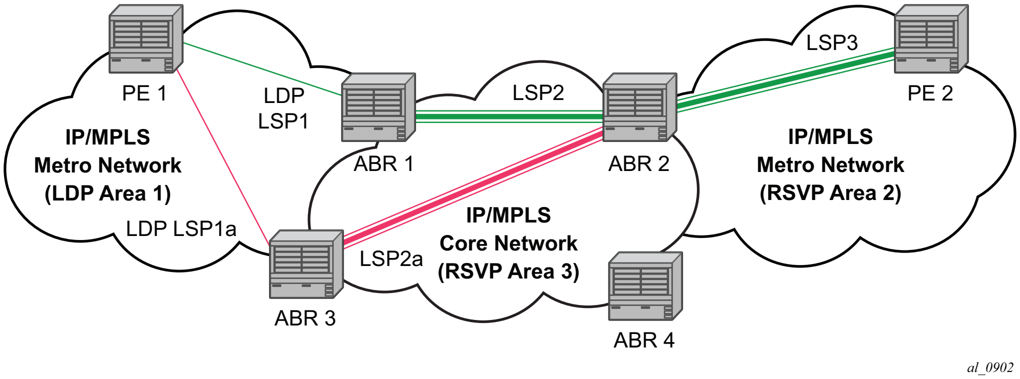
The design in Figure 73 allows a service provider to build and expand each area independently without requiring a full mesh of RSVP LSPs between PEs across the three areas.

To participate in a VPRN service, the PE1 and PE2 perform the autobind to LDP. The LDP label which represents the target PE loopback address is used below the RSVP LSP label. Therefore a 3 label stack is required.

In order to provide a VLL service, PE1 and PE2 are still required to set up a targeted LDP session directly between them. Again a 3 label stack is required, the RSVP LSP label, followed by the LDP label for the loopback address of the destination PE, and finally the pseudowire label (VC label).

This implementation supports a variation of the application in Figure 73, in which area 1 is an LDP area. In that case, PE1 will push a two label stack while ABR1 will swap the LDP label and push the RSVP label as illustrated in Figure 74. LDP-over-RSVP tunnels can also be used as IGP shortcuts.

Figure 74 **LDP over RSVP Application Variant**



7.5.1 Signaling and Operation

7.5.1.1 LDP Label Distribution and FEC Resolution

The user creates a targeted LDP (T-LDP) session to an ABR or the destination PE. This results in LDP hellos being sent between the two routers. These messages are sent unlabeled over the IGP path. Next, the user enables LDP tunneling on this T-LDP session and optionally specifies a list of LSP names to associate with this T-LDP session. By default, all RSVP LSPs which terminate on the T-LDP peer are candidates for LDP-over-RSVP tunnels. At this point in time, the LDP FECs resolving to RSVP LSPs are added into the Tunnel Table Manager as tunnel-in-tunnel type.

If LDP is running on regular interfaces also, the prefixes LDP learns are going to be distributed over both the T-LDP session as well as regular IGP interfaces. LDP FEC prefixes with a subnet mask lower or equal than 32 will be resolved over RSVP LSPs. The policy controls which prefixes go over the T-LDP session, for example, only /32 prefixes, or a particular prefix range.

LDP-over-RSVP works with both OSPF and ISIS. These protocols include the advertising router when adding an entry to the RTM. LDP-over-RSVP tunnels can be used as shortcuts for BGP next-hop resolution.

7.5.1.2 Default FEC Resolution Procedure

When LDP tries to resolve a prefix received over a T-LDP session, it performs a lookup in the Routing Table Manager (RTM). This lookup returns the next hop to the destination PE and the advertising router (ABR or destination PE itself). If the next-hop router advertised the same FEC over link-level LDP, LDP will prefer the LDP tunnel by default unless the user explicitly changed the default preference using the system wide prefer-tunnel-in-tunnel command. If the LDP tunnel becomes unavailable, LDP will select an LDP-over-RSVP tunnel if available.

When searching for an LDP-over-RSVP tunnel, LDP selects the advertising router(s) with best route. If the advertising router matches the T-LDP peer, LDP then performs a second lookup for the advertising router in the Tunnel Table Manager (TTM) which returns the user configured RSVP LSP with the best metric. If there are more than one configured LSP with the best metric, LDP selects the first available LSP.

If all user configured RSVP LSPs are down, no more action is taken. If the user did not configure any LSPs under the T-LDP session, the lookup in TTM will return the first available RSVP LSP which terminates on the advertising router with the lowest metric.

7.5.1.3 FEC Resolution Procedure When prefer-tunnel-in-tunnel is Enabled

When LDP tries to resolve a prefix received over a T-LDP session, it performs a lookup in the Routing Table Manager (RTM). This lookup returns the next hop to the destination PE and the advertising router (ABR or destination PE itself).

When searching for an LDP-over-RSVP tunnel, LDP selects the advertising router(s) with best route. If the advertising router matches the targeted LDP peer, LDP then performs a second lookup for the advertising router in the Tunnel Table Manager (TTM) which returns the user configured RSVP LSP with the best metric. If there are more than one configured LSP with the best metric, LDP selects the first available LSP.

If all user configured RSVP LSPs are down, then an LDP tunnel will be selected if available.

If the user did not configure any LSPs under the T-LDP session, a lookup in TTM will return the first available RSVP LSP which terminates on the advertising router. If none are available, then an LDP tunnel will be selected if available.

7.5.2 Rerouting Around Failures

Every failure in the network can be protected against, except for the ingress and egress PEs. All other constructs have protection available. These constructs are LDP-over-RSVP tunnel and ABR.

7.5.2.1 LDP-over-RSVP Tunnel Protection

An RSVP LSP can deal with a failure in two ways:

- If the LSP is a loosely routed LSP, then RSVP will find a new IGP path around the failure, and traffic will follow this new path. This may involve some churn in the network if the LSP comes down and then gets re-routed. The tunnel damping feature was implemented on the LSP so that all the dependent protocols and applications do not flap unnecessarily.
- If the LSP is a CSPF-computed LSP with the fast reroute option enabled, then RSVP will switch to the detour path very quickly. From that point, a new LSP will be attempted from the head-end (global revertive). When the new LSP is in place, the traffic switches over to the new LSP with make-before-break.

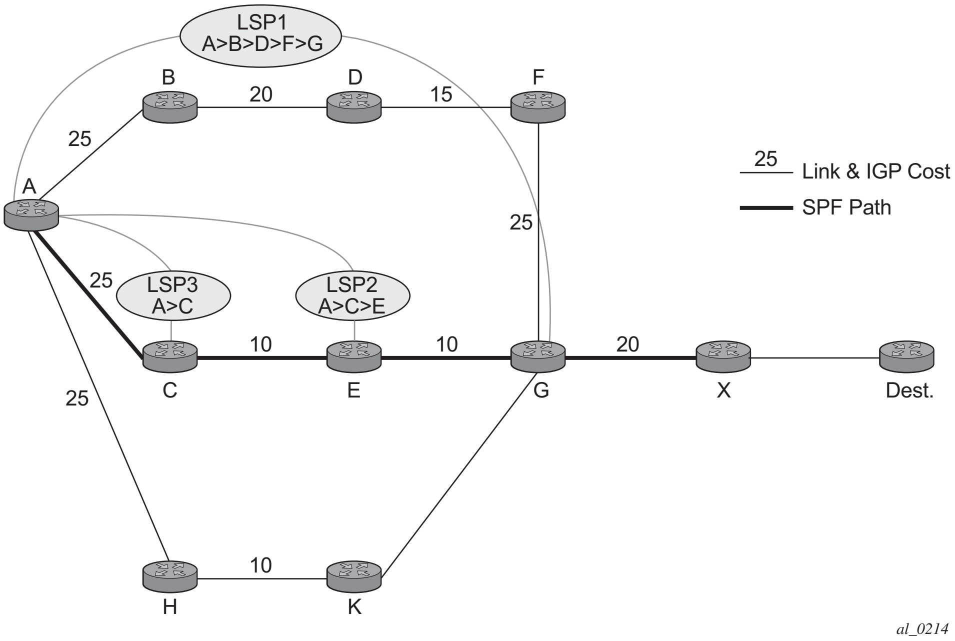
7.5.2.2 ABR Protection

If an ABR fails, then routing around the ABR requires that a new next-hop LDP-over-RSVP tunnel be found to a backup ABR. If an ABR fails, then the T-LDP adjacency fails. Eventually, the backup ABR becomes the new next hop (after SPF converges), and LDP learns of the new next-hop and can reprogram the new path.

7.6 LDP over RSVP Without Area Boundary

The LDP over RSVP capability set includes the ability to stitch LDP-over-RSVP tunnels at internal (non-ABR) OSPF and IS-IS routers.

Figure 75 **LDP over RSVP Without ABR Stitching Point**



In Figure 75, assume that the user wants to use LDP over RSVP between router A and destination “Dest”. The first thing that happens is that either OSPF or IS-IS will perform an SPF calculation resulting in an SPF tree. This tree specifies the lowest possible cost to the destination. In the example shown, the destination “Dest” is reachable at the lowest cost through router X. The SPF tree will have the following path: A>C>E>G>X.

Using this SPF tree, router A will search for the endpoint that is closest (farthest/highest cost from the origin) to “Dest” that is eligible. Assuming that all LSPs in the above diagram are eligible, LSP endpoint G will be selected as it terminates on router G while other LSPs only reach routers C and E, respectively.

IGP and LSP metrics associated with the various LSP are ignores; only tunnel endpoint matters to IGP. The endpoint that terminates closest to “Dest” (highest IGP path cost) will be selected for further selection of the LDP over RSVP tunnels to that endpoint. The explicit path the tunnel takes may not match the IGP path that the SPF computes.

If router A and G have an additional LSP terminating on router G, there would now be two tunnels both terminating on the same router closest to the final destination. For IGP, it does not make any difference on the numbers of LDPs to G, only that there is at least one LSP to G. In this case, the LSP metric will be considered by LDP when deciding which LSP to stitch for the LDP over RSVP connection.

The IGP only passes endpoint information to LDP. LDP looks up the tunnel table for all tunnels to that endpoint and picks up the one with the least tunnel metric. There may be many tunnels with the same least cost. LDP FEC prefixes with a subnet mask lower or equal than 32 will be resolved over RSVP LSPs within an area.

7.6.1 LDP over RSVP and ECMP

ECMP for LDP over RSVP is supported (also see ECMP Support for LDP). If ECMP applies, all LSP endpoints found over the ECMP IGP path will be installed in the routing table by the IGP for consideration by LDP. IGP costs to each endpoint may differ because IGP selects the farthest endpoint per ECMP path.

LDP will choose the endpoint that is highest cost in the route entry and will do further tunnel selection over those endpoints. If there are multiple endpoints with equal highest cost, then LDP will consider all of them.

7.7 Weighted Load Balancing for LDP over RSVP

Weighted load balancing (Weighted ECMP) is supported for LDP over RSVP (LoR), when the LDP next hop resolves to an IGP shortcut tunnel over RSVP, when it resolves to a static route with next hops which in turn uses RSVP tunnels, and where the **tunneling** option is configured for the LDP peer (classical LDP over RSVP). Weighted load balancing is supported for both push and swap NHLFEs.

At a high level, the feature operates as follows.

- All of the RSVP LSPs in the ECMP set should have a **load-balancing-weight** configured, otherwise non-weighted ECMP behavior is used.
- The normalized weight of each RSVP LSP is calculated based on its configured load-balancing weight. The calculation is performed by LDP to a resolution of 64. These next hops are then populated in TTM.
- RTM entries are updated accordingly for LDP shortcuts.
- When weighted ECMP is configured for LDP, the normalized weight is downloaded to the IOM when the LDP route is resolved. This occurs for both push and swap NHLFEs.
- LDP labeled packets are then sprayed in proportion to the normalized weight of the RSVP LSPs that they are forwarded over.
- There is no per-service differentiation between packets. LDP labeled packets from all services are sprayed in proportion to the normalized weight.
- Tunnel-in-tunnel takes precedence over the existence of a static route with a tunneled next hop. That is, if tunneling is configured, then LDP uses these LSPs rather than those used by the static route. This means that LDP may use different tunnels to those pointed to by static routes.

Weighted ECMP for LDP over RSVP, when using IGP shortcuts or static routes, is enabled as follows:

```
config
   router
      ldp
         [no] weighted-ecmp
```

However, in case of classic LoR, weighted ECMP only needs to be configured under LDP. The maximum number of ECMP tunnels is taken from the lower of the **config**>**router**>**ecmp** and **config**>**router**>**ldp**>**max-ecmp-routes** commands.

The following configuration illustrates the case of LDP resolving to a static route with one or more indirect next hops and a set of RSVP tunnels specified in the resolution filter:

```
config>router
   static-route-entry 192.0.2.102/32
      indirect 192.0.2.2
        tunnel-next-hop
           resolution-filter
              rsvp-te
                 lsp "LSP-ABR-1-1"
                 lsp "LSP-ABR-1-2"
                 lsp "LSP-ABR-1-3"
                 exit
           exit
     indirect 192.0.2.3
        tunnel-next-hop
           resolution-filter
              rsvp-te
                 lsp "LSP-ABR-2-1"
                 lsp "LSP-ABR-2-2"
                 lsp "LSP-ABR-2-3"
                 exit
           exit
           no shutdown
       exit
```

If both **config>router>weighted-ecmp** and **config>router>ldp>weighted-ecmp** are configured, then the weights of all of the RSVP tunnels for the static route are normalized to 64 and these are used to spray LDP labeled packets across the set of LSPs. This applies across all shortcuts (static and IGP) to which a route is resolved to the far-end prefix.

7.7.1 Interaction with Class-Based Forwarding

Class Based Forwarding (CBF) is not supported together with Weighted ECMP in LoR.

If both weighted ECMP and class-forwarding are configured under LDP, then LDP uses weighted ECMP only if all LSP next hops have non-default-weighted values configured. If any of the ECMP set LSP next hops do not have the weight configured, then LDP uses CBF. Otherwise, LDP uses CBF if possible. If weighted ECMP is configured for both LDP and the IGP shortcut for the RSVP tunnel, (**config**>**router**>**weighted-ecmp**), then weighted ECMP is used.

LDP resolves and programs FECs according to the weighted ECMP information if the following conditions are met.

- LDP has both CBF and weighted ECMP fully configured.
- All LSPs in ECMP set have both a load-balancing weight and CBF information configured.
- **weighted-ecmp** is enabled under **config**>**router**.

Subsequently, deleting the CBF configuration has no effect; however, deleting the weighted ECMP configuration causes LDP to resolve according to CBF, if complete, consistent CBF information is available. Otherwise LDP sprays over all the LSPs equally, using non-weighted ECMP behavior.

If the IGP shortcut tunnel using the RSVP LSP does not have complete weighted ECMP information (for example, **config**>**router**>**weighted-ecmp** is not configured or one or more of the RSVP tunnels has **no load-balancing-weight**) then LDP attempts CBF resolution. If the CBF resolution is complete and consistent, then LDP programs that resolution. If a complete, consistent CBF resolution is not received, then LDP sprays over all the LSPs equally, using regular ECMP behavior.

Where entropy labels are supported on LoR, the entropy label (both insertion and extraction at LER for the LDP label and hashing at LSR for the LDP label) is supported when weighted ECMP is in use.

7.8 Class-based Forwarding of LDP Prefix Packets over IGP Shortcuts

Within large ISP networks, services are typically required from any PE to any PE and can traverse multiple domains. Also, within a service, different traffic classes can co-exist, each with specific requirements on latency and jitter.

SR OS provides a comprehensive set of Class Based Forwarding capabilities. Specifically the following can be performed:

- Class-based forwarding, in conjunction with ECMP, for incoming unlabeled traffic resolving to an LDP FEC, over IGP IPv4 shortcuts (LER role)
- Class-based forwarding, in conjunction with ECMP, for incoming labeled LDP traffic, over IGP IPv4 shortcuts (LSR role)
- Class-based forwarding, in conjunction with ECMP, of GRT IPv4/IPv6 prefixes over IGP IPv4 shortcuts

  Refer to chapter IP Router Configuration, Section 2.3 in *7450 ESS, 7750 SR, 7950 XRS, and VSR Router Configuration Guide*, for a description of this case.
- Class-based forwarding, in conjunction with ECMP, of VPN-v4/-v6 prefixes over RSVP-TE or SR-TE

  Refer to chapter Virtual Private Routed Network Service, Section 3.2.27 in *7450 ESS, 7750 SR, 7950 XRS, and VSR Layer 3 Services Guide: IES and VPRN*, for a description of this case.

IGP IPv4 shortcuts, in all four cases, refer to MPLS RSVP-TE or SR-TE LSPs.

7.8.1 Configuration and Operation

The class-based forwarding feature enables service providers to control which LSPs, of a set of ECMP tunnel next hops that resolve an LDP FEC prefix, to forward packets that were classified to specific forwarding classes, as opposed to normal ECMP spraying where packets are sprayed over the whole set of LSPs.

To activate CBF, the user should enable the following:

- IGP shortcuts or forwarding adjacencies in the routing instance
- ECMP
- advertisement of unicast prefix FECs on the Targeted LDP session to the peer
- class-based forwarding in the LDP context (LSR role, LER role or both)

The **FC-to-Set based configuration** mode is controlled by the following commands:

config>router>mpls>class-forwarding-policy *policy-name*

config>router>mpls>class-forwarding-policy>fc> {**be** | **l2** | **af** | **l1** | **h2** | **ef** | **h1** | **nc**} **forwarding-set** *value*

config>router>mpls>class-forwarding-policy>default-set *value*

config>router>mpls>lsp>class-forwarding>forwarding-set policy *policy-name* **set** *set-id*

The last command applies to the **lsp-template** context. So, LSPs that are created from that template, acquire the assigned CBF configurations.

Multiple FCs can be assigned to a given set. Also, multiple LSPs can map to the same (policy, set) pair. However, an LSP cannot map to more than one (policy, set) pair.

Both configuration modes are mutually exclusive on a per LSP basis.

The CBF behavior depends on the configuration used, and on whether CBF was enabled for the LER and/or LSR role(s). The table below illustrates the different modes of operation of Class Based Forwarding depending on the node functionality where enabled, and on the type of configuration present in the ECMP set.

These modes of operation are explained in following sections.

7.8.1.1 LSR and/or LER Roles with FC-to-Set Configuration

Both LSR and LER roles behave in the same way with this type of configuration.

Before installing CBF information in the forwarding path, the system performs a consistency check on the CBF information of the ECMP set of tunnel next hops that resolve an LDP prefix FEC.

If no LSP, in the full ECMP set, has been assigned with a class forwarding policy configuration, the set is considered as inconsistent from a CBF perspective. The system, then, programs in the forwarding path, the whole ECMP set without any CBF information, and regular ECMP spraying occurs over the full set.

If the ECMP set is assigned to more than one class forwarding policy, the set is inconsistent from a CBF perspective. Then, the system programs, in the forwarding path, the whole ECMP set without any CBF information, and regular ECMP spraying occurs over the full set.

A full ECMP set is consistent from a CBF perspective when the ECMP:

- is assigned to a single class forwarding policy
- contains either an LSP assigned to the default set (implicit or explicit), or
- contains an LSP assigned to a non-default set that has explicit FC mappings

If there is no default set in a consistent ECMP set, the system automatically selects one set as the default one. The selected set is one set with the lowest ID among those referenced by the LSPs of the ECMP set.

If the ECMP set is consistent from a CBF perspective, the system programs in the forwarding path all the LSPs which have CBF configuration, and packets classified to a given FC are forwarded by using the LSPs of the corresponding forwarding set.

If there are more than one LSPs in a forwarding set, the system performs a modulo operation on these LSPs only to select one. As a result, ECMP spraying occurs for multiple packets of this forwarding class. Also, the system programs, in the forwarding path, the remaining LSPs of the ECMP set, without any CBF information. These LSPs are not used for class-based forwarding.

If there is no operational LSP in a given forwarding set, the system forwards packets which have been classified to the corresponding forwarding class onto the default set. Additionally, if there is no operational LSP in the default set, the system reverts to regular ECMP spraying over the full ECMP set.

If the user changes (by adding, modifying or deleting) the CBF configuration associated to an LSP that was previously selected as part of an ECMP set, then the FEC resolution is automatically updated, and a CBF consistency check is performed. Moreover, the user changes can update the forwarding configuration.

The LSR role applies to incoming labeled LDP traffic whose FEC is resolved to IGP IPv4 shortcuts.

The LER role applies to the following:

- IPv4 and IPv6 prefixes in GRT (with an IPv4 BGP NH)
- VPN-v4 and VPN-v6 routes

However, LER does not apply to any service which uses either explicit binding to an SDP (static or T-LDP signaled services), or auto-binding to SDP (BGP-AD VPLS, BGP-VPLS, BGP-VPWS, Dynamic MS-PW).

For BGP-LU, ECMP+CBF is supported only in the absence of the VPRN label. Therefore, ECMP+CBF is not supported when a VPRN label runs on top of BGP-LU (itself running over LDPoRSVP).

The CBF capability is available with any system profile. The number of sets is limited to four with system profile None or A, and to six with system profile B. This capability does not apply to CPM generated packets, including OAM packets, which are looked-up in RTM, and which are forwarded over tunnel next hops. These packets are forwarded by using either regular ECMP, or by selecting one nexthop from the set.

7.9 LDP ECMP Uniform Failover

LDP ECMP uniform failover allows the fast re-distribution by the ingress data path of packets forwarded over an LDP FEC next-hop to other next-hops of the same FEC when the currently used next-hop fails. The switchover is performed within a bounded time, which does not depend on the number of impacted LDP ILMs (LSR role) or service records (ingress LER role). The uniform failover time is only supported for a single LDP interface or LDP next-hop failure event.

This feature complements the coverage provided by the LDP Fast-ReRoute (FRR) feature, which provides a Loop-Free Alternate (LFA) backup next-hop with uniform failover time. Prefixes that have one or more ECMP next-hop protection are not programmed with a LFA back-up next-hop, and vice-versa.

The LDP ECMP uniform failover feature builds on the concept of Protect Group ID (PG-ID) introduced in LDP FRR. LDP assigns a unique PG-ID to all FECs that have their primary Next-Hop Label Forwarding Entry (NHLFE) resolved to the same outgoing interface and next-hop.

When an ILM record (LSR role) or LSPid-to-NHLFE (LTN) record (LER role) is created on the IOM, it has the PG-ID of each ECMP NHLFE the FEC is using.

When a packet is received on this ILM/LTN, the hash routine selects one of the up to 32, or the ECMP value configured on the system, whichever is less, ECMP NHLFEs for the FEC based on a hash of the packet’s header. If the selected NHLFE has its PG-ID in DOWN state, the hash routine re-computes the hash to select a backup NHLFE among the first 16, or the ECMP value configured on the system, whichever is less, NHLFEs of the FEC, excluding the one that is in DOWN state. Packets of the subset of flows that resolved to the failed NHLFE are thus sprayed among a maximum of 16 NHLFEs.

LDP then re-computes the new ECMP set to exclude the failed path and downloads it into the IOM. At that point, the hash routine will update the computation and begin spraying over the updated set of NHLFEs.

LDP sends the DOWN state update of the PG-ID to the IOM when the outgoing interface or a specific LDP next-hop goes down. This can be the result of any of the following events:

- Interface failure detected directly.
- Failure of the LDP session detected via T-LDP BFD or LDP Keep-Alive.
- Failure of LDP Hello adjacency detected via link LDP BFD or LDP Hello.

In addition, PIP will send an interface down event to the IOM if the interface failure is detected by other means than the LDP control plane or BFD. In that case, all PG-IDs associated with this interface will have their state updated by the IOM.

When tunneling LDP packets over an RSVP LSP, it is the detection of the T-LDP session going down, via BFD or Keep-Alive, which triggers the LDP ECMP uniform failover procedures. If the RSVP LSP alone fails and the latter is not protected by RSVP FRR, the failure event will trigger the re-resolution of the impacted FECs in the slow path.

When a multicast LDP (mLDP) FEC is resolved over ECMP links to the same downstream LDP LSR, the PG-ID DOWN state will cause packets of the FEC resolved to the failed link to be switched to another link using the linear FRR switchover procedures.

The LDP ECMP uniform failover is not supported in the following forwarding contexts:

- VPLS BUM packets.
- Packets forwarded to an IES/VPRN spoke-interface.
- Packets forwarded towards VPLS spoke in routed VPLS.

Finally, the LDP ECMP uniform failover is only supported for a single LDP interface, LDP next-hop, or peer failure event.

7.10 LDP Fast-Reroute for IS-IS and OSPF Prefixes

LDP Fast Re-Route (FRR) is a feature which allows the user to provide local protection for an LDP FEC by pre-computing and downloading to the IOM or XCM both a primary and a backup NHLFE for this FEC.

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 RTM. The backup NHLFE corresponds to the label received for the same FEC from a Loop-Free Alternate (LFA) next-hop.

The LFA next-hop pre-computation by IGP is described in RFC 5286 – “Basic Specification for IP Fast Reroute: Loop-Free Alternates”. LDP FRR relies on using the label-FEC binding received from the LFA next-hop to forward traffic for a given prefix as soon as the primary next-hop is not available. This means that a node resumes forwarding LDP packets to a destination prefix without waiting for the routing convergence. The label-FEC binding is received from the loop-free alternate next-hop ahead of time and is stored in the Label Information Base since LDP on the router operates in the liberal retention mode.

This feature requires that IGP performs the Shortest Path First (SPF) computation of an LFA next-hop, in addition to the primary next-hop, for all prefixes used by LDP to resolve FECs. IGP also populates both routes in the Routing Table Manager (RTM).

7.10.1 LDP FRR Configuration

The user enables Loop-Free Alternate (LFA) computation by SPF under the IS-IS or OSPF routing protocol level:

config>**router**>**isis**>**loopfree-alternates**
config>**router**>**ospf**>**loopfree-alternates**

The above commands instruct the IGP SPF to attempt to pre-compute both a primary next-hop and an LFA next-hop for every learned prefix. When found, the LFA next-hop is populated into the RTM along with the primary next-hop for the prefix.

Next the user enables the use by LDP of the LFA next-hop by configuring the following option:

config>**router**>**ldp**>**fast-reroute**

When this command is enabled, LDP will use both the primary next-hop and LFA next-hop, when available, for resolving the next-hop of an LDP FEC against the corresponding prefix in the RTM. This will result in LDP programming a primary NHLFE and a backup NHLFE into the IOM or XCM for each next-hop of a FEC prefix for the purpose of forwarding packets over the LDP FEC.

Because LDP can detect the loss of a neighbor/next-hop independently, it is possible that it switches to the LFA next-hop while IGP is still using the primary next-hop. In order to avoid this situation, it is recommended to enable IGP-LDP synchronization on the LDP interface:

config>**router**>**if**>**ldp-sync-timer** *seconds*

7.10.1.1 Reducing the Scope of the LFA Calculation by SPF

The user can instruct IGP to not include all interfaces participating in a specific IS-IS level or OSPF area in the SPF LFA computation. This provides a way of reducing the LFA SPF calculation where it is not needed.

config>**router**>**isis**>**level**>**loopfree-alternate-exclude**
config>**router**>**ospf**>**area**>**loopfree-alternate-exclude**

If IGP shortcut are also enabled in LFA SPF, the LSPs with destination address in that IS-IS level or OSPF area are also not included in the LFA SPF calculation.

The user can also exclude a specific IP interface from being included in the LFA SPF computation by IS-IS or OSPF:

config>**router**>**isis**>**interface**> **loopfree-alternate-exclude**
config>**router**>**ospf**>**area**>**interface**> **loopfree-alternate-exclude**

When an interface is excluded from the LFA SPF in IS-IS, it is excluded in both level 1 and level 2. When the user excludes an interface from the LFA SPF in OSPF, it is excluded in all areas. However, the above OSPF command can only be executed under the area in which the specified interface is primary and once enabled, the interface is excluded in that area and in all other areas where the interface is secondary. If the user attempts to apply it to an area where the interface is secondary, the command will fail.

Finally, the user can apply the same above commands for an OSPF instance within a VPRN service:

config>**service**>**vprn**>**ospf**>**area**>**loopfree-alternate-exclude**
config>**service**>**vprn**>**ospf**>**area**>**interface**>**loopfree-alternate-exclude**

7.10.2 LDP FRR Procedures

The LDP FEC resolution when LDP FRR is not enabled operates as follows. When LDP receives a *FEC, label* binding for a prefix, it will resolve it by checking if the exact prefix, or a longest match prefix when the **aggregate-prefix-match option** is enabled in LDP, exists in the routing table and is resolved against a next-hop which is an address belonging to the LDP peer which advertised the binding, as identified by its LSR-id. When the next-hop is no longer available, LDP de-activates the FEC and de-programs the NHLFE in the data path. LDP will also immediately withdraw the labels it advertised for this FEC and deletes the ILM in the data path unless the user configured the **label-withdrawal-delay** option to delay this operation. Traffic that is received while the ILM is still in the data path is dropped. When routing computes and populates the routing table with a new next-hop for the prefix, LDP resolves again the FEC and programs the data path accordingly.

When LDP FRR is enabled and an LFA backup next-hop exists for the FEC prefix in RTM, or for the longest prefix the FEC prefix matches to when **aggregate-prefix-match** option is enabled in LDP, LDP will resolve the FEC as above but will program the data path with both a primary NHLFE and a backup NHLFE for each next-hop of the FEC.

In order perform a switchover to the backup NHLFE in the fast path, LDP follows the uniform FRR failover procedures which are also supported with RSVP FRR.

When any of the following events occurs, LDP instructs in the fast path the IOM on the line cards to enable the backup NHLFE for each FEC next-hop impacted by this event. The IOM line cards do that by simply flipping a single state bit associated with the failed interface or neighbor/next-hop:

1. An LDP interface goes operationally down, or is admin shutdown. In this case, LDP sends a neighbor/next-hop down message to the IOM line cards for each LDP peer it has adjacency with over this interface.
2. An LDP session to a peer went down as the result of the Hello or Keep-Alive timer expiring over a specific interface. In this case, LDP sends a neighbor/next-hop down message to the IOM line cards for this LDP peer only.
3. The TCP connection used by a link LDP session to a peer went down, due say to next-hop tracking of the LDP transport address in RTM, which brings down the LDP session. In this case, LDP sends a neighbor/next-hop down message to the IOM line cards for this LDP peer only.
4. A BFD session, enabled on a T-LDP session to a peer, times-out and as a result the link LDP session to the same peer and which uses the same TCP connection as the T-LDP session goes also down. In this case, LDP sends a neighbor/next-hop down message to the IOM line cards for this LDP peer only.

5. A BFD session enabled on the LDP interface to a directly connected peer, times-out and brings down the link LDP session to this peer. In this case, LDP sends a neighbor/next-hop down message to the IOM line cards for this LDP peer only. BFD support on LDP interfaces is a new feature introduced for faster tracking of link LDP peers.

The tunnel-down-dump-time option or the label-withdrawal-delay option, when enabled, does not cause the corresponding timer to be activated for a FEC as long as a backup NHLFE is still available.

7.10.2.1 ECMP Considerations

Whenever the SPF computation determined that there is more than one primary next-hop for a prefix, it will not program any LFA next-hop in RTM. In this case, the LDP FEC will resolve to the multiple primary next-hops, which provides the required protection.

Also, when the system ECMP value is set to **ecmp=1** or to **no ecmp**, which translates to the same and is the default value, SPF can use the overflow ECMP links as LFA next-hops in these two cases.

7.10.2.2 LDP FRR and LDP Shortcut

When LDP FRR is enabled in LDP and the ldp-shortcut option is enabled in the router level, in transit IPv4 packets and specific CPM generated IPv4 control plane packets with a prefix resolving to the LDP shortcut are protected by the backup LDP NHLFE.

7.10.2.3 LDP FRR and LDP-over-RSVP

When LDP-over-RSVP is enabled, the RSVP LSP is modeled as an endpoint, that is, the destination node of the LSP, and not as a link in the IGP SPF. Thus, it is not possible for IGP to compute a primary or alternate next-hop for a prefix which FEC path is tunneled over the RSVP LSP. Only LDP is aware of the FEC tunneling but it cannot determine on its own a loop-free backup path when it resolves the FEC to an RSVP LSP.

As a result, LDP does not activate the LFA next-hop it learned from RTM for a FEC prefix when the FEC is resolved to an RSVP LSP. LDP will activate the LFA next-hop as soon as the FEC is resolved to direct primary next-hop.

LDP FEC tunneled over an RSVP LSP due to enabling the LDP-over-RSVP feature will thus not support the LDP FRR procedures and will follow the slow path procedure of prior implementation.

When the user enables the **lfa-only** option for an RSVP LSP, as described in Loop-Free Alternate Calculation in the Presence of IGP shortcuts, the LSP will not be used by LDP to tunnel an LDP FEC even when IGP shortcut is disabled but LDP-over-RSVP is enabled in IGP.

7.10.2.4 LDP FRR and RSVP Shortcut (IGP Shortcut)

When an RSVP LSP is used as a shortcut by IGP, it is included by SPF as a P2P link and can also be optionally advertised into the rest of the network by IGP. Thus the SPF is able of using a tunneled next-hop as the primary next-hop for a given prefix. LDP is also able of resolving a FEC to a tunneled next-hop when the IGP shortcut feature is enabled.

When both IGP shortcut and LFA are enabled in IS-IS or OSPF, and LDP FRR is also enabled, then the following additional LDP FRR capabilities are supported:

1. A FEC which is resolved to a direct primary next-hop can be backed up by a LFA tunneled next-hop.
2. A FEC which is resolved to a tunneled primary next-hop will not have an LFA next-hop. It will rely on RSVP FRR for protection.

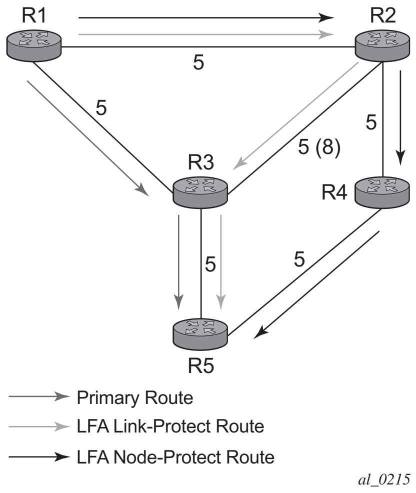
The LFA SPF is extended to use IGP shortcuts as LFA next-hops as explained in Loop-Free Alternate Calculation in the Presence of IGP shortcuts.

7.10.3 IS-IS and OSPF Support for Loop-Free Alternate Calculation

SPF computation in IS-IS and OSPF is enhanced to compute LFA alternate routes for each learned prefix and populate it in RTM.

Figure 76 illustrates a simple network topology with point-to-point (P2P) interfaces and highlights three routes to reach router R5 from router R1.

Figure 76 **Topology with Primary and LFA Routes**

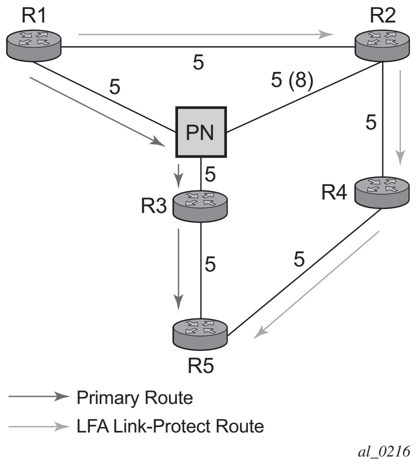


The primary route is by way of R3. The LFA route by way of R2 has two equal cost paths to reach R5. The path by way of R3 protects against failure of link R1-R3. This route is computed by R1 by checking that the cost for R2 to reach R5 by way of R3 is lower than the cost by way of routes R1 and R3. This condition is referred to as the *loop-free criterion*. R2 must be loop-free with respect to source node R1.

The path by way of R2 and R4 can be used to protect against the failure of router R3. However, with the link R2-R3 metric set to 5, R2 sees the same cost to forward a packet to R5 by way of R3 and R4. Thus R1 cannot guarantee that enabling the LFA next-hop R2 will protect against R3 node failure. This means that the LFA next-hop R2 provides link-protection only for prefix R5. If the metric of link R2-R3 is changed to 8, then the LFA next-hop R2 provides node protection since a packet to R5 will always go over R4. In other words it is required that R2 becomes loop-free with respect to both the source node R1 and the protected node R3.

Consider the case where the primary next-hop uses a broadcast interface as illustrated in Figure 77.

Figure 77 **Example Topology with Broadcast Interfaces**



In order for next-hop R2 to be a link-protect LFA for route R5 from R1, it must be loop-free with respect to the R1-R3 link's Pseudo-Node (PN). However, since R2 has also a link to that PN, its cost to reach R5 by way of the PN or router R4 are the same. Thus R1 cannot guarantee that enabling the LFA next-hop R2 will protect against a failure impacting link R1-PN since this may cause the entire subnet represented by the PN to go down. If the metric of link R2-PN is changed to 8, then R2 next-hop will be an LFA providing link protection.

The following are the detailed rules for this criterion as provided in RFC 5286:

- **Rule 1**: Link-protect LFA backup next-hop (primary next-hop R1-R3 is a P2P interface):
  Distance_opt(R2, R5) < Distance_opt(R2, R1) + Distance_opt(R1, R5)
  and,
  Distance_opt(R2, R5) ≥ Distance_opt(R2, R3) + Distance_opt(R3, R5)
- **Rule 2**: Node-protect LFA backup next-hop (primary next-hop R1-R3 is a P2P interface):
  Distance_opt(R2, R5) < Distance_opt(R2, R1) + Distance_opt(R1, R5)
  and,
  Distance_opt(R2, R5) < Distance_opt(R2, R3) + Distance_opt(R3, R5)
- **Rule 3**: Link-protect LFA backup next-hop (primary next-hop R1-R3 is a broadcast interface):
  Distance_opt(R2, R5) < Distance_opt(R2, R1) + Distance_opt(R1, R5)
  and,
  Distance_opt(R2, R5) < Distance_opt(R2, PN) + Distance_opt(PN, R5)
  where; PN stands for the R1-R3 link Pseudo-Node.

For the case of P2P interface, if SPF finds multiple LFA next-hops for a given primary next-hop, it follows the following selection algorithm:

1. It will pick the node-protect type in favor of the link-protect type.
2. If there is more than one LFA next-hop within the selected type, then it will pick one based on the least cost.
3. If more than one LFA next-hop with the same cost results from Step B, then SPF will select the first one. This is not a deterministic selection and will vary following each SPF calculation.

For the case of a broadcast interface, a node-protect LFA is not necessarily a link protect LFA if the path to the LFA next-hop goes over the same PN as the primary next-hop. Similarly, a link protect LFA may not guarantee link protection if it goes over the same PN as the primary next-hop.

The selection algorithm when SPF finds multiple LFA next-hops for a given primary next-hop is modified as follows:

1. The algorithm splits the LFA next-hops into two sets:
   - The first set consists of LFA next-hops which *do not* go over the PN used by primary next-hop.
   - The second set consists of LFA next-hops which *do* go over the PN used by the primary next-hop.
2. If there is more than one LFA next-hop in the first set, it will pick the node-protect type in favor of the link-protect type.
3. If there is more than one LFA next-hop within the selected type, then it will pick one based on the least cost.
4. If more than one LFA next-hop with equal cost results from Step C, SPF will select the first one from the remaining set. This is not a deterministic selection and will vary following each SPF calculation.
5. If no LFA next-hop results from Step D, SPF will rerun Steps B-D using the second set.

This algorithm is more flexible than strictly applying Rule 3 above; the link protect rule in the presence of a PN and specified in RFC 5286. A node-protect LFA which does not avoid the PN; does not guarantee link protection, can still be selected as a last resort. The same thing, a link-protect LFA which does not avoid the PN may still be selected as a last resort. Both the computed primary next-hop and LFA next-hop for a given prefix are programmed into RTM.

7.10.3.1 Loop-Free Alternate Calculation in the Presence of IGP shortcuts

In order to expand the coverage of the LFA backup protection in a network, RSVP LSP based IGP shortcuts can be placed selectively in parts of the network and be used as an LFA backup next-hop.

When IGP shortcut is enabled in IS-IS or OSPF on a given node, all RSVP LSP originating on this node and with a destination address matching the router-id of any other node in the network are included in the main SPF by default.

In order to limit the time it takes to compute the LFA SPF, the user must explicitly enable the use of an IGP shortcut as LFA backup next-hop using one of a couple of new optional argument for the existing LSP level IGP shortcut command:

config>**router**>**mpls**>**lsp**>**igp-shortcut** [**lfa-protect** | **lfa-only**]

The **lfa-protect** option allows an LSP to be included in both the main SPF and the LFA SPFs. For a given prefix, the LSP can be used either as a primary next-hop or as an LFA next-hop but not both. If the main SPF computation selected a tunneled primary next-hop for a prefix, the LFA SPF will not select an LFA next-hop for this prefix and the protection of this prefix will rely on the RSVP LSP FRR protection. If the main SPF computation selected a direct primary next-hop, then the LFA SPF will select an LFA next-hop for this prefix but will prefer a direct LFA next-hop over a tunneled LFA next-hop.

The **lfa-only** option allows an LSP to be included in the LFA SPFs only such that the introduction of IGP shortcuts does not impact the main SPF decision. For a given prefix, the main SPF always selects a direct primary next-hop. The LFA SPF will select a an LFA next-hop for this prefix but will prefer a direct LFA next-hop over a tunneled LFA next-hop.

Thus the selection algorithm when SPF finds multiple LFA next-hops for a given primary next-hop is modified as follows:

1. The algorithm splits the LFA next-hops into two sets:
   - the first set consists of direct LFA next-hops
   - the second set consists of tunneled LFA next-hops. after excluding the LSPs which use the same outgoing interface as the primary next-hop.
2. The algorithms continues with first set if not empty, otherwise it continues with second set.
3. If the second set is used, the algorithm selects the tunneled LFA next-hop which endpoint corresponds to the node advertising the prefix.

- If more than one tunneled next-hop exists, it selects the one with the lowest LSP metric.
- If still more than one tunneled next-hop exists, it selects the one with the lowest tunnel-id.
- If none is available, it continues with rest of the tunneled LFAs in second set.

4. Within the selected set, the algorithm splits the LFA next-hops into two sets:
    - The first set consists of LFA next-hops which do not go over the PN used by primary next-hop.
    - The second set consists of LFA next-hops which go over the PN used by the primary next-hop.
5. If there is more than one LFA next-hop in the selected set, it will pick the node-protect type in favor of the link-protect type.
6. If there is more than one LFA next-hop within the selected type, then it will pick one based on the least total cost for the prefix. For a tunneled next-hop, it means the LSP metric plus the cost of the LSP endpoint to the destination of the prefix.
7. If there is more than one LFA next-hop within the selected type (ecmp-case) in the first set, it will select the first direct next-hop from the remaining set. This is not a deterministic selection and will vary following each SPF calculation.
8. If there is more than one LFA next-hop within the selected type (ecmp-case) in the second set, it will pick the tunneled next-hop with the lowest cost from the endpoint of the LSP to the destination prefix. If there remains more than one, it will pick the tunneled next-hop with the lowest tunnel-id.

7.10.3.2 Loop-Free Alternate Calculation for Inter-Area/inter-Level Prefixes

When SPF resolves OSPF inter-area prefixes or IS-IS inter-level prefixes, it will compute an LFA backup next-hop to the same exit area/border router as used by the primary next-hop.

7.10.3.3 Loop-Free Alternate Shortest Path First (LFA SPF) Policies

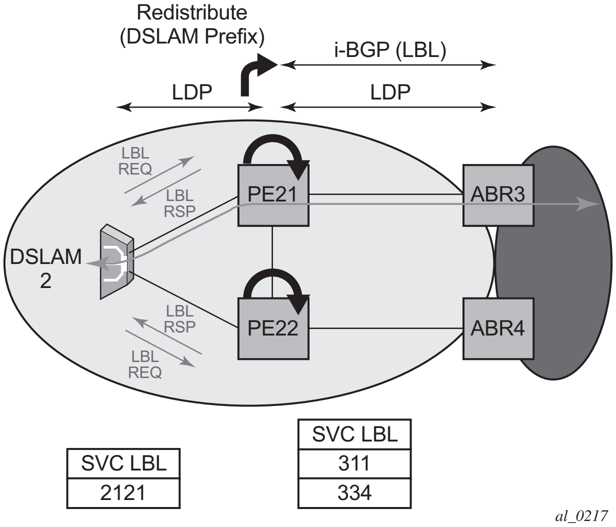
An LFA SPF policy allows the user to apply specific criteria, such as admin group and SRLG constraints, to the selection of a LFA backup next-hop for a subset of prefixes that resolve to a specific primary next-hop. See more details in the section titled “*Loop-Free Alternate Shortest Path First (LFA SPF) Policies*” in the *Routing Protocols Guide*.

7.11 LDP FEC to BGP Label Route Stitching

The stitching of an LDP FEC to a BGP labeled route allows the LDP capable PE devices to offer services to PE routers 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. Figure 78 shows a network with a core area and regional areas.

Figure 78 **Application of LDP to BGP FEC Stitching**



Specific /32 routes in a regional area are not redistributed into the core area. Therefore, only nodes within a regional area and the ABR nodes in the same area exchange LDP FECs. A PE router, for example, PE21, in a regional area learns the reachability of PE routers in other regional areas by way of RFC 3107 BGP labeled routes redistributed by the remote ABR nodes by way of the core area. The remote ABR then sets the next-hop self on the labeled routes before re-distributing them into the core area. The local ABR for PE2, for example, ABR3 may or may not set next-hop self when it re-distributes these labeled BGP routes from the core area to the local regional area.

When forwarding a service packet to the remote PE, PE21 inserts a VC label, the BGP route label to reach the remote PE, and an LDP label to reach either ABR3, if ABR3 sets next-hop self, or ABR1.

In the same network, an MPLS capable DSLAM also act as PE router for VLL services and will need to establish a PW to a PE in a different regional area by way of router PE21, acting now as an LSR. To achieve that, PE21 is required to perform the following operations:

- Translate the LDP FEC it learned from the DSLAM into a BGP labeled route and re-distribute it by way of Interior Border Gateway Protocol (IBGP) within its area. This is in addition to redistributing the FEC to its LDP neighbors in the same area.
- Translate the BGP labeled routes it learns through IBGP into an LDP FEC and re-distribute it to its LDP neighbors in the same area. In the application in Figure 78, the DSLAM requests the LDP FEC of the remote PE router using LDP Downstream on Demand (DoD).
- When a packet is received from the DSLAM, PE21 swaps the LDP label into a BGP label and pushes the LDP label to reach ABR3 or ABR1. When a packet is received from ABR3, the top label is removed and the BGP label is swapped for the LDP label corresponding to the DSLAM FEC.

7.11.1 Configuration

You enable stitching of routes between the LDP and BGP by configuring separately tunnel table route export policies in both protocols and enabling the advertising of RFC 3107 formatted labeled routes for prefixes learned from LDP FECs.

The route export policy in BGP instructs BGP to listen to LDP route entries in the CPM tunnel table. If a /32 LDP FEC prefix matches an entry in the export policy, BGP originates a BGP labeled route, stitches it to the LDP FEC, and re-distributes the BGP labeled route to its IBGP neighbors.

The user adds LDP FEC prefixes with the statement ‘from protocol ldp’ in the configuration of the existing BGP export policy at the global level, the peer-group level, or at the peer level using the commands:

- **config**>**router**>**bgp**>**export** *policy-name*
- **config**>**router**>**bgp**>**group**>**export** *policy-name*
- **config**>**router**>**bgp**>**group**>**neighbor**>**export** *policy-name*

To indicate to BGP to evaluate the entries with the ‘from protocol ldp’ statement in the export policy when applied to a specific BGP neighbor, use commands:

- **config**>**router**>**bgp**>**group**>**neighbor**>**family label-ipv4**
- **config**>**router**>**bgp**>**group**>**neighbor**>**advertise-ldp-prefix**

Without this, only core IPv4 routes learned from RTM are advertised as BGP labeled routes to this neighbor. And the stitching of LDP FEC to the BGP labeled route is not performed for this neighbor even if the same prefix was learned from LDP.

The tunnel table route export policy in LDP instructs LDP to listen to BGP route entries in the CPM Tunnel Table. If a /32 BGP labeled route matches a prefix entry in the export policy, LDP originates an LDP FEC for the prefix, stitches it to the BGP labeled route, and re-distributes the LDP FEC its IBGP neighbors.

The user adds BGP labeled route prefixes with the statement ‘from protocol bgp’ in the configuration of a new LDP tunnel table export policy using the command:

config>router>ldp>export-tunnel-table *policy-name*.

The ‘from protocol’ statement has an effect only when the protocol value is **ldp**. Policy entries with protocol values of **rsvp**, **bgp**, or any value other than **ldp** are ignored at the time the policy is applied to LDP.

7.11.2 Detailed LDP FEC Resolution

When an LSR receives a FEC-label binding from an LDP neighbor for a given specific FEC1 element, the following procedures are performed.

1. LDP installs the FEC if:
   - It was able to perform a successful exact match or a longest match, if aggregate-prefix-match option is enabled in LDP, of the FEC /32 prefix with a prefix entry in the routing table.
   - The advertising LDP neighbor is the next-hop to reach the FEC prefix.
2. When such a FEC-label binding has been installed in the LDP FIB, LDP will perform the following:
   - Program a push and a swap NHLFE entries in the egress data path to forward packets to FEC1.
   - Program the CPM tunnel table with a tunnel entry for the NHLFE.
   - Advertise a new FEC-label binding for FEC1 to all its LDP neighbors according to the global and per-peer LDP prefix export policies.
   - Install the ILM entry pointing to the swap NHLFE.
3. When BGP learns the LDP FEC by way of the CPM tunnel table and the FEC prefix exists in the BGP route export policy, it will perform the following:
   - Originate a labeled BGP route for the same prefix with this node as the next-hop and advertise it by way of IBGP to its BGP neighbors, for example, the local ABR/ASBR nodes, which have the **advertise-ldp-prefix** enabled.

- Install the ILM entry pointing to the swap NHLFE programmed by LDP.

7.11.3 Detailed BGP Labeled Route Resolution

When an LSR receives a BGP labeled route by way of IBGP for a given specific /32 prefix, the following procedures are performed.

1. BGP resolves and installs the route in BGP if:
    - An LDP LSP to the BGP neighbor exists, for example, the ABR or ASBR, which advertised it and which is the next-hop of the BGP labeled route.
2. When the BGP route is installed, BGP programs the following:
    - Push NHLFE in the egress data path to forward packets to this BGP labeled route.
    - The CPM tunnel table with a tunnel entry for the NHLFE.
3. When LDP learns the BGP labeled route from the CPM tunnel table and the prefix exists in the new LDP tunnel table route export policy, it does the following:
    - Advertise a new LDP FEC-label binding for the same prefix to its LDP neighbors according the global and per-peer LDP export prefix policies. If LDP already advertised a FEC for the same /32 prefix after receiving it from an LDP neighbor then no action is required. For LDP neighbors that negotiated LDP Downstream on Demand (DoD), the FEC is advertised only when this node receives a Label Request message for this FEC from its neighbor.
    - Install the ILM entry pointing to the BGP NHLFE if a new LDP FEC-label binding is advertised. If an ILM entry exists and points to an LDP NHLFE for the same prefix then no update to the ILM entry is performed. The LDP route is always preferred over the BGP labeled route.

7.11.4 Data Plane Forwarding

When a packet is received from an LDP neighbor, the LSR swaps the LDP label into a BGP label and pushes the LDP label to reach the BGP neighbor, for example, ABR/ASBR, which advertised the BGP labeled route with itself as the next-hop.

When a packet is received from a BGP neighbor such as an ABR/ASBR, the top label is removed and the BGP label is swapped for the LDP label to reach the next-hop for the prefix.

7.12 LDP-SR Stitching for IPv4 prefixes

This feature enables stitching between an LDP FEC and SR node-SID route for the same IPv4 /32prefix.

7.12.1 LDP-SR Stitching Configuration

The user enables the stitching between an LDP FEC and SR node-SID route for the same prefix by configuring the export of SR (LDP) tunnels from the CPM Tunnel Table Manager (TTM) into LDP (IGP).

In the LDP-to-SR data path direction, the existing tunnel table route export policy in LDP, which was introduced for LDP-BGP stitching, is enhanced to support the export of SR tunnels from the TTM to LDP. The user adds the **config**>**router**>p**olicy-options**>**policy-statement**>**entry**>**from protocol isis** [**instance** *instance-id*] or **protocol ospf** [**instance** *instance-id*] configuration to the LDP tunnel table export policy using the following command:

config>**router**>**ldp**>**export-tunnel-table** *policy-name*

The user can restrict the export to LDP of SR tunnels from a specific prefix list. The user can also restrict the export to a specific IGP instance by optionally specifying the instance ID in the **from** statement.

The **from protocol** statement has an effect only when the protocol value is **isis**, **ospf**, or **bgp**. Policy entries configured with any other value are ignored when the policy is applied. If the user configures multiple **from** statements in the same policy or does not include the **from** statement but adds a default **accept** action, then LDP will follow the TTM selection rules as described in the “Segment Routing Tunnel Management” in the *7450 ESS, 7750 SR, 7950 XRS, and VSR Unicast Routing Protocols Guide* to select a tunnel to which it will stitch the LDP ILM to:

- LDP selects the tunnel from the lowest TTM preference protocol.
- If IS-IS and BGP protocols have the same preference, then LDP uses the default TTM protocol preference to select the protocol.
- Within the same IGP protocol, LDP selects the lowest instance ID.

When this policy is enabled in LDP, LDP listens to SR tunnel entries in the TTM. If an LDP FEC primary next hop cannot be resolved using an RTM route and a SR tunnel of type SR-ISIS or SR-OSPF to the same destination exists in TTM, LDP programs an LDP ILM and stitches it to the SR node-SID tunnel endpoint. LDP also originates an FEC for the prefix and re-distributes it to its LDP and T-LDP peers. The latter allows an LDP FEC that is tunneled over a RSVP-TE LSP to have its ILM stitched to an SR tunnel endpoint. When a LDP FEC is stitched to a SR tunnel, forwarded packets benefit from the protection provided by the LFA/remote LFA backup next-hop of the SR tunnel.

When resolving a FEC, LDP prefers the RTM over the TTM when both resolutions are possible. That is, swapping the LDP ILM to a LDP NHLFE is preferred over stitching the LDP ILM to an SR tunnel endpoint.

In the SR-to-LDP data path direction, the SR mapping server provides a global policy for prefixes corresponding to the LDP FECs the SR needs to stitch to. See “Segment Routing Mapping Server” in the *7450 ESS, 7750 SR, 7950 XRS, and VSR Unicast Routing Protocols Guide* for more information. As a result, a tunnel table export policy is not required. Instead, you can export to an IGP instance the LDP tunnels for FEC prefixes advertised by the mapping server using the following commands:

- **config**>**router**>**isis**>**segment-routing**>**export-tunnel-table ldp**
- **config**>**router**>**ospf**>**segment-routing**>**export-tunnel-table ldp**

When this command is enabled in the segment-routing context of an IGP instance, IGP listens to LDP tunnel entries in the TTM. When a /32 LDP tunnel destination matches a prefix for which IGP has received a prefix-SID sub-TLV from a mapping server, IGP instructs the SR module to program the SR ILM and stitch it to the LDP tunnel endpoint. The SR ILM can stitch to an LDP FEC resolved over either link LDP or T-LDP. In the latter case, the stitching is performed to an LDP-over-RSVP tunnel. When an SR tunnel is stitched to an LDP FEC, packets forwarded will benefit from the FRR protection of the LFA backup next-hop of the LDP FEC.

When resolving a node SID, IGP will prefer a prefix SID received in an IP Reach TLV over a prefix SID received via the mapping server. That is, swapping the SR ILM to a SR NHLFE is preferred over stitching it to a LDP tunnel endpoint. Refer to the “Segment Routing Mapping Server Prefix SID Resolution” in the *7450 ESS, 7750 SR, 7950 XRS, and VSR Unicast Routing Protocols Guide* for more information about prefix SID resolution.

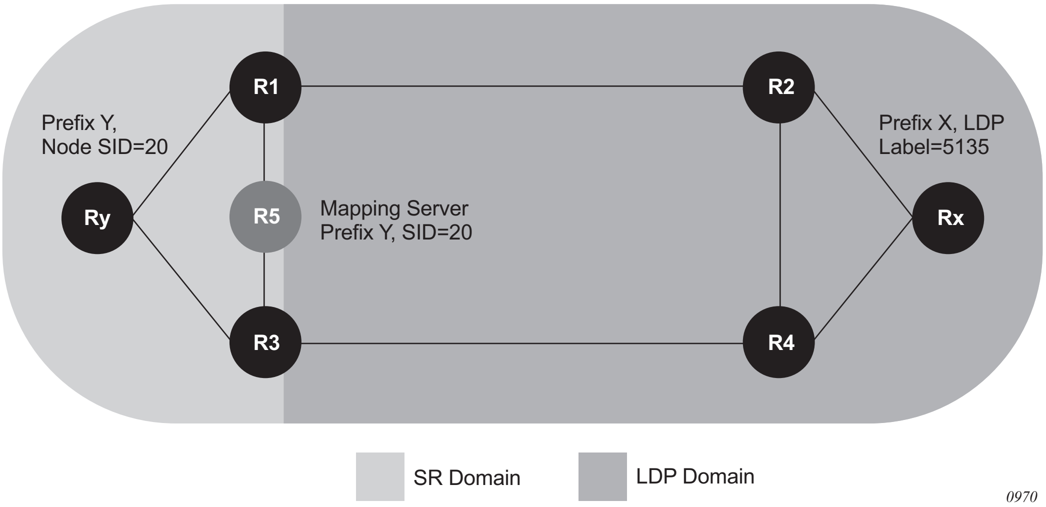
It is recommended to enable the **bfd-enable** option on the interfaces in both LDP and IGP instance contexts to speed up the failure detection and the activation of the LFA/remote-LFA backup next-hop in either direction. This is particularly true if the injected failure is a remote failure.

This feature is limited to IPv4 /32 prefixes in both LDP and SR.

7.12.2 Stitching in the LDP-to-SR Direction

Stitching in data-plane from the LDP-to-SR direction is based on the LDP module monitoring the TTM for a SR tunnel of a prefix matching an entry in the LDP TTM export policy.

Figure 79 **Stitching in the LDP-to-SR Direction**



In Figure 79, the boundary router R1 performs the following procedure to effect stitching:

Step 1. Router R1 is at the boundary between an SR domain and LDP domain and is configured to stitch between SR and LDP.

Step 2. Link R1-R2 is LDP-enabled, but router R2 does not support SR (or SR is disabled).

Step 3. Router R1 receives a prefix-SID sub-TLV in an IS-IS IP reachability TLV originated by router Ry for prefix Y.

Step 4. R1 resolves the prefix-SID and programs an NHLFE on the link towards the next-hop in the SR domain. R1 programs an SR ILM and points it to this NHLFE.

Step 5. Because R1 is programmed to stitch LDP to SR, the LDP in R1 discovers in TTM the SR tunnel to Y. LDP programs an LDP ILM and points it to the SR tunnel. As a result, both the SR ILM and LDP ILM now point to the SR tunnel, one via the SR NHLFE and the other via the SR tunnel endpoint.

Step 6. R1 advertises the LDP FEC for the prefix Y to all its LDP peers. R2 is now able to install a LDP tunnel towards Ry.

Step 7. If R1 finds multiple SR tunnels to destination prefix Y, it uses the following steps of the TTM tunnel selection rules to select the SR tunnel.

i. R1 selects the tunnel from the lowest preference IGP protocol.

ii. Select the protocol using the default TTM protocol preference.

iii. Within the same IGP protocol, R1 uses the lowest instance ID to select the tunnel.

Step 8. If the user concurrently configured the **from protocol ospf**, **from protocol isis**, and **from protocol bgp** statements in the same LDP tunnel table export policy, or did not include the from statement but added a default action of accept, R1 will select the tunnel to destination prefix Y to stitch the LDP ILM to using the TTM tunnel selection rules:

i. R1 selects the tunnel from the lowest preference protocol.

ii. If any two or all of IS-IS, OSPF, and BGP protocols have the same preference, then R1 selects the protocol using the default TTM protocol preference.

iii. Within the same IGP protocol, R1 uses the lowest instance ID to select the tunnel.

Note: If R1 has already resolved a LDP FEC for prefix Y, it has an ILM for it, but this ILM is not be updated to point towards the SR tunnel. This is because LDP resolves in RTM first before going to TTM and, therefore, prefers the LDP tunnel over the SR tunnel. Similarly, if an LDP FEC is received after the stitching is programmed, the LDP ILM is updated to point to the LDP NHLFE because LDP can resolve the LDP FEC in RTM.

Step 9. The user enables SR in R2. R2 resolves the prefix SID for Y and installs the SR ILM and the SR NHLFE. R2 is now able of forwarding packets over the SR tunnel to router Ry. No processing occurs in R1 because the SR ILM is already programmed.

Step 10. The user disables LDP on the interface R1-R2 (both directions) and the LDP FEC ILM and NHLFE are removed in R1. The same occurs in R2 which can then only forward using the SR tunnel towards Ry.

7.12.3 Stitching in the SR-to-LDP Direction

The stitching in data-plane from the SR-to-LDP direction is based on the IGP monitoring the TTM for a LDP tunnel of a prefix matching an entry in the SR TTM export policy.

In Figure 79, the boundary router R1 performs the following procedure to effect stitching:

Step 1. Router R1 is at the boundary between a SR domain and a LDP domain and is configured to stitch between SR and LDP.

Link R1-R2 is LDP enabled but router R2 does not support SR (or SR is disabled).

Step 2. R1 receives an LDP FEC for prefix X owned by router Rx further down in the LDP domain.

RTM in R1 shows that the interface to R2 is the next-hop for prefix X.

Step 3. LDP in R1 resolves this FEC in RTM and creates an LDP ILM for it with, for example, ingress label L1, and points it to an LDP NHLFE towards R2 with egress label L2.

Step 4. Later on, R1 receives a prefix-SID sub-TLV from the mapping server R5 for prefix X.

Step 5. IGP in R1 is resolving in its routing table the next-hop of prefix X to the interface to R2. R1 knows that R2 did not advertise support of Segment Routing and, thus, SID resolution for prefix X in routing table fails.

Step 6. IGP in R1 attempts to resolve prefix SID of X in TTM because it is configured to stitch SR-to-LDP. R1 finds a LDP tunnel to X in TTM, instructs the SR module to program a SR ILM with ingress label L3, and points it to the LDP tunnel endpoint, consequently stitching ingress L3 label to egress L2 label.

Note:

- Here, two ILMs, the LDP and SR, are pointing to the same LDP tunnel one via NHLFE and one via tunnel endpoint.
- No SR tunnel to destination X should be programmed in TTM following this resolution step.
- A trap will be generated for prefix SID resolution failure only after IGP fails to complete Step 5 and Step 6. The existing trap for prefix SID resolution failure is enhanced to state whether the prefix SID which failed resolution was part of mapping server TLV or a prefix TLV.

Step 7. The user enables segment routing on R2.

Step 8. IGP in R1 discovers that R2 supports SR via the SR capability.

Because R1 still has a prefix-SID for X from the mapping server R5, it maintains the stitching of the SR ILM for X to the LDP FEC unchanged.

Step 9. The operator disables the LDP interface between R1 and R2 (both directions) and the LDP FEC ILM and NHLFE for prefix X are removed in R1.

Step 10. This triggers the re-evaluation of the SIDs. R1 first attempts the resolution in routing table and since the next-hop for X now supports SR, IGP instructs the SR module to program a NHLFE for prefix-SID of X with egress label L4 and outgoing interface to R2. R1 installs a SR tunnel in TTM for destination X. R1 also changes the SR ILM with ingress label L3 to point to the SR NHLFE with egress label L4.

Router R2 now becomes the SR-LDP stitching router.

Step 11. Later, router Rx, which owns prefix X, is upgraded to support SR. R1 now receives a prefix-SID sub-TLV in a ISIS or OSPF prefix TLV originated by Rx for prefix X. The SID information may or may not be the same as the one received from the mapping server R5. In this case, IGP in R1 will prefer the prefix-SID originated by Rx and update the SR ILM and NHLFE with appropriate labels.

Step 12. Finally, the operator cleans up the mapping server and removes the mapping entry for prefix X, which then gets withdrawn by IS-IS.

7.13 LDP FRR LFA Backup using SR Tunnel for IPv4 Prefixes

The user enables the use of SR tunnel as a remote LFA or as a TI-LFA backup tunnel next hop by an LDP FEC via the following CLI command:

CLI Syntax:
```
config>router>ldp>fast-reroute [backup-sr-tunnel]
```

As a pre-requisite, the user must enable the stitching of LDP and SR in the LDP-to-SR direction as explained in LDP-SR Stitching Configuration. That is because the LSR must perform the stitching of the LDP ILM to SR tunnel when the primary LDP next-hop of the FEC fails. Thus, LDP must listen to SR tunnels programmed by the IGP in TTM, but the mapping server feature is not required.

Assume the **backup-sr-tunnel** option is enabled in LDP and the {**loopfree-alternates remote-lfa**} option and/or the {**loopfree-alternates ti-lfa**} option is enabled in the IGP instance, and that LDP was able to resolve the primary next hop of the LDP FEC in RTM. IGP SPF will run both the base LFA and the TI-LFA algorithms and if it does not find a backup next hop for a prefix of an LDP FEC, it will also run the remote LFA algorithm. If IGP finds a TI-LFA or a remote LFA tunnel next hop, LDP programs the primary next hop of the FEC using an LDP NHLFE and programs the LFA backup next hop using an LDP NHLFE pointing to the SR tunnel endpoint.

Note: The LDP packet is not “tunneled” over the SR tunnel. The LDP label is actually stitched to the segment routing label stack. LDP points both the LDP ILM and the LTN to the backup LDP NHLFE which itself uses the SR tunnel endpoint.

The behavior of the feature is similar to the LDP-to-SR stitching feature described in the LDP-SR Stitching for IPv4 prefixes section, except the behavior is augmented to allow the stitching of an LDP ILM/LTN to an SR tunnel for the LDP FEC backup NHLFE when the primary LDP NHLFE fails.

The following is the behavior of this feature:

- When LDP resolves a primary next hop in RTM and a TI-LFA or a remote LFA backup next hop using SR tunnel in TTM, LDP programs a primary LDP NHLFE as usual and a backup LDP NHLFE pointing to the SR tunnel, which has the TI-LFA or remote LFA backup for the same prefix.
- If the LDP FEC primary next hop failed and LDP has pre-programmed a TI-LFA or a remote LFA next hop with an LDP backup NHLFE pointing to the SR tunnel, the LDP ILM/LTN switches to it.

Note: If, for some reason, the failure impacted only the LDP tunnel primary next hop but not the SR tunnel primary next hop, the LDP backup NHLFE will effectively point to the primary next hop of the SR tunnel and traffic of the LDP ILM/LTN will follow this path instead of the TI-LFA or remote LFA next hop of the SR tunnel until the latter is activated.

- If the LDP FEC primary next hop becomes unresolved in RTM, LDP switches the resolution to a SR tunnel in TTM, if one exists, as per the LDP-to-SR stitching procedures described in Stitching in the LDP-to-SR Direction.
- If both the LDP primary next hop and a regular LFA next hop become resolved in RTM, the LDP FEC programs the primary and backup NHLFEs as usual.
- It is recommended to enable the **bfd-enable** option on the interfaces in both LDP and IGP instance contexts to speed up the failure detection and the activation of the LFA/TI-LFA/remote-LFA backup next hop in either direction.

7.14 LDP Remote LFA

LDP Remote LFA (rLFA) builds on the pre-existing capability to compute repair paths to a remote LFA node (or PQ node), which puts the packets onto the shortest path without looping them back to the node that forwarded them over the repair tunnel. Refer to *7450 ESS, 7750 SR, 7950 XRS, and VSR Unicast Routing Protocols Guide* section *Remote LFA with Segment Routing* for further information about rLFA computation. In SR OS, a repair tunnel can also be an SR tunnel, however this section describes an LDP-in-LDP tunnel.

A prerequisite for configuring LDP rLFA is to enable Remote LFA computation using the following command:

config>**router**>**isis**>**loopfree-alternates remote-lfa**

Finally, enable attaching rLFA information to RTM entries using the following command:

config>**router**>**isis**>**loopfree-alternates augment-route-table**

This command attaches rLFA-specific information to route entries that are necessary for LDP to program repair tunnels towards the PQ node using a specific neighbor.

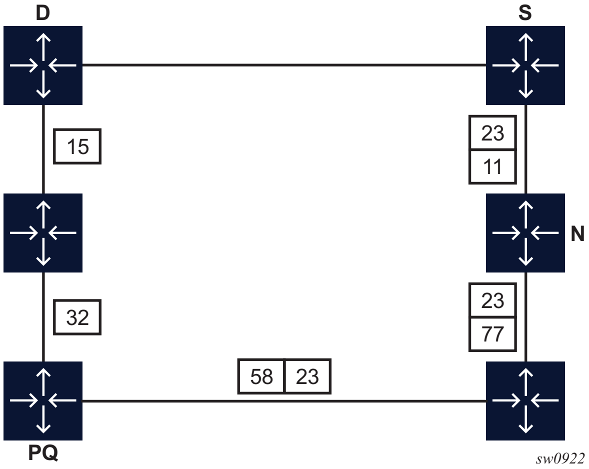
In addition, enable tunneling on both the PQ node and the source node using the following command:

config>**router**>**ldp**>**targ-session**>**peer**>**tunneling**

Note: LDP rLFA is only available with IS-IS.

Figure 80 shows the general principles of LDP rLFA operation.

Figure 80 **General Principles of LDP rLFA Operation**



In Figure 80, S is the source node and D is the destination node. The primary path is the direct link between S and D. The rLFA algorithm has determined the PQ node. In the event of a failure between S and D, for traffic not to loopback to S, the traffic must be sent directly to the PQ node. An LDP targeted session is required between PQ and S. Over that T-LDP session, the PQ node advertises label 23 for FEC D. All other labels are link LDP bindings, which allow traffic to reach the PQ node. On S, LDP creates an NHLFE that has two labels, where label 23 is the inner label. Label 23 is tunneled up to the PQ node, which then forwards traffic on the shortest path to D.

Note: LDP rLFA applies to IPv4 FECs only. LDP rLFA requires the targeted sessions (between Source node and PQ node) to be manually configured beforehand (the system does not automatically set-up T-LDP sessions towards the PQ nodes that the rLFA algorithm has identified). These targeted sessions must be set up with router IDs that match the ones the rLFA algorithm uses. LDP rLFA is designed to be operated in LDP-only environments; as such, LDP does not establish rLFA backups when in the presence of LDP over RSVP-TE or LDP over SR-TE tunnels. OAM (**lsp-trace**) is not supported over the repair tunnels.

7.15 Automatic LDP rLFA

The manual LDP rLFA configuration method requires the user to specify beforehand, on each node, the list of peers with which a targeted session will be established. See LDP Remote LFA for information about the rLFA LDP tunneling technology, and how to configure LDP to establish targeted sessions.

This section describes the automatic LDP rLFA mechanisms used to automatically establish targeted LDP sessions without the need to specify, on each node, the list of peers with which the targeted sessions must be established. The automatic LDP rLFA method considerably minimizes overall configuration, and increases dynamic flexibility.

The basic principles of operation for the automatic LDP rLFA capability are described in LDP Remote LFA. In the example shown in Figure 80, considering a failure on the shortest path between S and D nodes, S needs a targeted LDP session towards the PQ node to learn the label-binding information configured on PQ node for FEC D. As a prerequisite, the LFA algorithm has run successfully and the PQ node information is attached to the route entries used by LDP.

Enable remote LFA computation using the following command:

config>**router**>**isis**>**loopfree-alternates**>**remote-lfa**

Enable attaching rLFA information to RTM entries using the following command:

config>**router**>**isis**>**loopfree-alternates**>**augment-route-table**

In the Figure 80 scenarios, because the S node requires the T-LDP session, it should initiate the T-LDP session request. The PQ node will receive the request for this session. Therefore, S node configuration is as follows:

```
configure router ldp targeted-session auto-tx ipv4
{
      tunneling false
      admin-state enable
}
```

And PQ node configuration is as follows:

```
configure router ldp targeted-session auto-rx ipv4
{
      tunneling true
      admin-state enable
}
```

Based on the preceding configurations, the S node, using the PQ node information attached to the route entries, automatically starts sending LDP targeted Hello messages to the PQ node. The PQ node accepts them and the T-LDP session is established. For the same reason, as in case of manual LDP rLFA, enabling tunneling at the PQ node is required to enable PQ to send to S the label that it is bound to FEC D. In such a simple configuration, if there is a change in both the network topology and the PQ node of S for FEC D, S will automatically kill the session to the previous PQ node and establish a new one (toward the new PQ node).

➡ **Note:** It is not possible to configure parameters specifically for automatic T-LDP sessions. The system inherits parameters, either those defined for the IPv4 family (under targeted-session) or the default parameters of the system. This applies to **hello**, **hello-reduction**, and **keepalive** configurations. Also, the automatic T-LDP session can use parameters defined for the **peer-transport** configuration if the specified address is the router ID of the peer.

In typical network deployments, each node is potentially the source node as well as the PQ node of a source node for a specific destination FEC. Therefore, all nodes may have both **auto-tx** and **auto-rx** configured and enabled. Nodes may also have other configurations defined (for example, peer, peer-template, and so on).

There are several implications (explicit or implicit) of having multiple configurations on a peer (either explicit or implicit).

One implication is that LDP operates using precedence levels. When a targeted session is established with a peer, LDP uses the session parameters with the highest precedence. The order of precedence is as follows (highest to lowest):

- peer
- template
- auto-tx
- auto-rx
- sdp

Let us consider the case where a T-LDP session is needed between nodes A (source) and B (PQ node). If A has **auto-tx** enabled and a per-peer configuration for B also exists, A will establish the session using the parameters defined in the per-peer configuration for B, instead of using those defined under **auto-tx**. The same applies on B. However, if B uses per-peer configuration for A and the chosen configuration does not enable tunneling, LDP rLFA will not work because the PQ node will not tunnel the FEC/label bindings. This mechanism also applies to **auto-tx** and **auto-rx**.

In a typical scenario in which the **auto-tx** and **auto-rx** modes are both enabled on a node that will act as the PQ node, and the node chooses the **auto-tx** configuration for the T-LDP session (because it has the higher precedence than **auto-rx**), LDP rLFA will only work if tunneling is enabled under **auto-tx**. The configuration from which the session parameters are taken is indicated in the **show**>**router**>**ldp**>**targ-peer detail** command (“creator” label).

Another implication is that redundant T-LDP sessions may remain up after a topology change when they are no longer required. The following **clear** command enables the operator to delete these redundant T-LDP sessions.

clear>**router**>**ldp**>**targeted-auto-rx**>**hold-time** *seconds*

The operator must run the command during a specific time window on all nodes on which **auto-rx** is configured. The **hold-time** value should be greater than the hello-timer value plus the time required to run the **clear** command on all applicable nodes. A system check will verify that a non-zero value is configured; no other checks are enforced. It is the responsibility of the operator to ensure that the configured non-zero value is long enough to meet the preceding criterion.

While the hold timer for the **clear** command is in progress, the remaining timeout value can be displayed using the **tools**>**dump**>**router**>**ldp**>**timers** command.

The **clear** command is not synchronized to the standby CPM. If an operator does a clear with a large hold-time value and the CPM does a switchover during this time, the operator needs to restart the clear on the newly active CPM.

Note: The following considerations apply when configuring automatic LDP rLFA:

- works with IS-IS only
- only supports IPv4 FECs
- **local-lsr-id** configuration and templates are not supported
- **lsp-trace** on backup path is not supported

7.16 Automatic Creation of a Targeted Hello Adjacency and LDP Session

This feature enables the automatic creation of a targeted Hello adjacency and LDP session to a discovered peer.

7.16.1 Feature Configuration

The user first creates a targeted LDP session peer parameter template:

config>router>ldp>targ-session>peer-template *template-name*

Inside the template the user configures the common T-LDP session parameters or options shared by all peers using this template. These are the following:

bfd-enable, hello, hello-reduction, keepalive, local-lsr-id, and **tunneling**.

The tunneling option does not support adding explicit RSVP LSP names. LDP will select RSVP LSP for an endpoint in LDP-over-RSVP directly from the Tunnel Table Manager (TTM).

Then the user references the peer prefix list which is defined inside a policy statement defined in the global policy manager.

config>router>ldp>targ-session>peer-template-map peer-template *template-name* **policy** *peer-prefix-policy*

Each application of a targeted session template to a given prefix in the prefix list will result in the establishment of a targeted Hello adjacency to an LDP peer using the template parameters as long as the prefix corresponds to a router-id for a node in the TE database. The targeted Hello adjacency will either trigger a new LDP session or will be associated with an existing LDP session to that peer.

Up to five (5) peer prefix policies can be associated with a single peer template at all times. Also, the user can associate multiple templates with the same or different peer prefix policies. Thus multiple templates can match with a given peer prefix. In all cases, the targeted session parameters applied to a given peer prefix are taken from the first created template by the user. This provides a more deterministic behavior regardless of the order in which the templates are associated with the prefix policies.

Each time the user executes the above command, with the same or different prefix policy associations, or the user changes a prefix policy associated with a targeted peer template, the system re-evaluates the prefix policy. The outcome of the re-evaluation will tell LDP if an existing targeted Hello adjacency needs to be torn down or if an existing targeted Hello adjacency needs to have its parameters updated on the fly.

If a /32 prefix is added to (removed from) or if a prefix range is expanded (shrunk) in a prefix list associated with a targeted peer template, the same prefix policy re-evaluation described above is performed.

The template comes up in the **no shutdown** state and as such it takes effect immediately. Once a template is in use, the user can change any of the parameters on the fly without shutting down the template. In this case, all targeted Hello adjacencies are updated.

7.16.2 Feature Behavior

Whether the prefix list contains one or more specific /32 addresses or a range of addresses, an external trigger is required to indicate to LDP to instantiate a targeted Hello adjacency to a node which address matches an entry in the prefix list. The objective of the feature is to provide an automatic creation of a T-LDP session to the same destination as an auto-created RSVP LSP to achieve automatic tunneling of LDP-over-RSVP. The external trigger is when the router with the matching address appears in the Traffic Engineering database. In the latter case, an external module monitoring the TE database for the peer prefixes provides the trigger to LDP. As a result of this, the user must enable the **traffic-engineering** option in ISIS or OSPF.

Each mapping of a targeted session peer parameter template to a policy prefix which exists in the TE database will result in LDP establishing a targeted Hello adjacency to this peer address using the targeted session parameters configured in the template. This Hello adjacency will then either get associated with an LDP session to the peer if one exists or it will trigger the establishment of a new targeted LDP session to the peer.

The SR OS supports multiple ways of establishing a targeted Hello adjacency to a peer LSR:

- User configuration of the peer with the targeted session parameters inherited from the **config**>**router**>**ldp**>**targ-session**>**ipv4** in the top level context or explicitly configured for this peer in the **config**>**router**>**ldp**>**targ-session**>**peer** context and which overrides the top level parameters shared by all targeted peers. Let us refer to the top level configuration context as the global context. Some parameters only exist in the global context; their value will always be inherited by all targeted peers regardless of which event triggered it.
- User configuration of an SDP of any type to a peer with the **signaling tldp** option enabled (default configuration). In this case the targeted session parameter values are taken from the global context.
- User configuration of a (FEC 129) PW template binding in a BGP-VPLS service. In this case the targeted session parameter values are taken from the global context.
- User configuration of a (FEC 129 type II) PW template binding in a VLL service (dynamic multi-segment PW). In this case the target session parameter values are taken from the global context
- User configuration of a mapping of a targeted session peer parameter template to a prefix policy when the peer address exists in the TE database. In this case, the targeted session parameter values are taken from the template.
- Features using an LDP LSP, which itself is tunneled over an RSVP LSP (LDP-over-RSVP), as a shortcut do not trigger automatically the creation of the targeted Hello adjacency and LDP session to the destination of the RSVP LSP. The user must configure manually the peer parameters or configure a mapping of a targeted session peer parameter template to a prefix policy. These features are:
  - BGP shortcut (**next-hop-resolution shortcut-tunnel** option in BGP),
  - IGP shortcut (**igp-shortcut** option in IGP),
  - LDP shortcut for IGP routes (**ldp-shortcut** option in router level),
  - static route LDP shortcut (**ldp** option in a static route),
  - VPRN service (**auto-bind-tunnel resolution-filter ldp** option), and

Since the above triggering events can occur simultaneously or in any arbitrary order, the LDP code implements a priority handling mechanism in order to decide which event overrides the active targeted session parameters. The overriding trigger will become the owner of the targeted adjacency to a given peer and will be shown in **show router ldp targ-peer**.

Table 34 summarizes the triggering events and the associated priority.

Table 34 **Targeted LDP Adjacency Triggering Events and Priority**

Triggering Event	Automatic Creation of Targeted Hello Adjacency	Active Targeted Adjacency Parameter Override Priority
Manual configuration of peer parameters (creator=manual)	Yes	1
Mapping of targeted session template to prefix policy (creator=template)	Yes	2
Manual configuration of SDP with **signaling tldp** option enabled (creator=service manager)	Yes	3
PW template binding in BGP-AD VPLS (creator=service manager)	Yes	3
PW template binding in FEC 129 VLL (creator=service manager)	Yes	3
LDP-over-RSVP as a BGP/IGP/LDP/Static shortcut	No	—
LDP-over-RSVP in VPRN auto-bind	No	—
LDP-over-RSVP in BGP Label Route resolution	No	—

Any parameter value change to an active targeted Hello adjacency caused by any of the above triggering events is performed by having LDP immediately send a Hello message with the new parameters to the peer without waiting for the next scheduled time for the Hello message. This allows the peer to adjust its local state machine immediately and maintains both the Hello adjacency and the LDP session in UP state. The only exceptions are the following:

- The triggering event caused a change to the **local-lsr-id** parameter value. In this case, the Hello adjacency is brought down which will also cause the LDP session to be brought down if this is the last Hello adjacency associated with the session. A new Hello adjacency and LDP session will then get established to the peer using the new value of the local LSR ID.
- The triggering event caused the targeted peer **shutdown** option to be enabled. In this case, the Hello adjacency is brought down which will also cause the LDP session to be brought down if this is the last Hello adjacency associated with the session.

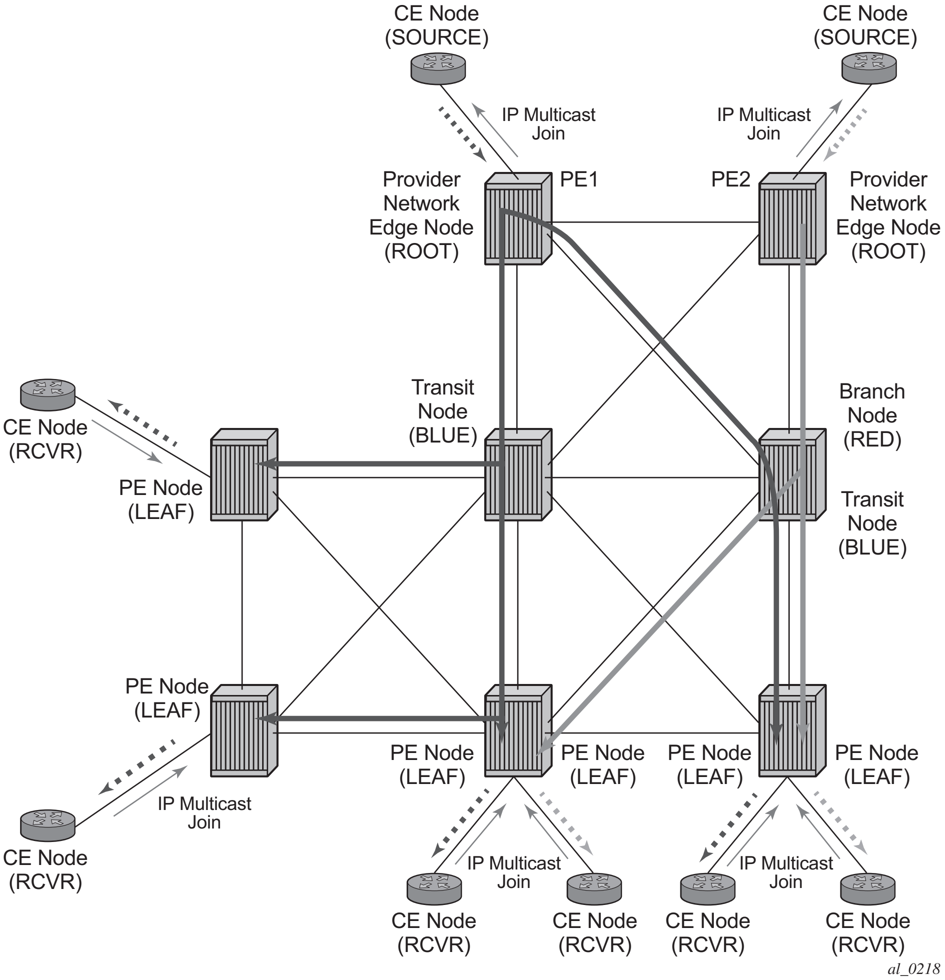
Finally, the value of any LDP parameter which is specific to the LDP/TCP session to a peer is inherited from the **config>router>ldp>session-params>peer** context. This includes MD5 authentication, LDP prefix per-peer policies, label distribution mode (DU or DOD), and so on.

7.17 Multicast P2MP LDP for GRT

The P2MP LDP LSP setup is initiated by each leaf node of multicast tree. A leaf PE node learns to initiate a multicast tree setup from client application and sends a label map upstream towards the root node of the multicast tree. On propagation of label map, intermediate nodes that are common on path for multiple leaf nodes become branch nodes of the tree.

Figure 81 illustrates wholesale video distribution over P2MP LDP LSP. Static IGMP entries on edge are bound to P2MP LDP LSP tunnel-interface for multicast video traffic distribution.

Figure 81 **Video Distribution using P2MP LDP**



7.18 LDP P2MP Support

7.18.1 LDP P2MP Configuration

A node running LDP also supports P2MP LSP setup using LDP. By default, it would advertise the capability to a peer node using P2MP capability TLV in LDP initialization message.

This configuration option per interface is provided to restrict/allow the use of interface in LDP multicast traffic forwarding towards a downstream node. Interface configuration option does not restrict/allow exchange of P2MP FEC by way of established session to the peer on an interface, but it would only restrict/allow use of next-hops over the interface.

7.18.2 LDP P2MP Protocol

Only a single generic identifier range is defined for signaling multipoint tree for all client applications. Implementation on the 7750 SR or 7950 XRS reserves the range (1..8292) of generic LSP P2MP-ID on root node for static P2MP LSP.

7.18.3 Make Before Break (MBB)

When a transit or leaf node detects that the upstream node towards the root node of multicast tree has changed, it follows graceful procedure that allows make-before-break transition to the new upstream node. Make-before-break support is optional. If the new upstream node does not support MBB procedures then the downstream node waits for the configured timer before switching over to the new upstream node.

7.18.4 ECMP Support

If multiple ECMP paths exist between two adjacent nodes on the then the upstream node of the multicast receiver programs all entries in forwarding plane. Only one entry is active based on ECMP hashing algorithm.

7.18.5 Inter-AS Non-segmented mLDP

This feature allows multicast services to use segmented protocols and span them over multiple autonomous systems (ASs), like in unicast services. As IP VPN or GRT services span multiple IGP areas or multiple ASs, either due to a network designed to deal with scale or as result of commercial acquisitions, operators may require inter-AS VPN (unicast) connectivity. For example, an inter-AS VPN can break the IGP, MPLS, and BGP protocols into access segments and core segments, allowing higher scaling of protocols by segmenting them into their own islands. SR OS allows for similar provision of multicast services and for spanning these services over multiple IGP areas or multiple ASs.

mLDP supports non-segmented mLDP trees for inter-AS solutions, applicable for multicast services in the GRT (Global Routing Table) where they need to traverse mLDP point-to-multipoint tunnels as well as NG-MVPN services.

7.18.5.1 In-band Signaling with Non-segmented mLDP Trees in GRT

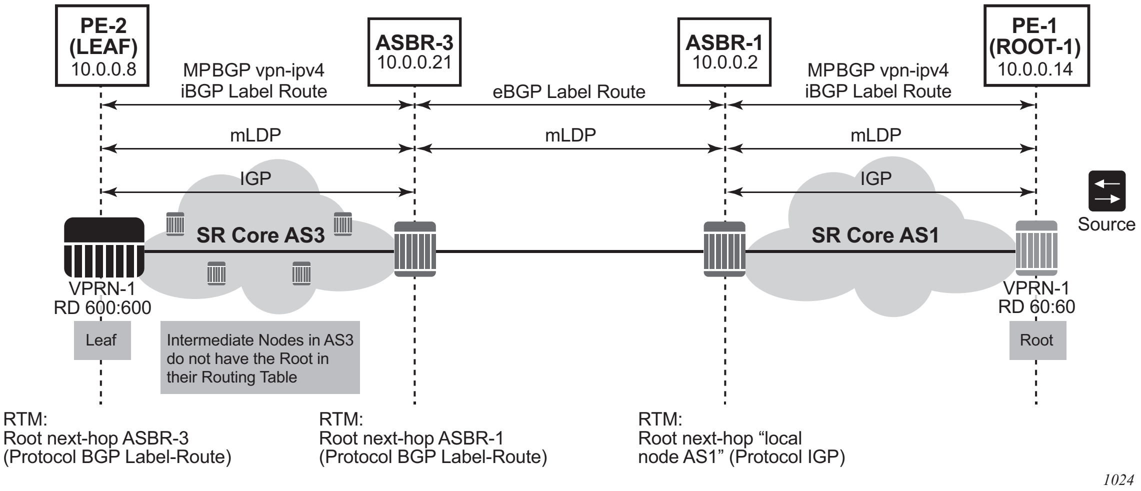
mLDP can be used to transport multicast in GRT. For mLDP LSPs to be generated, a multicast request from the leaf node is required to force mLDP to generate a downstream unsolicited (DU) FEC toward the root to build the P2MP LSPs.

For inter-AS solutions, the root might not be in the leaf’s RTM or, if it is present, it is installed using BGP with ASBRs acting as the leaf’s local AS root. Therefore, the leaf’s local AS intermediate routers might not know the path to the root.

Control protocols used for constructing P2MP LSPs contain a field that identifies the address of a root node. Intermediate nodes are expected to be able to look up that address in their routing tables; however, this is not possible if the route to the root node is a BGP route and the intermediate nodes are part of a BGP-free core (for example, if they use IGP).

To enable an mLDP LSP to be constructed through a BGP-free segment, the root node address is temporarily replaced by an address that is known to the intermediate nodes and is on the path to the true root node. For example, Figure 82 shows the procedure when the PE-2 (leaf) receives the route for root through ASBR3. This route resembles the root next-hop ASBR-3. The leaf, in this case, generates an LDP FEC which has an opaque value, and has the root address set as ASBR-3. This opaque value has additional information needed to reach the root from ASBR-3. As a result, the SR core AS3 only needs to be able to resolve the local AS ASBR-3 for the LDP FEC. The ASBR-3 uses the LDP FEC opaque value to find the path to the root.

Figure 82 **Inter-AS Option C**



Because non-segmented d-mLDP requires end-to-end mLDP signaling, the ASBRs support both mLDP and BGP signaling between them.

7.18.5.2 LDP Recursive FEC Process

For inter-AS networks where the leaf node does not have the root in the RTM or where the leaf node has the root in the RTM using BGP, and the leaf’s local AS intermediate nodes do not have the root in their RTM because they are not BGP-enabled, RFC 6512 defines a recursive opaque value and procedure for LDP to build an LSP through multiple ASs.

For mLDP to be able to signal through a multiple-AS network where the intermediate nodes do not have a routing path to the root, a recursive opaque value is needed. The LDP FEC root resolves the local ASBR, and the recursive opaque value contains the P2MP FEC element, encoded as specified in RFC 6513, with a type field, a length field, and a value field of its own.

RFC 6826 section 3 defines the Transit IPv4 opaque for P2MP LDP FEC, where the leaf in the local AS wants to establish an LSP to the root for P2MP LSP. Figure 83 shows this FEC representation.

Figure 83 mLDP FEC for Single AS with Transit IPv4 Opaque

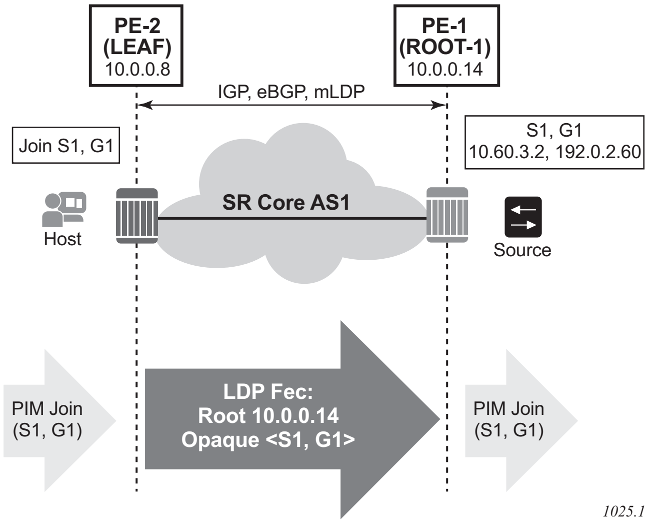
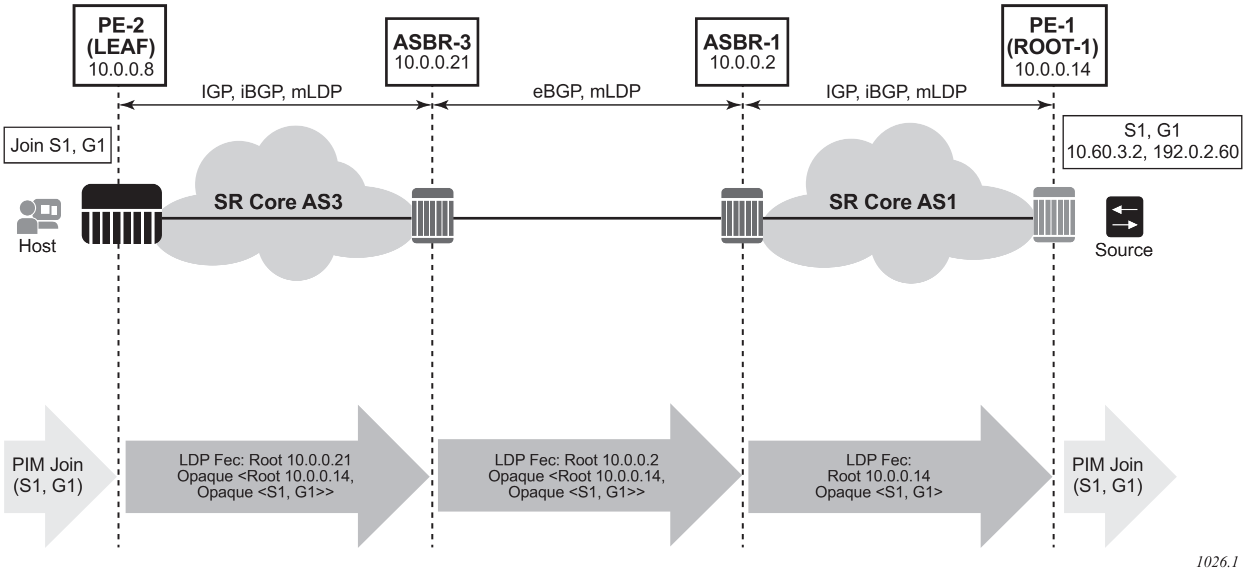


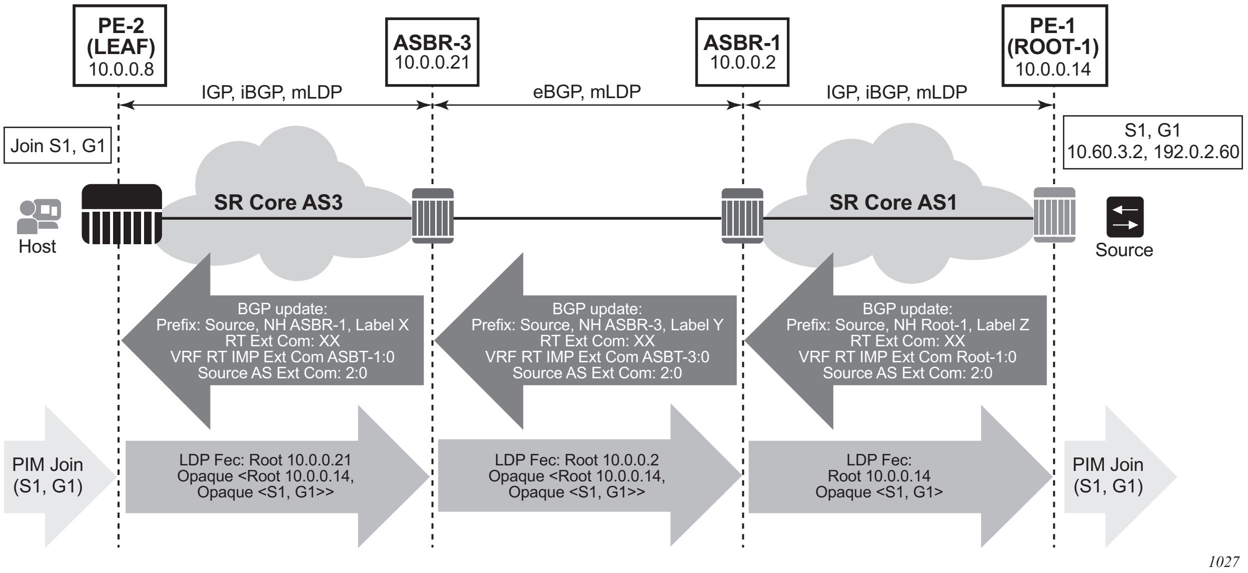
Figure 84 shows an inter-AS FEC with recursive opaque based on RFC 6512.

Figure 84 mLDP FEC for Inter-AS with Recursive Opaque Value



As shown in Figure 84, the root “10.0.0.21” is an ASBR and the opaque value contains the original mLDP FEC. As such, in the leaf’s AS where the actual root “10.0.0.14” is not known, the LDP FEC can be routed using the local root of ASBR. When the FEC arrives at an ASBR that co-locates in the same AS as the actual root, an LDP FEC with transit IPv4 opaque is generated. The end-to-end picture for inter-AS mLDP for non-VPN multicast is shown in Figure 85.

Figure 85 **Non-VPN mLDP with Recursive Opaque for Inter-AS**



As shown in Figure 85, the leaf is in AS3s where the AS3 intermediate nodes do not have the ROOT-1 in their RTM. The leaf has the S1 installed in the RTM via BGP. All ASBRs are acting as next-hop-self in the BGP domain. The leaf resolving the S1 via BGP will generate an mLDP FEC with recursive opaque, represented as:

Leaf FEC: <Root=ASBR-3, opaque-value=<Root=Root-1, <opaque-value = S1,G1>>>

This FEC will be routed through the AS3 Core to ASBR-3.

Note: AS3 intermediate nodes do not have ROOT-1 in their RTM; that is, are not BGP-capable.

At ASBR-3 the FEC will be changed to:

ASBR-3 FEC: <Root=ASBR-1, opaque-value=<Root=Root-1, <opaque-value = S1,G1>>>

This FEC will be routed from ASBR-3 to ASBR-1. ASBR-1 is co-located in the same AS as ROOT-1. Therefore, the ASBR-1 does not need a FEC with a recursive opaque value.

ASBR-1 FEC: <Root=Root-1, <opaque-value =S1,G1>>

This process allows all multicast services to work over inter-AS networks. All d-mLDP opaque types can be used in a FEC with a recursive opaque value.

7.18.5.3 Supported Recursive Opaque Values

A recursive FEC is built using the Recursive Opaque Value and VPN-Recursive Opaque Value types (opaque values 7 and 8 respectively). All SR non-recursive opaque values can be recursively embedded into a recursive opaque.

Table 35 displays all supported opaque values in SR OS.

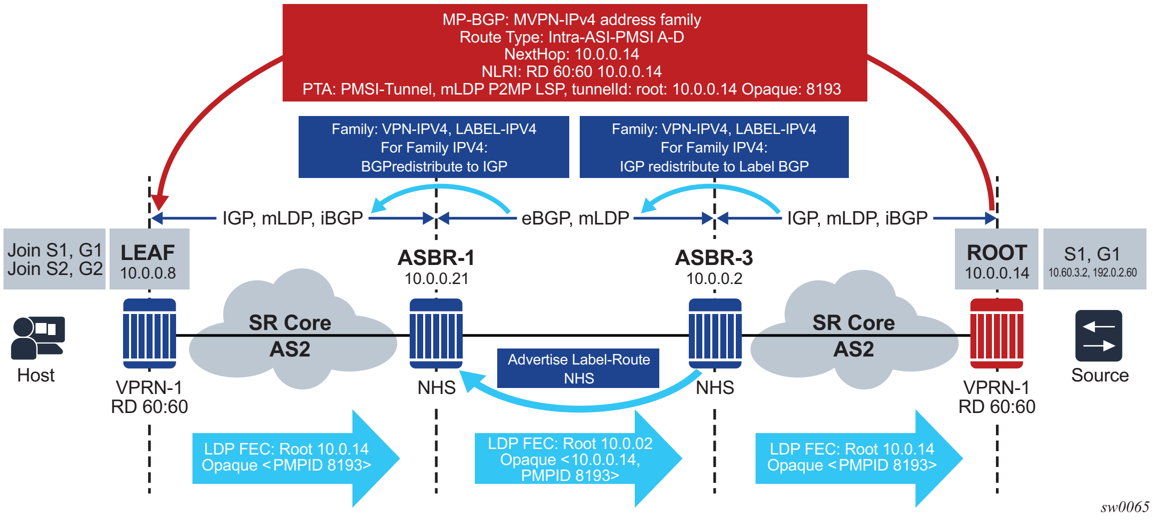
Table 35 **Opaque Types Supported By SR OS**

Opaque Type	Opaque Name	RFC	SR OS Use	FEC Representation
1	Generic LSP Identifier	RFC 6388	VPRN Local AS	<Root, Opaque<P2MPID>>
3	Transit IPv4 Source TLV Type	RFC 6826	IPv4 multicast over mLDP in GRT	<Root, Opaque<SourceIPv4, GroupIPv4>>
4	Transit IPv6 Source TLV Type	RFC 6826	IPv6 multicast over mLDP in GRT	<Root, Opaque<SourceIPv6, GroupIPv6>>
7	Recursive Opaque Value	RFC 6512	Inter-AS IPv4 multicast over mLDP in GRT	<ASBR, Opaque<Root, Opaque<SourceIPv4, GroupIPv4>>>
			Inter-AS IPv6 multicast over mLDP in GRT	<ASBR, Opaque<Root, Opaque<SourceIPv6, GroupIPv6>>>
			Inter-AS Option C MVPN over mLDP	<ASBR, Opaque<Root, Opaque<P2MPID>>>
8	VPN-Recursive Opaque Value	RFC 6512	Inter-AS Option B MVPN over mLDP	<ASBR, Opaque <RD, Root, P2MPID>>
250	Transit VPNv4 Source TLV Type	RFC 7246	In-band signaling for VPRN	<Root, Opaque<SourceIPv4 or RPA, GroupIPv4, RD>>
251	Transit VPNv6 Source TLV Type	RFC 7246	In-band signaling for VPRN	<Root, Opaque<SourceIPv6 or RPA, GroupIPv6, RD>>

7.18.5.4 Optimized Option C and Basic FEC Generation for Inter-AS

Not all leaf nodes can support label route or recursive opaque, so recursive opaque functionality can be transferred from the leaf to the ASBR, as shown in Figure 86.

Figure 86 **Optimized Option C — Leaf Router Not Responsible for Recursive FEC**



In Figure 86, the root advertises its unicast routes to ASBR-3 using IGP, and the ASBR-3 advertises these routes to ASBR-1 using label-BGP. ASBR-1 can redistribute these routes to IGP with next-hop ASBR-1. The leaf will resolve the actual root 10.0.0.14 using IGP and will create a type 1 opaque value <Root 10.0.0.14, Opaque <8193>> to ASBR-1. In addition, all P routers in AS 2 will know how to resolve the actual root because of BGP-to-IGP redistribution within AS 2.

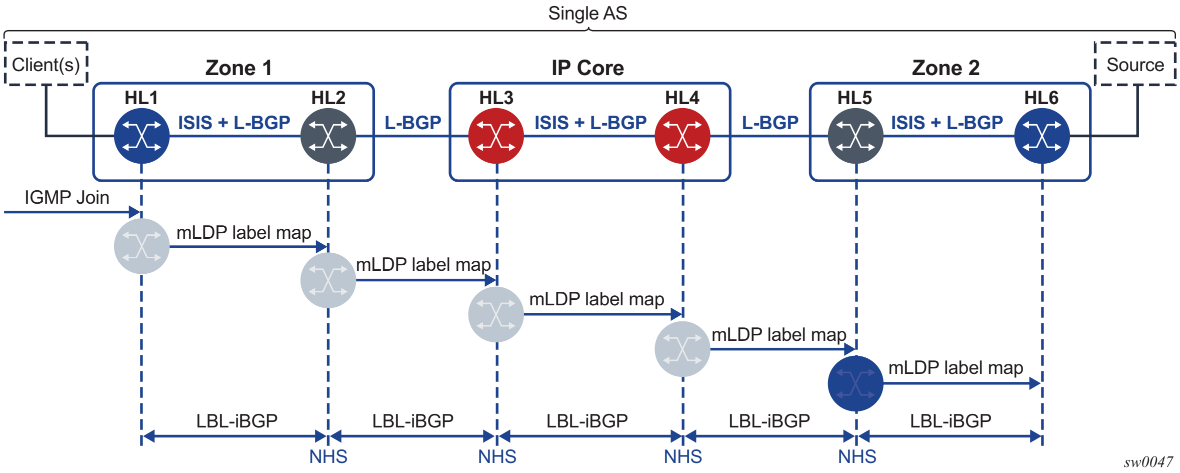
ASBR-1 will attempt to resolve the 10.0.0.14 actual route via BGP, and will create a recursive type 7 opaque value <Root 10.0.0.2, Opaque <10.0.0.14, 8193>>.

7.18.5.5 Basic Opaque Generation When Root PE is Resolved Using BGP

For inter-AS or intra-AS MVPN, the root PE (the PE on which the source resides) loopback IP address is usually not advertised into each AS or area. As such, the P routers in the ASs or areas that the root PE is not part of are not able to resolve the root PE loopback IP address. To resolve this issue, the leaf PE, which has visibility of the root PE loopback IP address using BGP, creates a recursive opaque with an outer root address of the local ASBR or ABR and an inner recursive opaque of the actual root PE.

Some non-Nokia routers do not support recursive opaque FEC when the root node loopback IP address is resolved using IBGP or EBGP. These routers will accept and generate a basic opaque type. In such cases, there should not be any P routers between a leaf PE and ASBR or ABR, or any P routers between ASBR or ABR and the upstream ASBR or ABR. Figure 87 shows an example of this situation.

Figure 87 Example AS



In Figure 87, the leaf HL1 is directly attached to ABR HL2, and ABR HL2 is directly attached to ABR HL3. In this case, it is possible to generate a non-recursive opaque simply because there is no P router that cannot resolve the root PE loopback IP address in between any of the elements. All elements are BGP-speaking and have received the root PE loopback IP address via IBGP or EBGP.

In addition, SR OS does not generate a recursive FEC. The global **generate-basic-fec-only** command disables recursive opaque FEC generation when the provider desires basic opaque FEC generation on the node. In Figure 87, the basic non-recursive FEC is generated even if the root node HL6 is resolved via BGP (IBGP or EBGP).

Currently, when the root node HL6 systemIP is resolved via BGP, a recursive FEC is generated by the leaf node HL1:

HL1 FEC = <HL2, <HL6, OPAQUE>>

When the **generate-basic-fec-only** command is enabled on the leaf node or any ABR, they will generate a basic non-recursive FEC:

HL1 FEC = <HL6, OPAQUE>

When this FEC arrives at HL2, if the **generate-basic-fec-only** command is enabled then HL2 will generate the following FEC:

HL2 FEC = <HL6, OPAQUE>

If there are any P routers between the leaf node and an ASBR or ABR, or any P routers between ASBRs or ABRs that do not have the root node (HL6) in their RTM, then this type 1 opaque FEC will not be resolved and forwarded upstream, and the solution will fail.

7.18.5.5.1 Leaf and ABR Behavior

When **generate-basic-fec-only** is enabled on a leaf node, LDP generates a basic opaque type 1 FEC.

When **generate-basic-fec-only** is enabled on the ABR, LDP will accept a lower FEC of basic opaque type 1 and generate a basic opaque type 1 upper FEC. LDP then stitches the lower and upper FECs together to create a cross connect.

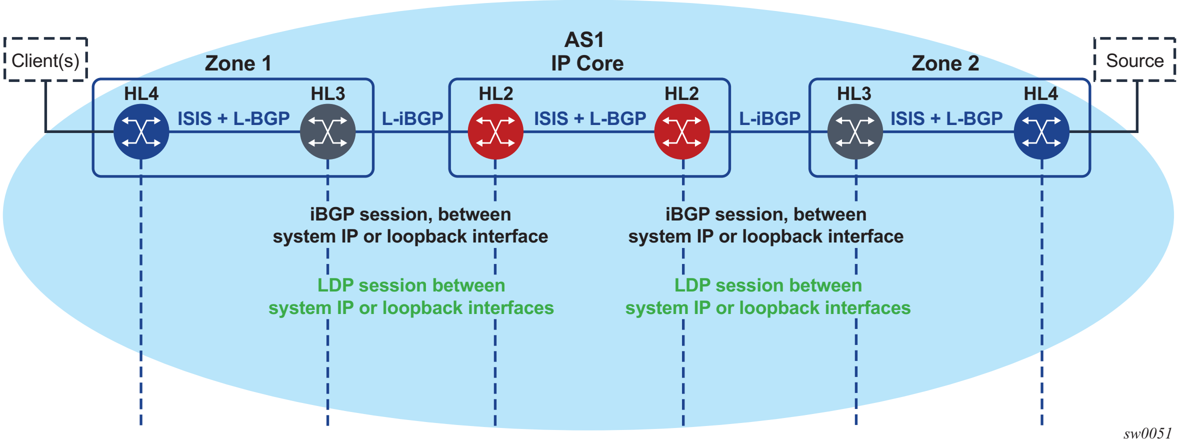
When **generate-basic-fec-only** is enabled and the ABR receives a lower FEC of:

a. recursive FEC with type 7 opaque — The ABR will stitch the lower FEC to an upper FEC with basic opaque type 1.
b. any FEC type other than a recursive FEC with type 7 opaque or a non-recursive FEC with type 1 basic opaque — ABR will process the packet in the same manner as when **generate-basic-fec-only** is disabled.

7.18.5.5.2 Intra-AS Support

ABR uses IBGP and peers between systemIP or loopback IP addresses, as shown in Figure 88.

Figure 88 **ABR and IBGP**



The **generate-basic-fec-only** command is supported on leaf PE and ABR nodes. The **generate-basic-fec-only** command only interoperates with intra-AS as option C, or opaque type 7 with inner opaque type 1. No other opaque type is supported.

7.18.5.5.3 Opaque Type Behavior with Basic FEC Generation

Table 36 describes the behavior of different opaque types when the **generate-basic-fec-only** command is enabled or disabled.

Table 36 **Opaque Type Behavior with Basic FEC Generation**

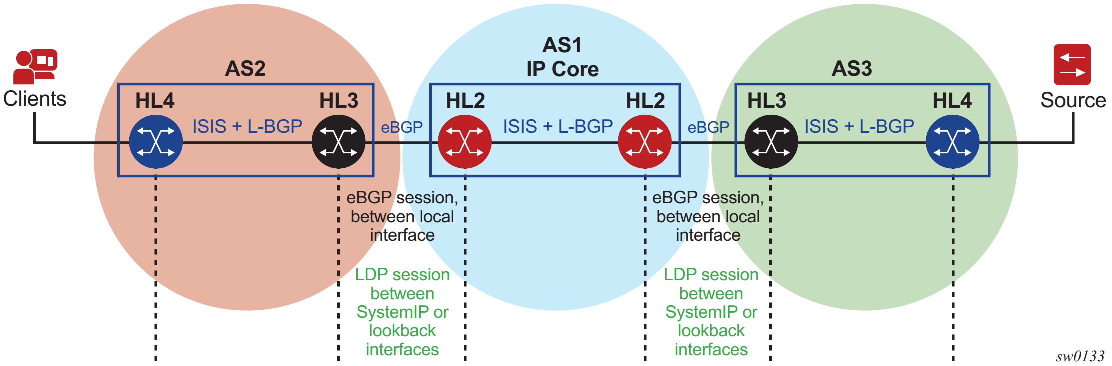
FEC Opaque Type	generate-basic-fec-only Enabled
1	Generate type 1 basic opaque when FEC is resolved using BGP route
3	Same behavior as when **generate-basic-fec-only** is disabled
4	Same behavior as when **generate-basic-fec-only** is disabled
7 with inner type 1	Generate type 1 basic opaque
7 with inner type 3 or 4	Same behavior as when **generate-basic-fec-only** is disabled
8 with inner type 1	Same behavior as when **generate-basic-fec-only** is disabled

7.18.5.5.4 Inter-AS Support

In the inter-AS case, the ASBRs use EBGP as shown in Figure 89.

The two ASBRs become peers via local interface. The **generate-basic-fec-only** command can be used on the LEAF or the ASBR to force SR OS to generate a basic opaque FEC when the actual ROOT is resolved via BGP. The opaque type behavior is on par with the intra-AS scenario as shown in Figure 88.

Figure 89 **ASBR and EBGP**

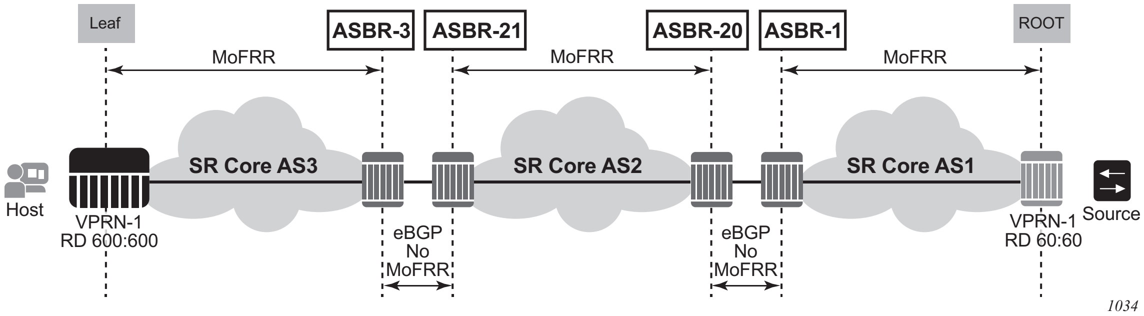

Clients
AS2
HL4
HL3
ISIS + L-BGP
eBGP
AS1
IP Core
HL2
HL2
ISIS + L-BGP
eBGP
AS3
HL3
HL4
ISIS + L-BGP
Source
eBGP session, between local interface
LDP session between SystemIP or lookback interfaces
eBGP session, between local interface
LDP session between SystemIP or lookback interfaces
sw0133

The **generate-basic-fec-only** command is supported on LEAF PE and ASBR nodes in case of inter-AS. The **generate-basic-fec-only** command only interoperates with inter-AS as option C and opaque type 7 with inner opaque type 1.

7.18.5.6 Redundancy and Resiliency

For mLDP, MoFRR is supported with the IGP domain; for example, ASBRs that are not directly connected. MoFRR is not supported between directly connected ASBRs, such as ASBRs that are using EBGP without IGP.

Figure 90 **ASBRs Using EBGP Without IGP**



7.18.5.7 ASBR Physical Connection

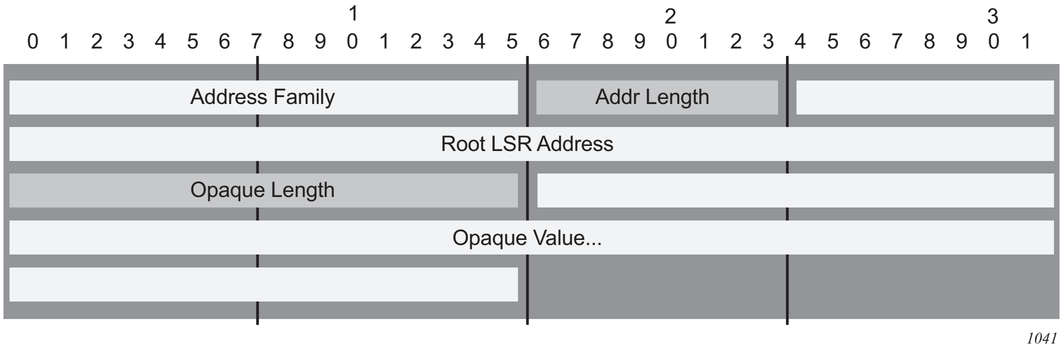
Non-segmented mLDP functions with ASBRs directly connected or connected via an IGP domain, as shown in Figure 90.

7.18.5.8 OAM

LSPs are unidirectional tunnels. When an LSP ping is sent, the echo request is transmitted via the tunnel and the echo response is transmitted via the vanilla IP to the source. Similarly, for a **p2mp-lsp-ping**, on the root, the echo request is transmitted via the mLDP P2MP tunnel to all leafs and the leafs use vanilla IP to respond to the root.

The echo request for mLDP is generated carrying a root Target FEC Stack TLV, which is used to identify the multicast LDP LSP under test at the leaf. The Target FEC Stack TLV must carry an mLDP P2MP FEC Stack Sub-TLV from RFC 6388 or RFC 6512. See Figure 91.

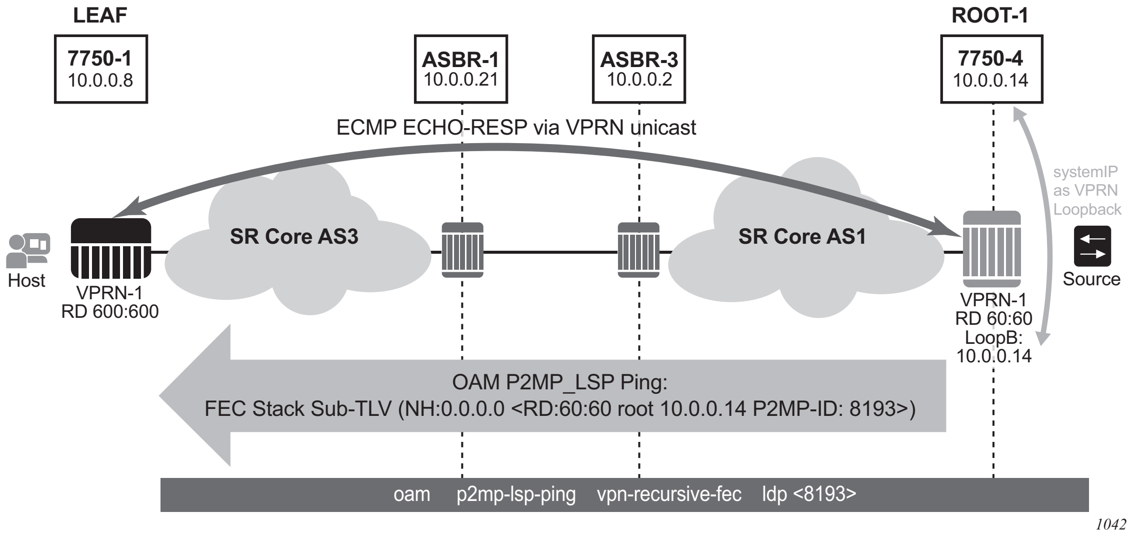
Figure 91 **ECHO Request Target FEC Stack TLV**



The same concept applies to inter-AS and non-segmented mLDP. The leafs in the remote AS should be able to resolve the root via GRT routing. This is possible for inter-AS Option C where the root is usually in the leaf RTM, which is a next-hop ASBR.

For inter-AS Option B where the root is not present in the leaf RTM, the echo reply cannot be forwarded via the GRT to the root. To solve this problem, for inter-AS Option B, the SR OS uses VPRN unicast routing to transmit the echo reply from the leaf to the root via VPRN.

Figure 92 **MVPN Inter-AS Option B OAM**



As shown in Figure 92, the echo request for VPN recursive FEC is generated from the root node by executing the **p2mp-lsp-ping** with the **vpn-recursive-fec** option. When the echo request reaches the leaf, the leaf uses the sub-TLV within the echo request to identify the corresponding VPN via the FEC which includes the RD, the root, and the P2MP-ID.

After identifying the VPRN, the echo response is sent back via the VPRN and unicast routes. There should be a unicast route (for example, root 10.0.0.14, as shown in Figure 92) present in the leaf VPRN to allow the unicast routing of the echo reply back to the root via VPRN. To distribute this root from the root VPRN to all VPRN leafs, a loopback interface should be configured in the root VPRN and distributed to all leafs via MP-BGP unicast routes.

The OAM functionality for Options B and C is summarized in Table 37.

Notes:

1. For SR OS, all P2MP mLDP FEC types will respond to the **vpn-recursive-fec** echo request. Leafs in the local-AS and inter-AS Option C will respond to the recursive-FEC TLV echo request in addition to the leafs in the inter-AS Option B.
   a. For non inter-AS Option B where the root system IP is visible through the GRT, the echo reply will be sent via the GRT, that is, not via the VPRN.
2. This **vpn-recursive-fec** is a Nokia proprietary implementation, and therefore third-party routers will not recognize the recursive FEC and will not generate an echo respond.

a. The user can generate the **p2mp-lsp-ping** without the **vpn-recursive-fec** to discover non-Nokia routers in the local-AS and inter-AS Option C, but not the inter-AS Option B leafs.

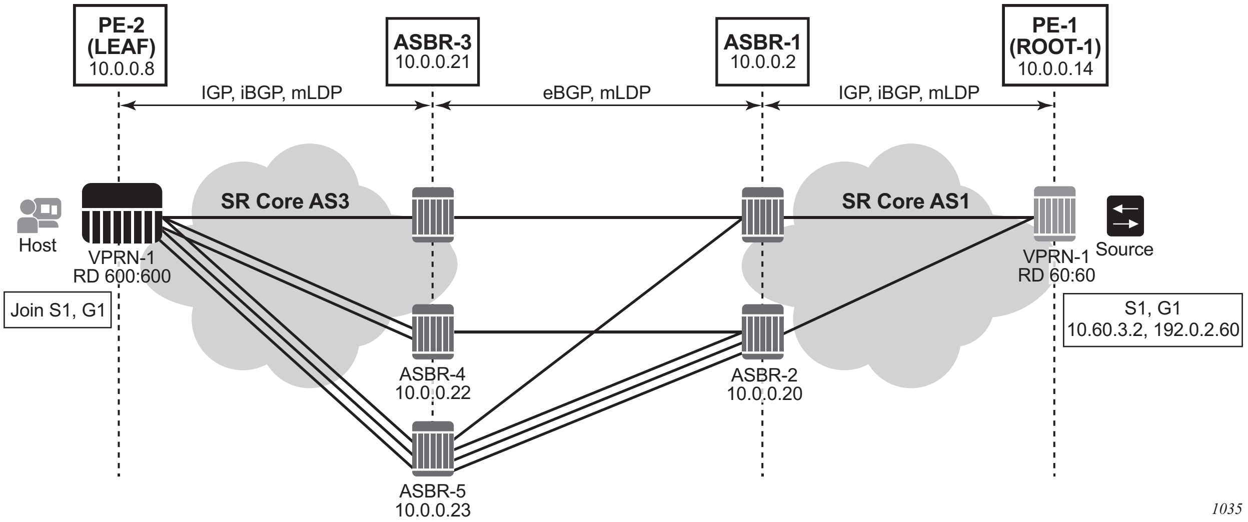
Table 37 **OAM Functionality for Options B and C**

OAM Command (for mLDP)	Leaf and Root in Same AS	Leaf and Root in Different AS (Option B)	Leaf and Root in Different AS (Option C)
p2mp-lsp-ping ldp	✓		✓
p2mp-lsp-ping ldp-ssm	✓		✓
p2mp-lsp-ping ldp vpn-recursive-fec	✓	✓	✓
p2mp-lsp-trace			

7.18.5.9 ECMP Support

In Figure 93, the leaf discovers the ROOT-1 from all three ASBRs (ASBR-3, ASBR-4 and ASBR-5).

Figure 93 **ECMP Support**



The leaf chooses which ASBR will be used for the multicast stream using the following process.

1. The leaf determines the number of ASBRs that should be part of the hash calculation.

   The number of ASBRs that are part of the hash calculation comes from the configured ECMP (**config**>**router**>**ecmp**). For example, if the ECMP value is set to 2, only two of the ASBRs will be part of the hash algorithm selection.

2. After deciding the upstream ASBR, the leaf determines whether there are multiple equal cost paths between it and the chosen ASBR.

   - If there are multiple ECMP paths between the leaf and the ASBR, the leaf performs another ECMP selection based on the configured value in **config**>**router**>**ecmp**. This is a recursive ECMP lookup.
   - The first lookup chooses the ASBR and the second lookup chooses the path to that ASBR.

     For example, if the ASBR 5 was chosen in Figure 93, there are three paths between the leaf and ASBR-5. As such, a second ECMP decision is made to choose the path.

3. At ASBR-5, the process is repeated. For example, in Figure 93, ASBR-5 will go through steps 1 and 2 to choose between ASBR-1 and ASBR-2, and a second recursive ECMP lookup to choose the path to that ASBR.

When there are several candidate upstream LSRs, the LSR must select one upstream LSR. The algorithm used for the LSR selection is a local matter. If the LSR selection is done over a LAN interface and the Section 6 procedures are applied, the procedure described in ECMP Hash Algorithm should be applied to ensure that the same upstream LSR is elected among a set of candidate receivers on that LAN.

The ECMP hash algorithm ensures that there is a single forwarder over the LAN for a particular LSP.

7.18.5.9.1 ECMP Hash Algorithm

The ECMP hash algorithm requires the opaque value of the FEC (see Table 35) and is based on RFC 6388 section 2.4.1.1.

- The candidate upstream LSRs are numbered from lower to higher IP addresses.
- The following hash is performed: *H* **= (CRC32 (Opaque Value)) modulo** *N*, where *N* is the number of upstream LSRs. The “Opaque Value” is the field identified in the FEC element after “Opaque Length”. The “Opaque Length” indicates the size of the opaque value used in this calculation.
- The selected upstream LSR U is the LSR that has the number *H* above.

7.18.5.10 Dynamic mLDP and Static mLDP Co-existing on the Same Node

When creating a static mLDP tunnel, the user must configure the P2MP tunnel ID.

Example:

```
*A:SwSim2>config>router# tunnel-interface
  no tunnel-interface ldp-p2mp p2mp-id sender sender-address
  tunnel-interface ldp-p2mp p2mp-id sender sender-address [root-node]
```

This **p2mp-id** can coincide with a dynamic mLDP **p2mp-id** (the dynamic mLDP is created by the PIM automatically without manual configuration required). If the node has a static mLDP and dynamic mLDP with same label and **p2mp-id**, there will be collisions and OAM errors.

Do not use a static mLDP and dynamic mLDP on same node. If it is necessary to do so, ensure that the **p2mp-id** is not the same between the two tunnel types.

Static mLDP FECs originate at the leaf node. If the FEC is resolved using BGP, it will not be forwarded downstream. A static mLDP FEC will only be created and forwarded if it is resolved using IGP. For optimized Option C, the static mLDP can originate at the leaf node because the root is exported from BGP to IGP at the ASBR; therefore the leaf node resolves the root using IGP.

In the optimized Option C scenario, it is possible to have a static mLDP FEC originate from a leaf node as follows:

static-mLDP <Root: ROOT-1, Opaque: <p2mp-id-1>>

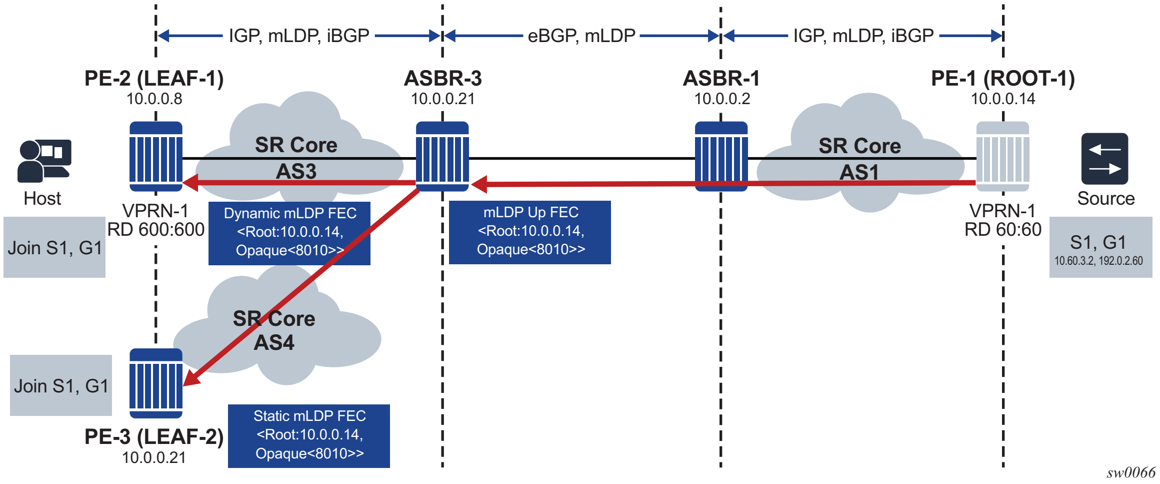
A dynamic mLDP FEC can also originate from a separate leaf node with the same FEC:

dynamic-mLDP <Root: ROOT-1, Opaque: <p2mp-id-1>>

In this case, the tree and the up-FEC will merge the static mLDP and dynamic mLDP traffic at the ASBR. The user must ensure that the static mLDP **p2mp-id** is not used by any dynamic mLDP LSPs on the path to the root.

Figure 94 illustrates the scenario where one leaf (LEAF-1) is using dynamic mLDP for NG-MVPN and a separate leaf (LEAF-2) is using static mLDP for a tunnel interface.

Figure 94 **Static and Dynamic mLDP Interaction**

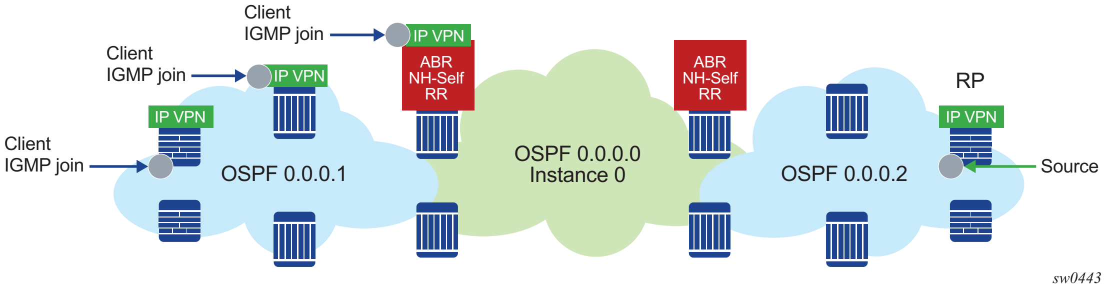


In Figure 94, both FECs generated by LEAF-1 and LEAF-2 are identical, and the ASBR-3 will merge the FECs into a single upper FEC. Any traffic arriving from ROOT-1 to ASBR-3 over VPRN-1 will be forked to LEAF-1 and LEAF-2, even if the tunnels were signaled for different services.

7.18.6 Intra-AS Non-segmented mLDP

Non-segmented mLDP intra-AS (inter-area) is supported on option Band C only. Figure 95 shows a typical intra-AS topology. With a backbone IGP area 0 and access non- backbone IGP areas 1 and 2. In these topologies, the ABRs usually does next-hop-self for BGP label routes, which requires recursive FEC.

Figure 95 **Intra-AS Non-segmented Topology**



For option B, the ABR routers change the next hop of the MVPN AD routes to the ABR system IP or Loopback IP. The **next-hop-self** command for BGP does not change the next hop of the MVPN AD routes. Instead, a BGP policy can be used to change the MVPN AD routes next hop at the ABR.

In the meantime a BGP policy can be used to change the MVPN AD routes nexthop at the ABR.

7.18.6.1 ABR MoFRR for Intra-AS

With ABR MoFRR in the intra-AS environment, the leaf will choose a local primary ABR and a backup ABR, with separate mLDP signaling toward these two ABRs. In addition, each path from a leaf to the primary ABR and from a leaf to the backup ABR will support IGP MoFRR. This behavior is similar to ASBR MoFRR in the inter-AS environment; for more details, see ASBR MoFRR.

MoFRR is only supported for intra-AS option C, with or without RR.

7.18.6.2 Interaction with an Inter-AS Non-segmented mLDP Solution

Intra-AS option C will be supported in conjunction to inter-AS option B or C. Intra-AS option C with inter-AS option B is not supported.

7.18.6.3 Intra-AS/Inter-AS Option B

For intra/inter-AS option B the root is not visible on the leaf. LDP is responsible for building the recursive FEC and signaling the FEC to ABR/ASBR on the leaf. The ABR/ASBR needs to have the PMSI AD router to re-build the FEC (recursive or basic) depending on if they are connected to another ABR/ASBR or to a root node. LDP must import the MVPN PMSI AD routes. To reduce resource usage, importing of the MVPN PMSI AD routes is done manually using the **configure router ldp import-pmsi-routes mvpn** command. When enabled, LDP will request BGP to provide the LDP task with all of the MVPN PMSI AD routes and LDP will cache these routes internally. If **import-pmsi-routes mvpn** is disabled, MVPN will discard the cached routes to save resources.

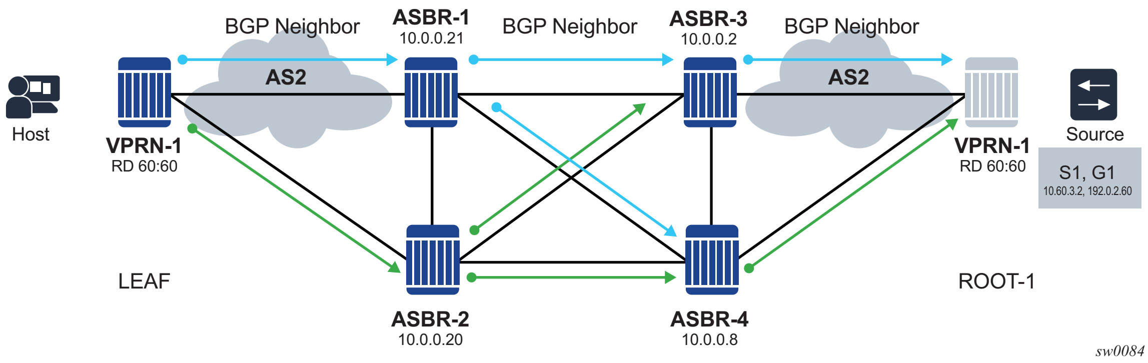
The **import-pmsi-routes mvpn** command is enabled if there is an upgrade from a software version that does not support this inter-AS case. Otherwise, by default import-pmsi-routes **mvpn** is disabled for MVPN inter-AS, MVPN intra-AS, and EVPN, so LDP does not cache any MVPN PMSI AD routes.

7.18.7 ASBR MoFRR

ASBR MoFRR in the inter-AS environment allows the leaf PE to signal a primary path to the remote root through the first ASBR and a backup path through the second ASBR, so that there is an active LSP signaled from the leaf node to the first local root (ASBR-1 in Figure 96) and a backup LSP signaled from the leaf node to the second local root (ASBR-2 in Figure 96) through the best IGP path in the AS.

Using Figure 96 as an example, ASBR-1 and ASBR-2 are local roots for the leaf node, and ASBR-3 and ASBR-4 are local roots for ASBR-1 or ASBR-2. The actual root node (ROOT-1) is also a local root for ASBR-3 and ASBR-4.

Figure 96 **BGP Neighboring for MoFRR**

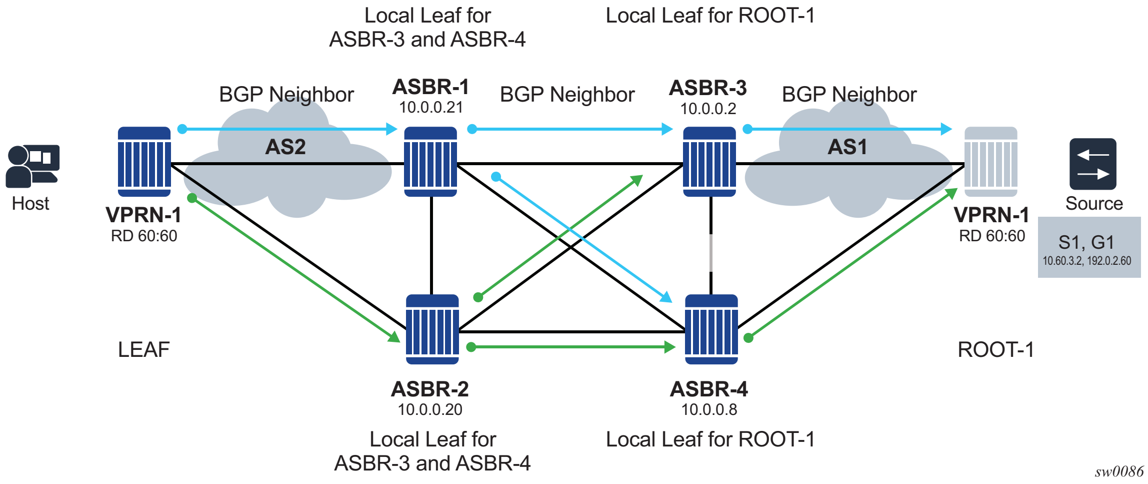


In Figure 96, ASBR-2 is a disjointed ASBR; with the AS spanning from the leaf to the local root, which is the ASBR selected in the AS, the traditional IGP MoFRR is used. ASBR MoFRR is used from the leaf node to the local root, and IGP MoFRR is used for any P router that connects the leaf node to the local root.

7.18.7.1 IGP MoFRR Versus BGP (ASBR) MoFRR

The local leaf can be the actual leaf node that is connected to the host, or an ASBR node that acts as the local leaf for the LSP in that AS, as illustrated in Figure 97.

Figure 97 ASBR Node Acting as Local Leaf

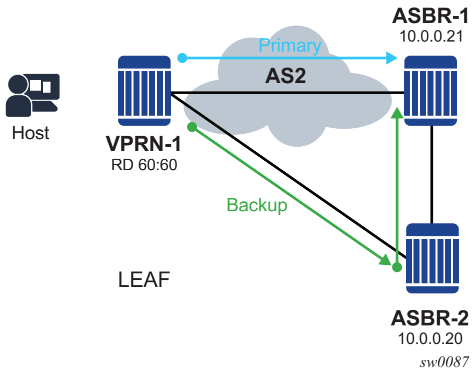


Two types of MoFRR can exist in a unique AS:

- IGP MoFRR — When the **mcast-upstream-frr** command is enabled for LDP, the local leaf selects a single local root, either ASBR or actual, and creates a FEC towards two different upstream LSRs using LFA/ECMP for the ASBR route. If there are multiple ASBRs directed towards the actual root, the local leaf only selects a single ASBR; for example, ASBR-1 in Figure 98. In this example, LSPs are not set up for ASBR-2. The local root ASBR-1 is selected by the local leaf and the primary path is set up to ASBR-1, while the backup path is set up through ASBR-2.

  For more information, see Multicast LDP Fast Upstream Switchover.

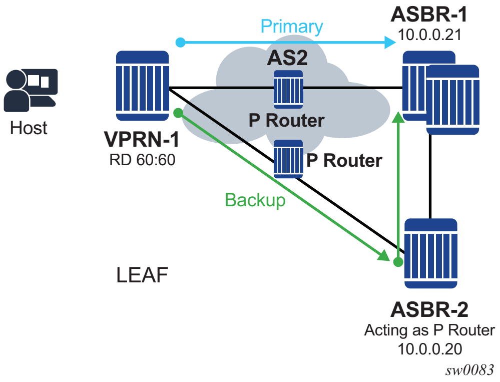
Figure 98 IGP MoFRR



- ASBR MoFRR — When the **mcast-upstream-asbr-frr** command is enabled for LDP, and the **mcast-upstream-frr** command is not enabled, the local leaf will select a single ASBR as the primary ASBR and another ASBR as the backup ASBR. The primary and backup LSPs will be set to these two ASBRs, as shown in Figure 99. Because the **mcast-upstream-frr** command is not configured, IGP MoFRR will not be enabled in the AS2, and therefore none of the P routers will perform local IGP MoFRR.

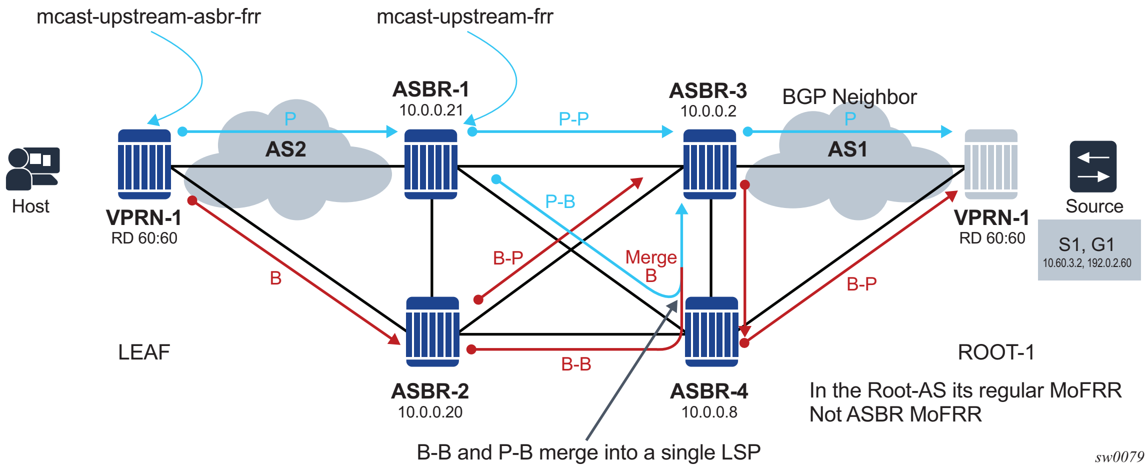
  BGP neighboring and sessions can be used to detect BGP peer failure from the local leaf to the ASBR, and can cause a MoFRR switch from the primary LSP to the backup LSP. Multihop BFD can be used between BGP neighbors to detect failure more quickly and remove the primary BGP peer (ASBR-1 in Figure 99) and its routes from the routing table so that the leaf can switch to the backup LSP and backup ASBR.

Figure 99 **ASBR MoFRR**



The **mcast-upstream-frr** and **mcast-upstream-asbr-frr** commands can be configured together on the local leaf of each AS to create a high-resilience MoFRR solution. When both commands are enabled, the local leaf will set up ASBR MoFRR first and set up a primary LSP to one ASBR (ASBR-1 in Figure 100) and a backup LSP to another ASBR (ASBR-2 in Figure 100). In addition, the local leaf will protect each LSP using IGP MoFRR through the P routers in that AS.

Figure 100 **ASBR MoFRR and IGP MoFRR**



Note: Enabling both the **mcast-upstream-frr** and **mcast-upstream-asbr-frr** commands can cause extra multicast traffic to be created. Ensure that the network is designed and the appropriate commands are enabled to meet network resiliency needs.

At each AS, either command can be configured; for example, in Figure 100, the leaf is configured with **mcast-upstream-asbr-frr** enabled and will set up a primary LSP to ASBR-1 and a backup LSP to ASBR-2. ASBR-1 and ASBR-2 are configured with **mcast-upstream-frr** enabled, and will both perform IGP MoFRR to ASBR-3 only. ASBR-2 can select ASBR-3 or ASBR-4 as its local root for IGP MoFRR; in this example, ASBR-2 has selected ASBR-3 as its local root.

There are no ASBRs in the root AS (AS-1), so IGP MoFRR will be performed if **mcast-upstream-frr** is enabled on ASBR-3.

The **mcast-upstream-frr** and **mcast-upstream-asbr-frr** commands work separately depending on the desired behavior. If there is more than one local root, then **mcast-upstream-asbr-frr** can provide extra resiliency between the local ASBRs, and **mcast-upstream-frr** can provide extra redundancy between the local leaf and the local root by creating a disjointed LSP for each ASBR.

If the **mcast-upstream-asbr-frr** command is disabled and **mcast-upstream-frr** is enabled, and there is more than one local root, only a single local root will be selected and IGP MoFRR can provide local AS resiliency.

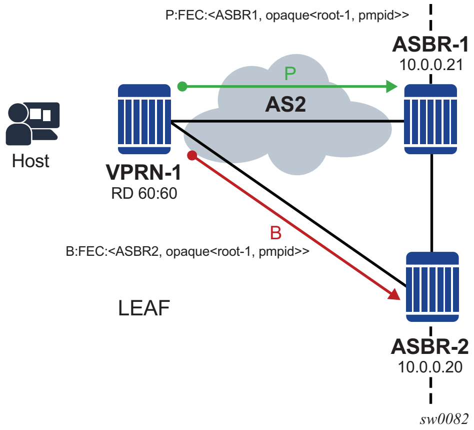
In the actual root AS, only the **mcast-upstream-frr** command needs to be configured.

7.18.7.2 ASBR MoFRR Leaf Behavior

With inter-AS MoFRR at the leaf, the leaf will select a primary ASBR and a backup ASBR. These ASBRs are disjointed ASBRs.

The primary and backup LSPs will be set up using the primary and backup ASBRs, as illustrated in Figure 101.

Figure 101 **ASBR MoFRR Leaf Behavior**


P:FEC:<ASBR1, opaque<root-1, pmpid>>
ASBR-1
10.0.0.21
P
AS2
Host
VPRN-1
RD 60:60
B
B:FEC:<ASBR2, opaque<root-1, pmpid>>
LEAF
ASBR-2
10.0.0.20
sw0082

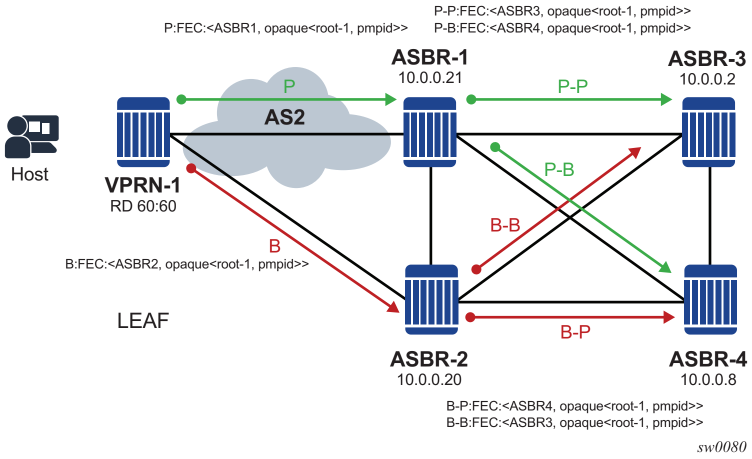
Note: Using Figure 101 as a reference, ensure that the paths to ASBR-1 and ASBR-2 are disjointed from the leaf. MLDP does not support TE and cannot create two disjointed LSPs from the leaf to ASBR-1 and ASBR-2. The operator and IGP architect must define the disjointed paths.

7.18.7.3 ASBR MoFRR ASBR Behavior

Each LSP at the ASBR will create its own primary and backup LSPs.

As shown in Figure 102, the primary LSP from the leaf to ASBR-1 will generate a primary LSP to ASBR-3 (P-P) and a backup LSP to ASBR-4 (P-B). The backup LSP from the leaf also generates a backup primary to ASBR-4 (B-P) and a backup backup to ASBR-3 (B-B). When two similar FECs of an LSP intersect, the LSPs will merge.

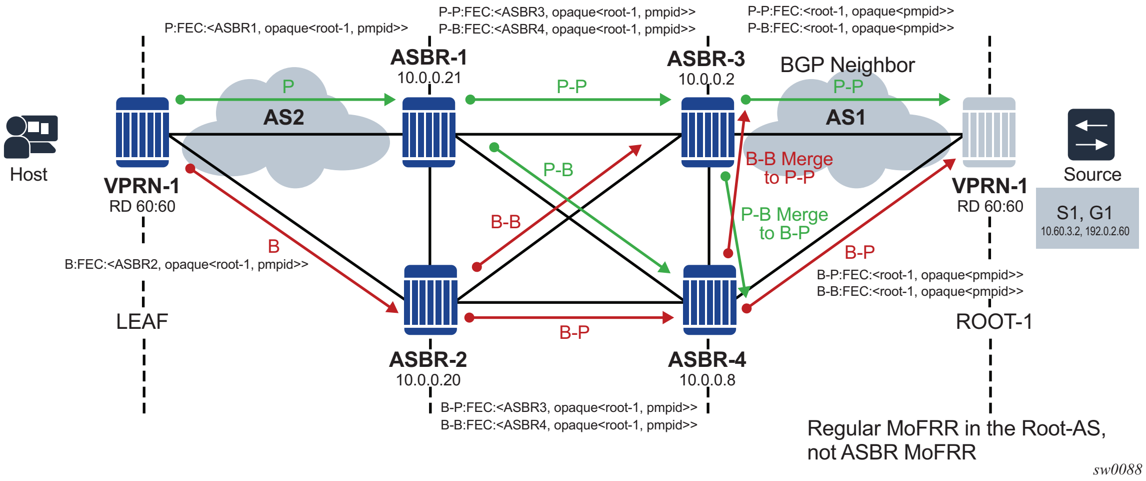
Figure 102   ASBR MoFRR ASBR Behavior


P-P:FEC:<ASBR3, opaque<root-1, pmpid>>
P-B:FEC:<ASBR4, opaque<root-1, pmpid>>
P:FEC:<ASBR1, opaque<root-1, pmpid>>
ASBR-1
10.0.0.21
ASBR-3
10.0.0.2
P
P-P
AS2
Host
VPRN-1
RD 60:60
P-B
B-B
B
B:FEC:<ASBR2, opaque<root-1, pmpid>>
LEAF
B-P
ASBR-2
10.0.0.20
ASBR-4
10.0.0.8
B-P:FEC:<ASBR4, opaque<root-1, pmpid>>
B-B:FEC:<ASBR3, opaque<root-1, pmpid>>
sw0080

7.18.7.4 MoFRR Root AS Behavior

In the root AS, MoFRR is based on regular IGP MoFRR. At the root, there are primary and backup LSPs for each of the primary and backup LSPs that arrive from the neighboring AS, as shown in Figure 103.

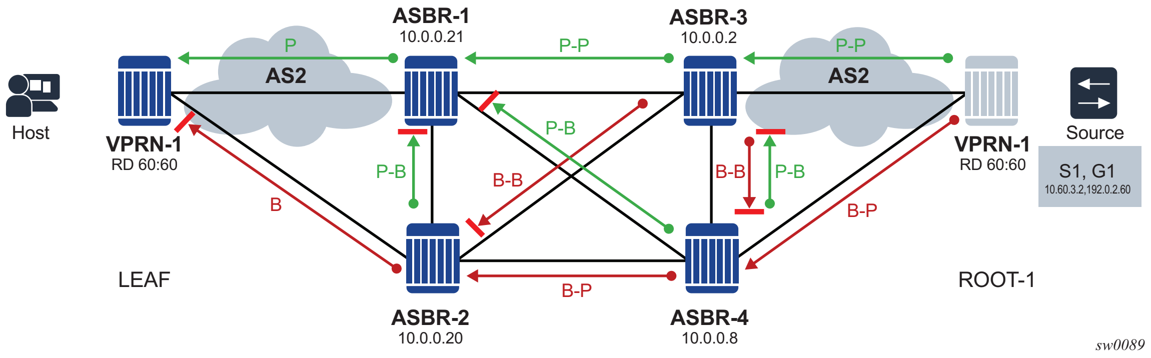
Figure 103   MoFRR Root AS Behavior


P:FEC:<ASBR1, opaque<root-1, pmpid>>
P-P:FEC:<ASBR3, opaque<root-1, pmpid>>
P-B:FEC:<ASBR4, opaque<root-1, pmpid>>
P-P:FEC:<root-1, opaque<pmpid>>
P-B:FEC:<root-1, opaque<pmpid>>
ASBR-1
10.0.0.21
ASBR-3
10.0.0.2
BGP Neighbor
P
P-P
P-P
AS2
AS1
Host
VPRN-1
RD 60:60
P-B
B-B Merge
to P-P
Source
VPRN-1
RD 60:60
S1, G1
10.60.3.2, 192.0.2.60
B-B
P-B Merge
to B-P
B
B:FEC:<ASBR2, opaque<root-1, pmpid>>
B-P
B-P:FEC:<root-1, opaque<pmpid>>
B-B:FEC:<root-1, opaque<pmpid>>
LEAF
ROOT-1
B-P
ASBR-2
10.0.0.20
ASBR-4
10.0.0.8
B-P:FEC:<ASBR3, opaque<root-1, pmpid>>
B-B:FEC:<ASBR4, opaque<root-1, pmpid>>
Regular MoFRR in the Root-AS,
not ASBR MoFRR
sw0088

7.18.7.5 Traffic Flow

Figure 104 illustrates traffic flow based on the LSP setup. The backup LSPs of the primary and backup LSPs (B-B, P-B) will be blocked in the non-leaf AS.

Figure 104 **Traffic Flow**



7.18.7.6 Failure Detection and Handling

Failure detection can be achieved by using either of the following:

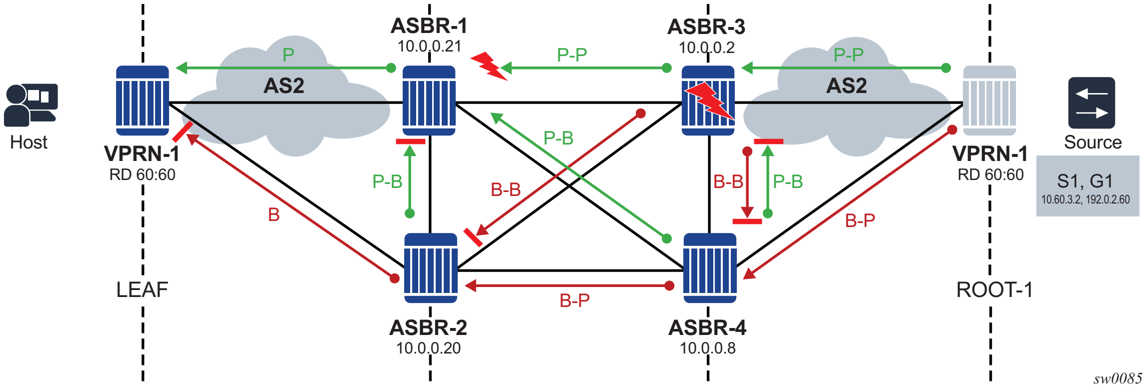
- IGP failure detection
  - Enabling BFD is recommended for IGP protocols or static route (if static route is used for IGP forwarding). This enables faster IGP failure detection.
  - IGP can detect P router failures for IGP MoFRR (single AS).
  - If the ASBR fails, IGP can detect the failure and converge the route table to the local leaf. The local leaf in an AS can be either the ASBR or the actual leaf.
  - IGP routes to the ASBR address must be deleted for IGP failure to be handled.
- BGP failure detection
  - BGP neighboring must be established between the local leaf and each ASBR. Using multi-hop BFD for ASBR failure is recommended.
  - Each local leaf will attempt to calculate a primary ASBR or backup ASBR. The local leaf will set up a primary LSP to the primary ASBR and a backup LSP to the backup ASBR. If the primary ASBR has failed, the local leaf will remove the primary ASBR from the next-hop list and will allow traffic to be processed from the backup LSP and the backup ASBR.
  - BGP MoFRR can offer faster ASBR failure detection than IGP MoFRR.

- BGP MoFRR can also be activated via IGP changes, such as if the node detects a direct link failure, or if IGP removes the BGP neighbor system IP address from the routing table. These events can cause a switch from the primary ASBR to a backup ASBR. It is recommended to deploy IGP and BFD in tandem for fast failure detection.

7.18.7.7 Failure Scenario

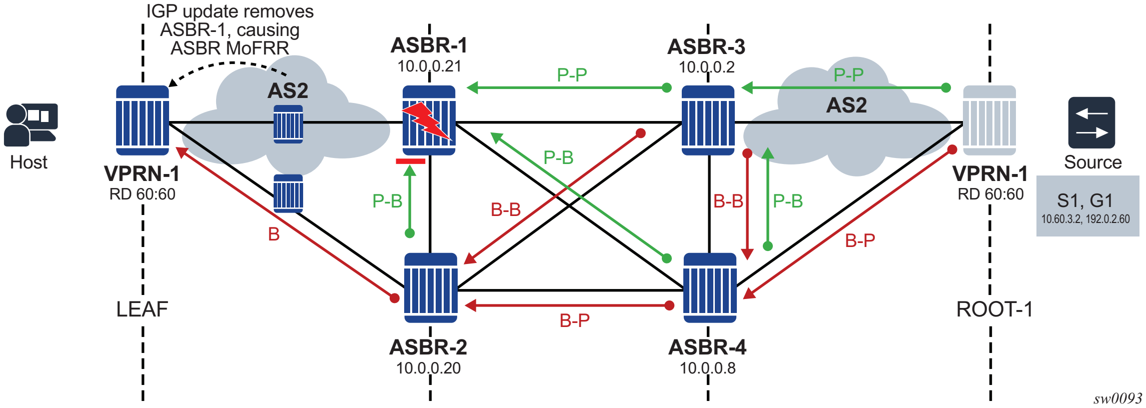
As shown in Figure 105, when ASBR-3 fails, ASBR-1 will detect the failure using ASBR MoFRR and will enable the primary backup path (P-B). This is the case for every LSP that has been set up for ASBR MoFRR in any AS.

Figure 105 **Failure Scenario 1**



In another example, as shown in Figure 106, failure on ASBR-1 will cause the attached P router to generate a route update to the leaf, removing the ASBR-1 from the routing table and causing an ASBR-MoFRR on the leaf node.

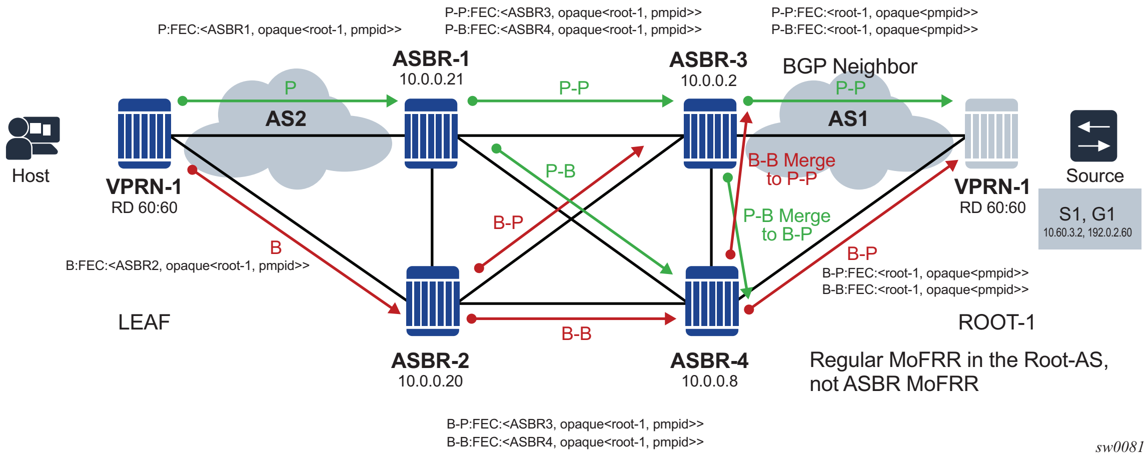
Figure 106 **Failure Scenario 2**



7.18.7.8 ASBR MoFRR Consideration

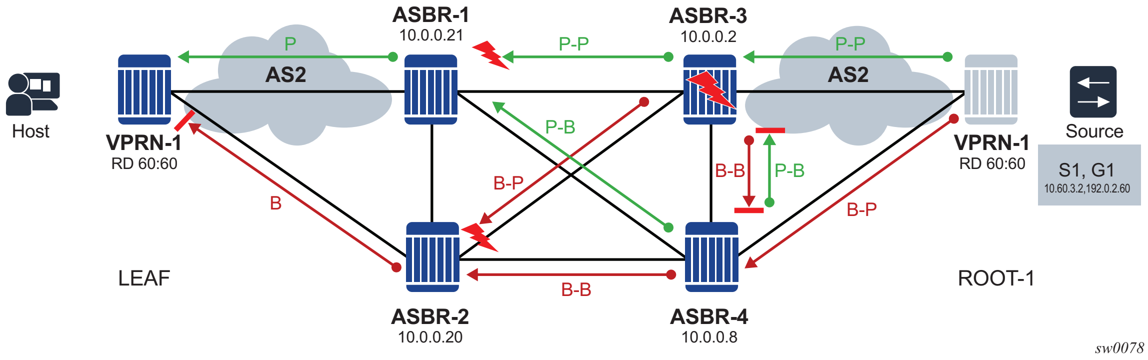
As illustrated in Figure 107, it is possible for the ASBR-1 primary-primary (P-P) LSP to be resolved using ASBR-3, and for the ASBR-2 backup-primary (B-P) LSP to be resolved using the same ASBR-3.

Figure 107 **Resolution via ASBR-3**


P:FEC:<ASBR1, opaque<root-1, pmpid>>
P-P:FEC:<ASBR3, opaque<root-1, pmpid>>
P-B:FEC:<ASBR4, opaque<root-1, pmpid>>
P-P:FEC:<root-1, opaque<pmpid>>
P-B:FEC:<root-1, opaque<pmpid>>
ASBR-1
10.0.0.21
ASBR-3
10.0.0.2
BGP Neighbor
P
AS2
P-P
P-P
AS1
Host
VPRN-1
RD 60:60
P-B
B-B Merge
to P-P
P-B Merge
to B-P
VPRN-1
RD 60:60
Source
S1, G1
10.60.3.2, 192.0.2.60
B-P
B
B:FEC:<ASBR2, opaque<root-1, pmpid>>
B-P
B-P:FEC:<root-1, opaque<pmpid>>
B-B:FEC:<root-1, opaque<pmpid>>
LEAF
B-B
ROOT-1
ASBR-2
10.0.0.20
ASBR-4
10.0.0.8
Regular MoFRR in the Root-AS,
not ASBR MoFRR
B-P:FEC:<ASBR3, opaque<root-1, pmpid>>
B-B:FEC:<ASBR4, opaque<root-1, pmpid>>
sw0081

In this case, both the backup-primary LSP and primary-primary LSP will be affected when a failure occurs on ASBR-3, as illustrated in Figure 108.

Figure 108 **ASBR-3 Failure**



In Figure 108, the MoFRR can switch to the primary-backup LSP between ASBR-4 and ASBR-1 by detecting BGP MoFRR failure on ASBR-3.

It is strongly recommended that LDP signaling be enabled on all links between the local leaf and local roots, and that all P routers enable ASBR MoFRR and IGP MoFRR. If only LDP signaling is configured, the routing table may resolve a next-hop for LDP FEC when there is no LDP signaling and the primary or backup MoFRR LSPs may not be set up.

ASBR MoFRR guarantees that ASBRs will be disjointed, but does not guarantee that the path from the local leaf to the local ASBR will be disjointed. The primary and backup LSPs take the best paths as calculated by IGP, and if IGP selects the same path for the primary ASBR and the backup ASBR, then the two LSPs will not be disjointed. Ensure that 2 disjointed paths are created to the primary and backup ASBRs.

7.18.7.9 ASBR MoFRR Opaque Support

Table 38 lists the FEC opaque types that are supported by ASBR MoFRR.

Table 38 **ASBR MoFRR Opaque Support**

FEC Opaque Type	Supported for ASBR MoFRR
Type 1	✓
Type 3	
Type 4	
Type 7, inner type 1	✓
Type 7, inner type 3 or 4	
Type 8, inner type 1	✓
Type 250	
Type 251	

7.18.8 MBB for MoFRR

Any optimization of the MoFRR primary LSP should be performed by the Make Before Break (MBB) mechanism. For example, if the primary LSP fails, a switch to the backup LSP will occur and the primary LSP will be signaled. After the primary LSP is successfully re-established, MoFRR will switch from the backup LSP to the primary LSP.

MBB is performed from the leaf node to the root node, and as such it is not performed per autonomous system (AS); the MBB signaling must be successful from the leaf PE to the root PE, including all ASBRs and P routers in between.

The conditions of MBB for mLDP LSPs are:

- re-calculation of the SFP
- failure of the primary ASBR

If the primary ASBR fails and a switch is made to the backup ASBR, and the backup ASBR is the only other ASBR available, the MBB mechanism will not signal any new LSP and will use this backup LSP as the primary.

7.18.9 Add-path for Route Reflectors

If the ASBRs and the local leaf are connected by a route reflector, the BGP **add-path** command must be enabled on the route reflector for **mcast-vpn-ipv4** and **mcast-vpn-ipv6**, or for **label-ipv4** if Option C is used. The **add-path** command forces the route reflector to advertise all ASBRs to the local leaf as the next hop for the actual root.

If the **add-path** command is not enabled for the route reflector, only a single ASBR will be advertised to the local root, and ASBR MoFRR will not be available.

7.19 Multicast LDP Fast Upstream Switchover

This feature allows a downstream LSR of a multicast LDP (mLDP) FEC to perform a fast switchover and source the traffic from another upstream LSR while IGP and LDP are converging due to a failure of the upstream LSR which is the primary next-hop of the root LSR for the P2MP FEC. In essence it provides an upstream Fast-Reroute (FRR) node-protection capability for the mLDP FEC packets. It does it at the expense of traffic duplication from two different upstream nodes into the node which performs the fast upstream switchover.

The detailed procedures for this feature are described in *draft-pdutta-mpls-mldp-up-redundancy*.

7.19.1 Feature Configuration

The user enables the mLDP fast upstream switchover feature by configuring the following option in CLI:

config>router>ldp>mcast-upstream-frr

When this command is enabled and LDP is resolving a mLDP FEC received from a downstream LSR, it checks if an ECMP next-hop or a LFA next-hop exist to the root LSR node. If LDP finds one, it programs a primary ILM on the interface corresponding to the primary next-hop and a backup ILM on the interface corresponding to the ECMP or LFA next-hop. LDP then sends the corresponding labels to both upstream LSR nodes. In normal operation, the primary ILM accepts packets while the backup ILM drops them. If the interface or the upstream LSR of the primary ILM goes down causing the LDP session to go down, the backup ILM will then start accepting packets.

In order to make use of the ECMP next-hop, the user must configure the **ecmp** value in the system to at least two (2) using the following command:

config>router>ecmp

In order to make use of the LFA next-hop, the user must enable LFA using the following commands:

config>router>isis>loopfree-alternates

config>router>ospf>loopfree-alternates

Enabling IP FRR or LDP FRR using the following commands is not strictly required since LDP only needs to know where the alternate next-hop to the root LSR is to be able to send the Label Mapping message to program the backup ILM at the initial signaling of the tree. Thus enabling the LFA option is sufficient. If however, unicast IP and LDP prefixes need to be protected, then these features and the mLDP fast upstream switchover can be enabled concurrently:

config>router>ip-fast-reroute

config>router>ldp>fast-reroute

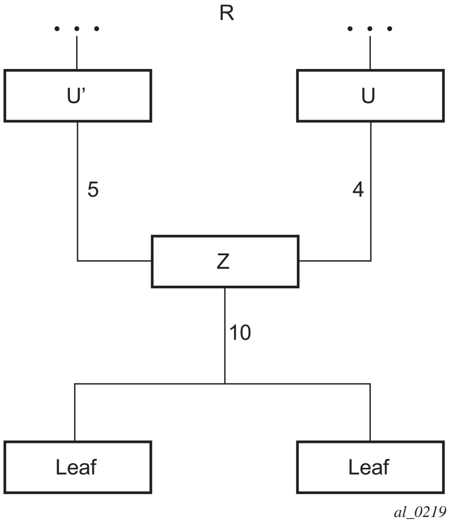
Caution: The mLDP FRR fast switchover relies on the fast detection of loss of **LDP session** to the upstream peer to which the primary ILM label had been advertised. We strongly recommend that you perform the following:

1. Enable BFD on all LDP interfaces to upstream LSR nodes. When BFD detects the loss of the last adjacency to the upstream LSR, it will bring down immediately the LDP session which will cause the IOM to activate the backup ILM.
2. If there is a concurrent TLDP adjacency to the same upstream LSR node, enable BFD on the T-LDP peer in addition to enabling it on the interface.
3. Enable ldp-sync-timer option on all interfaces to the upstream LSR nodes. If an LDP session to the upstream LSR to which the primary ILM is resolved goes down for any other reason than a failure of the interface or of the upstream LSR, routing and LDP will go out of sync. This means the backup ILM will remain activated until the next time SPF is rerun by IGP. By enabling IGP-LDP synchronization feature, the advertised link metric will be changed to max value as soon as the LDP session goes down. This in turn will trigger an SPF and LDP will likely download a new set of primary and backup ILMs.

7.19.2 Feature Behavior

This feature allows a downstream LSR to send a label binding to a couple of upstream LSR nodes but only accept traffic from the ILM on the interface to the primary next-hop of the root LSR for the P2MP FEC in normal operation, and accept traffic from the ILM on the interface to the backup next-hop under failure. Obviously, a candidate upstream LSR node must either be an ECMP next-hop or a Loop-Free Alternate (LFA) next-hop. This allows the downstream LSR to perform a fast switchover and source the traffic from another upstream LSR while IGP is converging due to a failure of the LDP session of the upstream peer which is the primary next-hop of the root LSR for the P2MP FEC. In a sense it provides an upstream Fast-Reroute (FRR) node-protection capability for the mLDP FEC packets.

Figure 109 **mLDP LSP with Backup Upstream LSR Nodes**



Upstream LSR U in Figure 109 is the primary next-hop for the root LSR **R** of the P2MP FEC. This is also referred to as primary upstream LSR. Upstream LSR **U**' is an ECMP or LFA backup next-hop for the root LSR **R** of the same P2MP FEC. This is referred to as backup upstream LSR. Downstream LSR **Z** sends a label mapping message to both upstream LSR nodes and programs the primary ILM on the interface to LSR **U** and a backup ILM on the interface to LSR **U**'. The labels for the primary and backup ILMs must be different. LSR **Z** thus will attract traffic from both of them. However, LSR **Z** will block the ILM on the interface to LSR **U**' and will only accept traffic from the ILM on the interface to LSR **U**.

In case of a failure of the link to LSR **U** or of the LSR **U** itself causing the LDP session to LSR **U** to go down, LSR **Z** will detect it and reverse the ILM blocking state and will immediately start receiving traffic from LSR **U**' until IGP converges and provides a new primary next-hop, and ECMP or LFA backup next-hop, which may or may not be on the interface to LSR **U**'. At that point LSR **Z** will update the primary and backup ILMs in the data path.

The LDP uses the interface of either an ECMP next-hop or a LFA next-hop to the root LSR prefix, whichever is available, to program the backup ILM. ECMP next-hop and LFA next-hop are however mutually exclusive for a given prefix. IGP installs the ECMP next-hop in preference to an LFA next-hop for a prefix in the Routing Table Manager (RTM).

If one or more ECMP next-hops for the root LSR prefix exist, LDP picks the interface for the primary ILM based on the rules of mLDP FEC resolution specified in RFC 6388:

1. The candidate upstream LSRs are numbered from lower to higher IP address.
2. The following hash is performed: **H = (CRC32(Opaque Value)) modulo N**, where **N** is the number of upstream LSRs. The **Opaque Value** is the field identified in the P2MP FEC Element right after 'Opaque Length' field. The 'Opaque Length' indicates the size of the opaque value used in this calculation.
3. The selected upstream LSR **U** is the LSR that has the number **H**.

LDP then picks the interface for the backup ILM using the following new rules:

```
if (H + 1 < NUM_ECMP) {

// If the hashed entry is not last in the next-hops then pick up the next as backup.

backup = H + 1;

} else {

// Wrap around and pickup the first.

   backup = 1;

}
```

In some topologies, it is possible that none of ECMP or LFA next-hop will be found. In this case, LDP programs the primary ILM only.

7.19.3 Uniform Failover from Primary to Backup ILM

When LDP programs the primary ILM record in the data path, it provides the IOM with the Protect-Group Identifier (PG-ID) associated with this ILM and which identifies which upstream LSR is protected.

In order for the system to perform a fast switchover to the backup ILM in the fast path, LDP applies to the primary ILM uniform FRR failover procedures similar in concept to the ones applied to an NHLFE in the existing implementation of LDP FRR for unicast FECs. There are however important differences to note. LDP associates a unique Protect Group ID (PG–ID) to all mLDP FECs which have their primary ILM on any LDP interface pointing **to the same upstream LSR**. This PG-ID is assigned per upstream LSR regardless of the number of LDP interfaces configured to this LSR. As such this PG-ID is different from the one associated with unicast FECs and which is assigned to each downstream LDP interface and next-hop. If however a failure caused an interface to go down and also caused the LDP session to upstream peer to go down, both PG-IDs have their state updated in the IOM and thus the uniform FRR procedures will be triggered for both the unicast LDP FECs forwarding packets towards the upstream LSR and the mLDP FECs receiving packets from the same upstream LSR.

When the mLDP FEC is programmed in the data path, the primary and backup ILM record thus contain the PG-ID the FEC is associated with. The IOM also maintains a list of PG-IDs and a state bit which indicates if it is UP or DOWN. When the PG-ID state is UP the primary ILM for each mLDP FEC is open and will accept mLDP packets while the backup ILM is blocked and drops mLDP packets. LDP sends a PG-ID DOWN notification to IOM when it detects that the LDP session to the peer is gone down. This notification will cause the backup ILMs associated with this PG-ID to open and accept mLDP packets immediately. When IGP re-converges, an updated pair of primary and backup ILMs is downloaded for each mLDP FEC by LDP into the IOM with the corresponding PG-IDs.

If multiple LDP interfaces exist to the upstream LSR, a failure of one interface will bring down the link Hello adjacency on that interface but not the LDP session which is still associated with the remaining link Hello adjacencies. In this case, the upstream LSR updates in IOM the NHLFE for the mLDP FEC to use one of the remaining links. The switchover time in this case is not managed by the uniform failover procedures.

7.20 Multi-Area and Multi-Instance Extensions to LDP

In order to extend LDP across multiple areas of an IGP instance or across multiple IGP instances, the current standard LDP implementation based on RFC 3036 requires that all /32 prefixes of PEs be leaked between the areas or instances. This is because an exact match of the prefix in the routing table is required to install the prefix binding in the LDP Forwarding Information Base (FIB). Although a router will do this by default when configured as Area Border Router (ABR), this increases the convergence of IGP on routers when the number of PE nodes scales to thousands of nodes.

Multi-area and multi-instance extensions to LDP provide an optional behavior by which LDP installs a prefix binding in the LDP FIB by simply performing a longest prefix match with an aggregate prefix in the routing table (RIB). That way, the ABR will be configured to summarize the /32 prefixes of PE routers. This method is compliant to RFC 5283, *LDP Extension for Inter-Area Label Switched Paths (LSPs)*.

7.20.1 LDP Shortcut for BGP Next-Hop Resolution

LDP shortcut for BGP next-hop resolution shortcuts allow for the deployment of a 'route-less core' infrastructure on the 7750 SR and 7950 XRS. Many service providers either have or intend to remove the IBGP mesh from their network core, retaining only the mesh between routers connected to areas of the network that require routing to external routes.

Shortcuts are implemented by utilizing Layer 2 tunnels (that is, MPLS LSPs) as next hops for prefixes that are associated with the far end termination of the tunnel. By tunneling through the network core, the core routers forwarding the tunnel have no need to obtain external routing information and are immune to attack from external sources.

The tunnel table contains all available tunnels indexed by remote destination IP address. LSPs derived from received LDP /32 route FECs will automatically be installed in the table associated with the advertising router-ID when IGP shortcuts are enabled.

Evaluating tunnel preference is based on the following order in descending priority:

1. LDP /32 route FEC shortcut
2. Actual IGP next-hop

If a higher priority shortcut is not available or is not configured, a lower priority shortcut is evaluated. When no shortcuts are configured or available, the IGP next-hop is always used. Shortcut and next-hop determination is event driven based on dynamic changes in the tunneling mechanisms and routing states.

Refer to the *7450 ESS, 7750 SR, 7950 XRS, and VSR Unicast Routing Protocols Guide* for details on the use of LDP FEC and RSVP LSP for BGP Next-Hop Resolution.

7.20.2 LDP Shortcut for IGP Routes

The LDP shortcut for IGP route resolution feature allows forwarding of packets to IGP learned routes using an LDP LSP. When LDP shortcut is enabled globally, IP packets forwarded over a network IP interface will be labeled with the label received from the next-hop for the route and corresponding to the FEC-prefix matching the destination address of the IP packet. In such a case, the routing table will have the shortcut next-hop as the best route. If such a LDP FEC does not exist, then the routing table will have the regular IP next-hop and regular IP forwarding will be performed on the packet.

An egress LER advertises and maintains a FEC, label binding for each IGP learned route. This is performed by the existing LDP fec-originate capability.

7.20.2.1 LDP Shortcut Configuration

The user enables the use of LDP shortcut for resolving IGP routes by entering the global command **config>router>ldp-shortcut.**

This command enables forwarding of user IP packets and specified control IP packets using LDP shortcuts over all network interfaces in the system which participate in the IS-IS and OSPF routing protocols. The default is to disable the LDP shortcut across all interfaces in the system.

7.20.2.2 IGP Route Resolution

When LDP shortcut is enabled, LDP populates the RTM with next-hop entries corresponding to all prefixes for which it activated an LDP FEC. For a given prefix, two route entries are populated in 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 has preference over the regular IP next-hop for forwarding user packets and specified control packets over a given outgoing interface to the route next-hop.

The prior activation of the FEC by LDP is done by performing an exact match with an IGP route prefix in RTM. It can also be done by performing a longest prefix-match with an IGP route in RTM if the aggregate-prefix-match option is enabled globally in LDP.

This feature is not restricted to /32 FEC prefixes. However only /32 FEC prefixes will be populated in the CPM Tunnel Table for use as a tunnel by services.

All user packets and specified control packets for which the longest prefix match in RTM yields the FEC prefix will be forwarded over the LDP LSP. Currently, the control packets that could be forwarded over the LDP LSP are ICMP ping and UDP-traceroute. The following is an example of the resolution process.

Assume the egress LER advertised a FEC for some /24 prefix using the fec-originate command. At the ingress LER, LDP resolves the FEC by checking in RTM that an exact match exists for this prefix. Once LDP activated the FEC, it programs the NHLFE in the egress data path and the LDP tunnel information in the ingress data path tunnel table.

Next, LDP provides the shortcut route to RTM which will associate it with the same /24 prefix. There will be two entries for this /24 prefix, the LDP shortcut next-hop and the regular IP next-hop. The latter was used by LDP to validate and activate the FEC. RTM then resolves all user prefixes which succeed a longest prefix match against the /24 route entry to use the LDP LSP.

Assume now the aggregate-prefix-match was enabled and that LDP found a /16 prefix in RTM to activate the FEC for the /24 FEC prefix. In this case, RTM adds a new more specific route entry of /24 and has the next-hop as the LDP LSP but it will still not have a specific /24 IP route entry. RTM then resolves all user prefixes which succeed a longest prefix match against the /24 route entry to use the LDP LSP while all other prefixes which succeed a longest prefix-match against the /16 route entry will use the IP next-hop.

7.20.2.3 LDP Shortcut Forwarding Plane

Once LDP activated a FEC for a given prefix and programmed RTM, it also programs the ingress Tunnel Table in forwarding engine with the LDP tunnel information.

When an IPv4 packet is received on an ingress network interface, or a subscriber IES interface, or a regular IES interface, the lookup of the packet by the ingress forwarding engine will result in the packet being sent labeled with the label stack corresponding to the NHLFE of the LDP LSP when the preferred RTM entry corresponds to an LDP shortcut.

If the preferred RTM entry corresponds to an IP next-hop, the IPv4 packet is forwarded unlabeled.

7.20.3 ECMP Considerations

When ECMP is enabled and multiple equal-cost next-hops exit for the IGP route, the ingress forwarding engine sprays the packets for this route based on hashing routine currently supported for IPv4 packets.

When the preferred RTM entry corresponds to an LDP shortcut route, spraying will be performed across the multiple next-hops for the LDP FEC. The FEC next-hops can either be direct link LDP neighbors or T-LDP neighbors reachable over RSVP LSPs in the case of LDP-over-RSVP but not both. This is as per ECMP for LDP in 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.

7.20.4 Disabling TTL Propagation in an LSP Shortcut

This feature provides the option for disabling TTL propagation from a transit or a locally generated IP packet header into the LSP label stack when an LDP LSP is used as a shortcut for BGP next-hop resolution, a static-route next-hop resolution, or for an IGP route resolution.

A transit packet is a packet received from an IP interface and forwarded over the LSP shortcut at ingress LER.

A locally-generated IP packet is any control plane packet generated from the CPM and forwarded over the LSP shortcut at ingress LER.

TTL handling can be configured for all LDP LSP shortcuts originating on an ingress LER using the following global commands:

config>router>ldp>[**no**] **shortcut-transit-ttl-propagate**

config>router>ldp>[**no**] **shortcut-local-ttl-propagate**

These commands apply to all LDP LSPs which are used to resolve static routes, BGP routes, and IGP routes.

When the **no** form of the above command is enabled for local packets, TTL propagation is disabled on all locally generated IP packets, including ICMP Ping, traceroute, and OAM packets that are destined to a route that is resolved to the LSP shortcut. In this case, a TTL of 255 is programmed onto the pushed label stack. This is referred to as pipe mode.

Similarly, when the **no** form is enabled for transit packets, TTL propagation is disabled on all IP packets received on any IES interface and destined to a route that is resolved to the LSP shortcut. In this case, a TTL of 255 is programmed onto the pushed label stack.

7.21 LDP Graceful Handling of Resource Exhaustion

This feature enhances the behavior of LDP when a data path or a CPM resource required for the resolution of a FEC is exhausted. In prior releases, the LDP module shuts down. The user is required to fix the issue causing the FEC scaling to be exceeded and to restart the LDP module by executing the **unshut** command.

7.21.1 LDP Base Graceful Handling of Resources

This feature implements a base graceful handling capability by which the LDP interface to the peer, or the targeted peer in the case of Targeted LDP (T-LDP) session, is shutdown. If LDP tries to resolve a FEC over a link or a targeted LDP session and it runs out of data path or CPM resources, it will bring down that interface or targeted peer which will bring down the Hello adjacency over that interface to the resolved link LDP peer or to the targeted peer. The interface is brought down in LDP context only and is still available to other applications such as IP forwarding and RSVP LSP forwarding.

Depending of what type of resource was exhausted, the scope of the action taken by LDP will be different. Some resource such as NHLFE have interface local impact, meaning that only the interface to the downstream LSR which advertised the label is shutdown. Some resources such as ILM have global impact, meaning that they will impact every downstream peer or targeted peer which advertised the FEC to the node. The following are examples to illustrate this.

- For NHLFE exhaustion, one or more interfaces or targeted peers, if the FEC is ECMP, will be shut down. ILM is maintained as long as there is at least one downstream for the FEC for which the NHLFE has been successfully programmed.
- For an exhaustion of an ILM for a unicast LDP FEC, all interfaces to peers or all target peers which sent the FEC will be shutdown. No deprogramming of data path is required since FEC is not programmed.
- An exhaustion of ILM for an mLDP FEC can happen during primary ILM programming, MBB ILM programming, or multicast upstream FRR backup ILM programming. In all cases, the P2MP index for the mLDP tree is deprogrammed and the interfaces to each downstream peer which sent a Label Mapping message associated with this ILM are shutdown.

After the user has taken action to free resources up, he/she will require manually unshut the interface or the targeted peer to bring it back into operation. This then re-establishes the Hello adjacency and resumes the resolution of FECs over the interface or to the targeted peer.

Detailed guidelines for using the feature and for troubleshooting a system which activated this feature are provided in the following sections.

This behavior is the default behavior and interoperates with the SR OS based LDP implementation and any other third party LDP implementation.

The following data path resources can trigger this mechanism:

- NHLFE
- ILM
- Label-to-NHLFE (LTN)
- Tunnel Index
- P2MP Index

The following CPM resources can trigger this mechanism:

- Label allocation

7.22 LDP Enhanced Graceful Handling of Resources

This feature is an enhanced graceful handling capability which is supported only among SR OS based implementations. If LDP tries to resolve a FEC over a link or a targeted session and it runs out of data path or CPM resources, it will put the LDP/T-LDP session into overload state. As a result, it will release to its LDP peer the labels of the FECs which it could not resolve and will also send an LDP notification message to all LDP peers with the new status load of overload for the FEC type which caused the overload. The notification of overload is per FEC type, that is, unicast IPv4, P2MP mLDP and so on, and not per individual FEC. The peer which caused the overload and all other peers will stop sending any new FECs of that type until this node updates the notification stating that it is no longer in overload state for that FEC type. FECs of this type previously resolved and other FEC types to this peer and all other peers will continue to forward traffic normally.

After the user has taken action to free resources up, he/she will require manually clear the overload state of the LDP/T-LDP sessions towards its peers.

The enhanced mechanism will be enabled instead of the base mechanism only if both LSR nodes advertise this new LDP capability at the time the LDP session is initialized. Otherwise, they will continue to use the base mechanism.

This feature operates among SR OS LSR nodes using a couple of private vendor LDP capabilities:

- The first one is the LSR Overload Status TLV to signal or clear the overload condition.
- The second one is the Overload Protection Capability Parameter which allows LDP peers to negotiate the use or not of the overload notification feature and hence the enhanced graceful handling mechanism.

When interoperating with an LDP peer which does not support the enhanced resource handling mechanism, the router reverts automatically to the default base resource handling mechanism.

The following are the details of the mechanism.

7.22.1 LSR Overload Notification

When an upstream LSR is overloaded for a FEC type, it notifies one or more downstream peer LSRs that it is overloaded for the FEC type.

When a downstream LSR receives overload status ON notification from an upstream LSR, it does not send further label mappings for the specified FEC type. When a downstream LSR receives overload OFF notification from an upstream LSR, it sends pending label mappings to the upstream LSR for the specified FEC type.

This feature introduces a new TLV referred to as *LSR Overload Status TLV*, shown below. This TLV is encoded using vendor proprietary TLV encoding as per RFC 5036. It uses a TLV type value of 0x3E02 and the Timetra OUI value of 0003FA.

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|U|F| Overload Status TLV Type  |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  Timetra OUI  = 0003FA                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|S|                        Reserved                             |

where:
   U-bit: Unknown TLV bit, as described in RFC 5036. The value MUST
   be 1 which means if unknown to receiver then receiver should ignore

   F-bit: Forward unknown TLV bit, as described in RFC RFC5036. The value
   of this bit MUST be 1 since a LSR overload TLV is sent only between
   two immediate LDP peers, which are not forwarded.

   S-bit: The State Bit. It indicates whether the sender is setting the
   LSR Overload Status ON or OFF. The State Bit value is used as
   follows:

   1 - The TLV is indicating LSR overload status as ON.

   0 - The TLV is indicating LSR overload status as OFF.
```

When a LSR that implements the procedures defined in this document generates LSR overload status, it must send LSR Overload Status TLV in a LDP Notification Message accompanied by a FEC TLV. The FEC TLV must contain one Typed Wildcard FEC TLV that specifies the FEC type to which the overload status notification applies.

The feature in this document re-uses the Typed Wildcard FEC Element which is defined in RFC 5918.

7.22.2 LSR Overload Protection Capability

To ensure backward compatibility with procedures in RFC 5036 an LSR supporting Overload Protection need means to determine whether a peering LSR supports overload protection or not.

An LDP speaker that supports the LSR Overload Protection procedures as defined in this document must inform its peers of the support by including a LSR Overload Protection Capability Parameter in its initialization message. The Capability parameter follows the guidelines and all Capability Negotiation Procedures as defined in RFC 5561. This TLV is encoded using vendor proprietary TLV encoding as per RFC 5036. It uses a TLV type value of 0x3E03 and the Timetra OUI value of 0003FA.

```
       0                   1                   2                   3
      0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |U|F| LSR Overload Cap TLV Type |            Length             |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                  Timetra OUI = 0003FA                         |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |S| Reserved    |
      +-+-+-+-+-+-+-+-+
Where:

   U and F bits : MUST be 1 and 0 respectively as per section 3 of LDP
   Capabilities [RFC5561].

   S-bit : MUST be 1 (indicates that capability is being advertised).
```

7.22.3 Procedures for LSR overload protection

The procedures defined in this document apply only to LSRs that support Downstream Unsolicited (DU) label advertisement mode and Liberal Label Retention Mode. An LSR that implements the LSR overload protection follows the following procedures:

1. An LSR must not use LSR overload notification procedures with a peer LSR that has not specified LSR Overload Protection Capability in Initialization Message received from the peer LSR.
2. When an upstream LSR detects that it is overloaded with a FEC type then it must initiate an LDP notification message with the S-bit ON in LSR Overload Status TLV and a FEC TLV containing the Typed Wildcard FEC Element for the specified FEC type. This message may be sent to one or more peers.
3. After it has notified peers of its overload status ON for a FEC type, the overloaded upstream LSR can send Label Release for a set of FEC elements to respective downstream LSRs to off load its LIB to below a certain watermark.
4. When an upstream LSR that was previously overloaded for a FEC type detects that it is no longer overloaded, it must send an LDP notification message with the S-bit OFF in LSR Overload Status TLV and FEC TLV containing the Typed Wildcard FEC Element for the specified FEC type.

5. When an upstream LSR has notified its peers that it is overloaded for a FEC type, then a downstream LSR must not send new label mappings for the specified FEC type to the upstream LSR.
6. When a downstream LSR receives LSR overload notification from a peering LSR with status OFF for a FEC type then the receiving LSR must send any label mappings for the FEC type which were pending to the upstream LSR for which are eligible to be sent now.
7. When an upstream LSR is overloaded for a FEC type and it receives Label Mapping for that FEC type from a downstream LSR then it can send Label Release to the downstream peer for the received Label Mapping with LDP Status Code as *No_Label_Resource*s as defined in RFC 5036.

7.23 LDP-IGP Synchronization

The SR OS supports the synchronization of an IGP and LDP based on a solution described in RFC 5443, which consists of setting the cost of a restored link to infinity to give both the IGP and LDP time to converge. When a link is restored after a failure, the IGP sets the link cost to infinity and advertises it. The actual value advertised in OSPF is 0xFFFF (65535). The actual value advertised in an IS-IS regular metric is 0x3F (63) and in IS-IS wide-metric is 0xFFFFFE (16777214). This synchronization feature is not supported on RIP interfaces.

When the LDP synchronization timer subsequently expires, the actual cost is put back and the IGP readvertises it and uses it at the next SPF computation. The LDP synchronization timer is configured using the following command:

config>router>if> [no] ldp-sync-timer *seconds*

The SR OS also supports an LDP End of LIB message, as defined in RFC 5919, that allows a downstream node to indicate to its upstream peer that it has advertised its entire label information base. The effect of this on the IGP-LDP synchronization timer is described below.

If an interface belongs to both IS-IS and OSPF, a physical failure will cause both IGPs to advertise an infinite metric and to follow the IGP-LDP synchronization procedures. If only one IGP bounces on this interface or on the system, then only the affected IGP advertises the infinite metric and follows the IGP-LDP synchronization procedures.

Next, an LDP Hello adjacency is brought up with the neighbor. The LDP synchronization timer is started by the IGP when the LDP session to the neighbor is up over the interface. This is to allow time for the label-FEC bindings to be exchanged.

When the LDP synchronization timer expires, the link cost is restored and is readvertised. The IGP will announce a new best next hop and LDP will use it if the label binding for the neighbor's FEC is available.

If the user changes the cost of an interface, the new value is advertised at the next flooding of link attributes by the IGP. However, if the LDP synchronization timer is still running, the new cost value will only be advertised after the timer expires. The new cost value will also be advertised after the user executes any of the following commands:

- **tools>perform>router>isis>ldp-sync-exit**
- **tools>perform>router>ospf>ldp-sync-exit**
- **config>router>if>no ldp-sync-timer**

- **config>router>ospf>disable-ldp-sync**
- **router>isis>disable-ldp-sync**

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.

If parallel links exist to the same neighbor, then the bindings and services should remain up as long as there is one interface that is up. However, the user-configured LDP synchronization timer still applies on the interface that failed and was restored. In this case, the router will only consider this interface for forwarding after the IGP readvertises its actual cost value.

The LDP End of LIB message is used by a node to signal completion of label advertisements, using a FEC TLV with the Typed Wildcard FEC element for all negotiated FEC types. This is done even if the system has no label bindings to advertise. The SR OS also supports the Unrecognized Notification TLV (RFC 5919) that indicates to a peer node that it will ignore unrecognized status TLVs. This indicates to the peer node that it is safe to send End of LIB notifications even if the node is not configured to process them.

The behavior of a system that receives an End of LIB status notification is configured through the CLI on a per-interface basis:

config>router>if>[no] ldp-sync-timer *seconds* **end-of-lib**

If the **end-of lib** option is not configured, then the LDP synchronization timer is started when the LDP Hello adjacency comes up over the interface, as described above. Any received End of LIB LDP messages are ignored.

If the **end-of-lib** option is configured, then the system will behave as follows on the receive side:

- The **ldp-sync-timer** is started.
- If LDP End of LIB Typed Wildcard FEC messages are received for every FEC type negotiated for a given session to an LDP peer for that IGP interface, the **ldp-sync-timer** is terminated and processing proceeds as if the timer had expired, that is, by restoring the IGP link cost.
- If the **ldp-sync-timer** expires before the LDP End of LIB messages are received for every negotiated FEC type, then the system restores the IGP link cost.
- The receive side will drop any unexpected End of LIB messages.

If the **end-of-lib** option is configured, then the system will also send out an End of LIB message for prefix and P2MP FECs once all FECs are sent for all peers that have advertised the Unrecognized Notification Capability TLV.

See the *7450 ESS, 7750 SR, 7950 XRS, and VSR Router Configuration Guide* for the CLI command descriptions for LDP-IGP Synchronization.

7.24 MLDP Resolution using Multicast RTM

When unicast services use IGP shortcuts, IGP shortcut next-hops are installed in the RTM. Therefore, for multicast P2MP MLDP, the leaf node will resolve the root using these IGP shortcuts. Currently MLDP can not be resolved using IGP shortcuts. To avoid this, MLDP does a lookup in the multicast RTM. IGP shortcuts are not installed in MRTM. The command **configure router ldp resolve-root-using** forces MLDP do next-hop lookups in the RTM or MRTM.

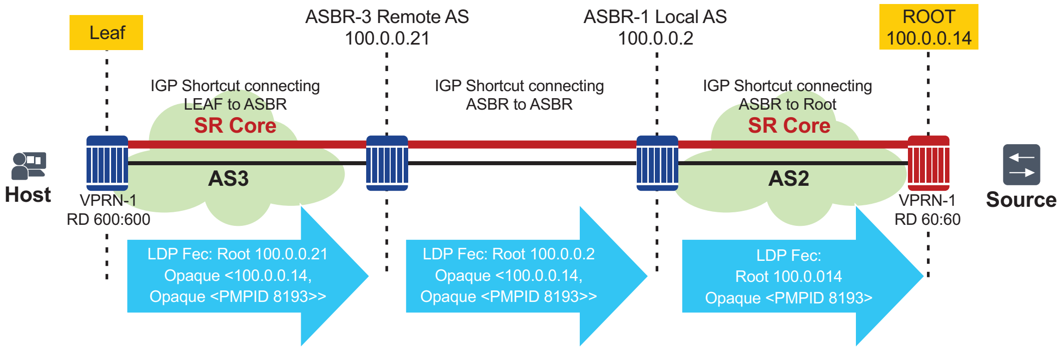
By default, the **configure router ldp resolve-root-using** command is set to **ucast-rtm** and MLDP uses the unicast RTM for resolution of the FEC in all cases. When MLDP uses the unicast RTM to resolve the FEC, it will not resolve the FEC if its next hop is resolved using an IGP shortcut.

To force MLDP resolution to use the multicast RTM, use the **configure router ldp resolve-root-using mcast-rtm** command. When this command is enabled:

- For FEC resolution using IGP, static or local, the ROOT in this FEC is resolved using the multicast RTM.
- A FEC being resolved using BGP is recursive, so the FEC next-hop (ASBR/ABR) is resolved using the multicast RTM first and, if this fails, it is resolved using the unicast RTM. This next-hop needs to be recursively resolved again using IGP/Static-Route or Local, this second resolution (recursive resolution) uses the multicast RTM only; see Figure 110.
- When **configure router ldp resolve-root-using ucast-rtm** is set, MLDP uses the unicast RTM to resolve the FEC and will not resolve the FEC if its next hop is resolved using an IGP shortcut.

For inter-AS or intra-AS, IGP shortcuts are limited to each AS or area connecting LEAF to ASBR, ASBR to ASBR, or ASBR to ROOT.

Figure 110 **Recursive FEC Behavior**



	LEAF	ASBR-3	ASBR-1
Inner Root 100.0.0.14 (First resolution using BGR to ASBR-3)	M-RTM first U-RTM second	M-RTM first U-RTM second	*
Inner Root recursive resolution of ASBR-3 100.0.0.21 for Inner Root	M-RTM		
Inner Root recursive resolution of ASBR-1 100.0.0.2 for Inner Root		M-RTM	
Outer Root 100.0.0.14			M-RTM

sw0442

In Figure 110, the FEC between LEAF and ASBR-3 is resolved using an IGP shortcut. When the **configure ldp resolve-root-using** is set to **mcast-rtm**, the inner Root 100.0.0.14 will be resolved using the multicast RTM first. If the multicast RTM lookup fails, then a second lookup for 100.0.0.14 is done in the unicast RTM. Resolution of 100.0.0.14 results in a next-hop of 100.0.0.21 which is ASBR-3, as such ASBR-3 100.0.0.21 is resolved only using multicast RTM when **mcast-rtm** is enabled.

7.24.1 Other Considerations for Multicast RTM MLDP Resolution

When **configure ldp resolve-root-using** is set to **mcast-rtm** and then changed to **ucast-rtm** there is traffic disruption. If MoFRR is enabled, by toggling from **mcast-rtm** to **ucast-rtm** (or the other way around) the MoFRR is not utilized. In fact, MoFRR is torn down and re-established using the new routing table.

The **mcast-rtm** only has a local effect. All MLDP routing calculations on this specific node will use MRTM and not RTM.

If **mcast-rtm** is enabled, all MLDP functionality will be based on MRTM. This includes MoFRR, ASBR-MoFRR, policy-based SPMSI, and non-segmented inter-AS.

7.25 Bidirectional Forwarding Detection for LDP LSPs

Bidirectional forwarding detection (BFD) for MPLS LSPs monitors the LSP between its LERs, irrespective of how many LSRs the LSP may traverse. This enables the detection of faults that are local to individual LSPs, whether or not they also affect forwarding for other LSPs or IP packet flows. BFD is ideal for monitoring LSPs that carry high-value services, where detection of forwarding failures in a minimal amount of time is critical. The system will raise an SNMP trap, as well as indicate the BFD session state in **show** and **tools dump** commands if an LSP BFD session goes down.

SR OS supports LSP BFD on RSVP and LDP LSPs. See MPLS and RSVP for information on using LSP BFD on RSVP LSPs. BFD packets are encapsulated in an MPLS label stack corresponding to the FEC that the BFD session is associated with, as described in RFC 5884, Section 7. SR OS does not support the monitoring of multiple ECMP paths that are associated with the same LDP FEC which is using multiple LSP BFD sessions simultaneously. However, LSP BFD still provides continuity checking for paths associated with a target FEC. LDP provides a single path to LSP BFD, corresponding with the first resolved lower if index next-hop, and the first resolved lower tid index for LDP-over-RSVP cases. The path may potentially change over the lifetime of the FEC, based on resolution changes. The system tracks the changing path and maintains the LSP BFD session.

Since LDP LSPs are unidirectional, a routed return path is used for the BFD control packets traveling from the egress LER to the ingress LER.

7.25.1 Bootstrapping and Maintaining LSP BFD Sessions

A BFD session on an LSP is bootstrapped using LSP ping. LSP ping is used to exchange the local and remote discriminator values to use for the BFD session for a particular MPLS LSP or FEC.

The process for bootstrapping an LSP BFD session for LDP is the same as for RSVP, as described in Bidirectional Forwarding Detection for MPLS LSPs.

SR OS supports the sending of periodic LSP ping messages on an LSP for which LSP BFD has been configured, as specified in RFC 5884. The ping messages are sent, along with the bootstrap TLV, at a configurable interval for LSPs on which **bfd-enable** has been configured. The default interval is 60 s, with a maximum interval of 300 s. The LSP ping echo request message uses the system IP address as the default source address. An alternative source address consisting of any routable address that is local to the node may be configured, and will be used if the local system IP address is not routable from the far-end node.

Note: SR OS does not take any action if a remote system fails to respond to a periodic LSP ping message. However, when the **show**>**test-oam**>**lsp-bfd** command is executed, it will display a return code of zero and a replying node address of 0.0.0.0 if the periodic LSP ping times out.

The periodic LSP ping interval is configured using the **config**>**router**>**ldp**>**lsp-bfd** *prefix-list*>**lsp-ping-interval** *seconds* command.

Configuring an LSP ping interval of 0 disables periodic LSP ping for LDP FECs matching the specified prefix list. The **no lsp-ping-interval** command reverts to the default of 60 s.

LSP BFD sessions are recreated after a high availability switchover between active and standby CPMs. However, some disruption may occur to LSP ping due to LSP BFD.

At the head end of an LSP, sessions are bootstrapped if the local and remote discriminators are not known. The sessions will experience jitter at 0 to 25% of a retry time of 5 seconds. A side effect is that the following current information will be lost from an active **show test-oam lsp-bfd** display:

- Replying Node
- Latest Return Code
- Latest Return SubCode
- Bootstrap Retry Count
- Tx Lsp Ping Requests
- Rx Lsp Ping Replies

If the local and remote discriminators are known, the system immediately begins generating periodic LSP pings. The pings will experience jitter at 0 to 25% of the **lsp-ping-interval** time of 60 to 300 seconds. The **lsp-ping-interval** time is synchronized across by LSP BFD. A side effect is that the following current information will be lost from an active **show test-oam lsp-bfd** display:

- Replying Node

- Latest Return Code
- Latest Return SubCode
- Bootstrap Retry Count
- Tx Lsp Ping Requests
- Rx Lsp Ping Replies

At the tail end of an LSP, sessions are recreated on the standby CPM following a switchover. A side effect is that the following current information will be lost from an active **tools dump test-oam lsp-bfd tail** display:

- handle
- seqNum
- rc
- rsc

Any new, incoming bootstrap requests will be dropped until LSP BFD has become active. When LSP BFD has finished becoming active, new bootstrap requests will be considered.

7.25.2 BFD Configuration on LDP LSPs

LSP BFD is configured for LDP using the following CLI commands:

CLI Syntax:

```
config
   router
      ldp
         [no] lsp-bfd prefix-list-name
            priority priority-level
            no priority
            bfd-template bfd-template-name
            no bfd-template
            source-address ip-address
            no source-address
            [no] bfd-enable
            lsp-ping-interval seconds
            no lsp-ping-interval
            exit
```

The **lsp-bfd** command creates the context for LSP BFD configuration for a set of LDP LSPs with a FEC matching the one defined by the *prefix-list-name* parameter. The default is **no lsp-bfd**. Configuring **no lsp-bfd** for a specified prefix list will remove LSP BFD for all matching LDP FECs except those that also match another LSP BFD prefix list. The *prefix-list-name* parameter refers to a named prefix list configured in the **config**>**router**>**policy-options** context.

Up to 16 instances of LSP BFD can be configured under LDP in the base router instance.

The optional **priority** command configures a priority value that is used to order the processing if multiple prefix lists are configured. The default value is 1.

If more than one prefix in a prefix list, or more than one prefix list, contains a prefix that corresponds to the same LDP FEC, then the system will test the prefix against the configured prefix lists in the following order:

1. numerically by *priority-level*
2. alphabetically by *prefix-list-name*

The system will use the first matching configuration, if one exists.

If an LSP BFD is removed for a prefix list, but there remains another LSP BFD configuration with a prefix list match, then any FECs matched against that prefix will be rematched against the remaining prefix list configurations in the same manner as described above.

A non-existent prefix list is equivalent to an empty prefix list. When a prefix list is created and populated with prefixes, LDP will match its FECs against that prefix list. It is not necessary to configure a named prefix list in the **config**>**router**>**policy-options** context before specifying a prefix list using the **config**>**router**>**ldp**>**lsp-bfd** command.

If a prefix list contains a longest match corresponding to one or more LDP FECs, the BFD configuration is applied to all of the matching LDP LSPs.

Only /32 IPv4 and /128 IPv6 host prefix FECs will be considered for BFD. BFD on PW FECs uses VCCV BFD.

The **source-address** command is used to configure the source address of periodic LSP ping packets and BFD control packets for LSP BFD sessions associated with LDP prefixes in the prefix list. The default value is the system IP address. If the system IP address is not routable from the far-end node of the BFD session, then an alternative routable IP address local to the source node should be used.

The system will not initialize an LSP BFD session if there is a mismatch between the address family of the source address and the address family of the prefix in the prefix list.

If the system has both IPv4 and IPv6 system IP addresses, and the **source-address** command is not configured, then the system will use a source address of the matching address family for IPv4 and IPv6 prefixes in the prefix list.

The **bfd-template** command applies the specified BFD template to the BFD sessions for LDP LSPs with FECs that match the prefix list. The default is **no bfd-template**. The named BFD template must first be configured using the **config**>**router**>**bfd**>**bfd-template** command before it can be referenced by LSP BFD, otherwise a CLI error is generated. The minimum receive interval and transmit interval supported for LSP BFD on LDP LSPs is 1 second.

The **bfd-enable** command enables BFD on the LDP LSPs with FECs that match the prefix list.

7.26 User Guidelines and Troubleshooting Procedures

7.26.1 Common Procedures

When troubleshooting a LDP resource exhaustion situation on an LSR, the user must first determine which of the LSR and its peers supports the enhanced handling of resources. This is done by checking if the local LSR or its peers advertised the LSR Overload Protection Capability:

```
*A:Sim>config>router>ldp# show router ldp status

===============================================================================
LDP Status for IPv4 LSR ID 0.0.0.0
               IPv6 LSR ID ::
===============================================================================
Created at         : 01/08/19 17:57:06
Last Change        : 01/08/19 17:57:06
Admin State        : Up
IPv4 Oper State    : Down                 IPv6 Oper State      : Down
IPv4 Down Time     : 0d 00:12:58          IPv6 Down Time       : 0d 00:12:58
IPv4 Oper Down Rea*: systemIpDown         IPv6 Oper Down Reason: systemIpDown
IPv4 Oper Down Eve*: 0                    IPv6 Oper Down Events: 0
Tunn Down Damp Time: 3 sec                Weighted ECMP        : Disabled
Label Withdraw Del*: 0 sec                Implicit Null Label  : Disabled
Short. TTL Local   : Enabled              Short. TTL Transit   : Enabled
ConsiderSysIPInGep : Disabled
Imp Ucast Policies :                      Exp Ucast Policies   :
    pol1                                      none
Imp Mcast Policies :
    pol1
    policy2
    policy-3
    policy-four
    pol-five
Tunl Exp Policies  : None                 Tunl Imp Policies    : None
FRR                : Disabled             Mcast Upstream FRR   : Disabled
Mcast Upst ASBR FRR: Disabled
```

7.26.2 Base Resource Handling Procedures

Step 1

If the peer OR the local LSR does not support the Overload Protection Capability it means that the associated adjacency [interface/peer] will be brought down as part of the base resource handling mechanism.

The user can determine which interface or targeted peer was shut down, by applying the following commands:

- [**show router ldp interface resource-failures**]

- [**show router ldp targ-peer resource-failures**]

```
show router ldp interface resource-failures
===============================================================================
LDP Interface Resource Failures
===============================================================================
srl                                     srr
sru4                                    sr4-1-5-1
===============================================================================

show router ldp targ-peer resource-failures
===============================================================================
LDP Peers Resource Failures
===============================================================================
10.20.1.22                              192.168.1.3
===============================================================================
```

A trap is also generated for each interface or targeted peer:

```
16 2013/07/17 14:21:38.06 PST MINOR: LDP #2003 Base LDP Interface Admin State
"Interface instance state changed - vRtrID: 1, Interface sr4-1-5-1, administrati
ve state: inService, operational state: outOfService"

13 2013/07/17 14:15:24.64 PST MINOR: LDP #2003 Base LDP Interface Admin State
"Interface instance state changed - vRtrID: 1, Peer 10.20.1.22, administrative s
tate: inService, operational state: outOfService"
```

The user can then check that the base resource handling mechanism has been applied to a specific interface or peer by running the following show commands:

- [**show router ldp interface detail**]

- [**show router ldp targ-peer detail**]

```
    show router ldp interface detail
===============================================================================
LDP Interfaces (Detail)
===============================================================================
-------------------------------------------------------------------------------
Interface "sr4-1-5-1"
-------------------------------------------------------------------------------
Admin State        : Up                   Oper State       : Down
Oper Down Reason   : noResources  <----- //link LDP resource exhaustion handled
Hold Time          : 45                   Hello Factor     : 3
Oper Hold Time     : 45
Hello Reduction    : Disabled             Hello Reduction *: 3
Keepalive Timeout  : 30                   Keepalive Factor : 3
Transport Addr     : System               Last Modified    : 07/17/13 14:21:38
```

```
Active Adjacencies : 0
Tunneling          : Disabled
Lsp Name           : None
Local LSR Type     : System
Local LSR          : None
BFD Status         : Disabled
Multicast Traffic  : Enabled
-------------------------------------------------------------------------------

show router ldp discovery interface "sr4-1-5-1" detail
===============================================================================
LDP Hello Adjacencies (Detail)
===============================================================================
-------------------------------------------------------------------------------
Interface "sr4-1-5-1"
-------------------------------------------------------------------------------
Local Address      : 192.168.2.110      Peer Address        : 192.168.0.2
Adjacency Type     : Link               State               : Down
===============================================================================


show router ldp targ-peer detail
===============================================================================
LDP Peers (Detail)
===============================================================================
-------------------------------------------------------------------------------
Peer 10.20.1.22
-------------------------------------------------------------------------------
Admin State        : Up              Oper State           : Down
Oper Down Reason   : noResources     <----- // T-LDP resource exhaustion handled
Hold Time          : 45              Hello Factor         : 3
Oper Hold Time     : 45
Hello Reduction    : Disabled        Hello Reduction Fact*: 3
Keepalive Timeout  : 40              Keepalive Factor     : 4
Passive Mode       : Disabled        Last Modified        : 07/17/13 14:15:24
Active Adjacencies : 0               Auto Created         : No
Tunneling          : Enabled
Lsp Name           : None
Local LSR          : None
BFD Status         : Disabled
Multicast Traffic  : Disabled
-------------------------------------------------------------------------------

show router ldp discovery peer 10.20.1.22 detail
===============================================================================
LDP Hello Adjacencies (Detail)
===============================================================================
-------------------------------------------------------------------------------
Peer 10.20.1.22
-------------------------------------------------------------------------------
Local Address      : 192.168.1.110      Peer Address        : 10.20.1.22
Adjacency Type     : Targeted           State               : Down   <-----
//T-LDP resource exhaustion handled
===============================================================================
```

Step 2

Besides interfaces and targeted peer, locally originated FECs may also be put into overload. These are the following:

- unicast fec-originate pop

- multicast local static p2mp-fec type=1 [on leaf LSR]

- multicast local Dynamic p2mp-fec type=3 [on leaf LSR]

The user can check if only remote and/or local FECs have been set in overload by the resource base resource exhaustion mechanism using the following command:

- [tools dump router ldp instance]

The relevant part of the output is described below:

```
{...... snip......}
Num OLoad Interfaces:      4     <----- //#LDP interfaces resource in exhaustion
Num Targ Sessions:         72          Num Active Targ Sess:  62
Num OLoad Targ Sessions:   7     <----- //#T-LDP peers in resource exhaustion
Num Addr FECs Rcvd:        0           Num Addr FECs Sent:    0
Num Addr Fecs OLoad:       1     <----- //# of local/remote unicast FECs in Overload
Num Svc FECs Rcvd:         0           Num Svc FECs Sent:     0
Num Svc FECs OLoad:        0     <----- // # of local/
remote service Fecs in Overload
Num mcast FECs Rcvd:       0           Num Mcast FECs Sent:   0
Num mcast FECs OLoad:      0     <----- // # of local/
remote multicast Fecs in Overload
{...... snip......}
```

When at least one local FEC has been set in overload the following trap will occur:

```
23 2013/07/17 15:35:47.84 PST MINOR: LDP #2002 Base LDP Resources Exhausted
"Instance
 state changed - vRtrID: 1, administrative state: inService, operationa l state:
 inService"
```

Step 3

After the user has detected that at least, one link LDP or T-LDP adjacency has been brought down by the resource exhaustion mechanism, he/she must protect the router by applying one or more of the following to free resources up:

- Identify the source for the [unicast/multicast/service] FEC flooding.
- Configure the appropriate [import/export] policies and/or delete the excess [unicast/multicast/service] FECs not currently handled.

Step 4

Next, the user has to manually attempt to clear the overload (no resource) state and allow the router to attempt to restore the link and targeted sessions to its peer.

Note: Because of the dynamic nature of FEC distribution and resolution by LSR nodes, one cannot predict exactly which FECs and which interfaces or targeted peers will be restored after performing the following commands if the LSR activates resource exhaustion again.

One of the following commands can be used:

- [**clear router ldp resource-failures**]

  - Clears the overload state and attempt to restore adjacency and session for LDP interfaces and peers.
  - Clear the overload state for the local FECs.

- [**clear router ldp interface ifName**]

- [**clear router ldp peer peerAddress]**

  - Clears the overload state and attempt to restore adjacency and session for LDP interfaces and peers.
  - These two commands ***DO NOT*** Clear the overload state for the local FECs.

7.26.3 Enhanced Resource Handling Procedures

Step 1

If the peer and the local LSR do support the Overload Protection Capability it means that the LSR will signal the overload state for the FEC type which caused the resource exhaustion as part of the enhanced resource handling mechanism.

In order to verify if the local router has received or sent the overload status TLV, perform the following:

```
-[show router ldp session detail]
show router ldp session 192.168.1.1 detail
-------------------------------------------------------------------------------
Session with Peer 192.168.1.1:0, Local 192.168.1.110:0
-------------------------------------------------------------------------------
Adjacency Type          : Both           State                  : Established
Up Time                 : 0d 00:05:48
Max PDU Length          : 4096           KA/Hold Time Remaining : 24
Link Adjacencies        : 1              Targeted Adjacencies   : 1
Local Address           : 192.168.1.110  Peer Address           : 192.168.1.1
Local TCP Port          : 51063          Peer TCP Port          : 646
Local KA Timeout        : 30             Peer KA Timeout        : 45
Mesg Sent               : 442            Mesg Recv              : 2984
FECs Sent               : 16             FECs Recv              : 2559
Addrs Sent              : 17             Addrs Recv             : 1054
```

```
GR State              : Capable        Label Distribution     : DU
Nbr Liveness Time     : 0              Max Recovery Time      : 0
Number of Restart     : 0              Last Restart Time      : Never
P2MP                  : Capable        MP MBB                 : Capable
Dynamic Capability    : Not Capable    LSR Overload           : Capable
Advertise             : Address/Servi* BFD Operational Status : inService
Addr FEC OverLoad Sent : Yes           Addr FEC OverLoad Recv : No     <----
// this LSR sent overLoad for unicast FEC type to peer
Mcast FEC Overload Sent: No            Mcast FEC Overload Recv: No
Serv FEC Overload Sent : No            Serv FEC Overload Recv : No
-------------------------------------------------------------------------------

show router ldp session 192.168.1.110 detail
-------------------------------------------------------------------------------
Session with Peer 192.168.1.110:0, Local 192.168.1.1:0
-------------------------------------------------------------------------------
Adjacency Type        : Both           State                  : Established
Up Time               : 0d 00:08:23
Max PDU Length        : 4096           KA/Hold Time Remaining : 21
Link Adjacencies      : 1              Targeted Adjacencies   : 1
Local Address         : 192.168.1.1    Peer Address           : 192.168.1.110
Local TCP Port        : 646            Peer TCP Port          : 51063
Local KA Timeout      : 45             Peer KA Timeout        : 30
Mesg Sent             : 3020           Mesg Recv              : 480
FECs Sent             : 2867           FECs Recv              : 16
Addrs Sent            : 1054           Addrs Recv             : 17
GR State              : Capable        Label Distribution     : DU
Nbr Liveness Time     : 0              Max Recovery Time      : 0
Number of Restart     : 0              Last Restart Time      : Never
P2MP                  : Capable        MP MBB                 : Capable
Dynamic Capability    : Not Capable    LSR Overload           : Capable
Advertise             : Address/Servi* BFD Operational Status : inService
Addr FEC OverLoad Sent : No            Addr FEC OverLoad Recv : Yes     <----
// this LSR received overLoad for unicast FEC type from peer
Mcast FEC Overload Sent: No            Mcast FEC Overload Recv: No
Serv FEC Overload Sent : No            Serv FEC Overload Recv : No
===============================================================================
```

A trap is also generated:

```
70002 2013/07/17 16:06:59.46 PST MINOR: LDP #2008 Base LDP Session State Change
"Session state is operational. Overload Notification message is sent to/from peer
  192.168.1.1:0 with overload state true for fec type prefixes"
```

Step 2

Besides interfaces and targeted peer, locally originated FECs may also be put into overload. These are the following:

- unicast fec-originate pop

- multicast local static p2mp-fec type=1 [on leaf LSR]

- multicast local Dynamic p2mp-fec type=3 [on leaf LSR]

The user can check if only remote and/or local FECs have been set in overload by the resource enhanced resource exhaustion mechanism using the following command:

- [**tools dump router ldp instance**]

The relevant part of the output is described below:

```
  Num Entities OLoad (FEC: Address Prefix  ):  Sent: 7           Rcvd: 0   <-----
// # of session in OvLd for fec-type=unicast
  Num Entities OLoad (FEC: PWE3            ):  Sent: 0           Rcvd: 0   <-----
// # of session in OvLd for fec-type=service
  Num Entities OLoad (FEC: GENPWE3         ):  Sent: 0           Rcvd: 0   <-----
// # of session in OvLd for fec-type=service
  Num Entities OLoad (FEC: P2MP            ):  Sent: 0           Rcvd: 0   <-----
// # of session in OvLd for fec-type=MulticastP2mp
  Num Entities OLoad (FEC: MP2MP UP        ):  Sent: 0           Rcvd: 0   <-----
// # of session in OvLd for fec-type=MulticastMP2mp
  Num Entities OLoad (FEC: MP2MP DOWN      ):  Sent: 0           Rcvd: 0   <-----
// # of session in OvLd for fec-type=MulticastMP2mp
  Num Active Adjacencies:    9
  Num Interfaces:            6           Num Active Interfaces: 6
  Num OLoad Interfaces:      0      <----- // link LDP interfaces in resource
 exhaustion
 should be zero when Overload Protection Capability is supported
  Num Targ Sessions:         72          Num Active Targ Sess:  67
  Num OLoad Targ Sessions:   0      <----- // T-LDP peers in resource exhaustion
 should be zero if Overload Protection Capability is supported
  Num Addr FECs Rcvd:        8667        Num Addr FECs Sent:    91
  Num Addr Fecs OLoad:       1                                    <-----
// # of local/remote unicast Fecs in Overload
  Num Svc FECs Rcvd:         3111        Num Svc FECs Sent:     0
  Num Svc FECs OLoad:        0                                    <-----
// # of local/remote service   Fecs in Overload
  Num mcast FECs Rcvd:       0           Num Mcast FECs Sent:   0
  Num mcast FECs OLoad:      0                                    <-----
// # of local/remote multicast Fecs in Overload
  Num MAC Flush Rcvd:        0           Num MAC Flush Sent:    0
```

When at least one local FEC has been set in overload the following trap will occur:

```
69999 2013/07/17 16:06:59.21 PST MINOR: LDP #2002 Base LDP Resources Exhausted
 "Instance state changed - vRtrID: 1, administrative state: inService, operational
 state: inService"
```

Step 3

After the user has detected that at least one overload status TLV has been sent or received by the LSR, he/she must protect the router by applying one or more of the following to free resources up:

- Identify the source for the [unicast/multicast/service] FEC flooding. This is most likely the LSRs which session received the overload status TLV.

- Configure the appropriate [import/export] policies and/or delete the excess [unicast/multicast/service] FECs from the FEC type in overload.

Step 4

Next, the user has to manually attempt to clear the overload state on the affected sessions and for the affected FEC types and allow the router to clear the overload status TLV to its peers.

Note: Because of the dynamic nature of FEC distribution and resolution by LSR nodes, one cannot predict exactly which sessions and which FECs will be cleared after performing the following commands if the LSR activates overload again.

One of the following commands can be used depending if the user wants to clear all sessions or at once or one session at a time:

- [**clear router ldp resource-failures**]

- Clears the overload state for the affected sessions and FEC types.
- Clear the overload state for the local FECs.

- [**clear router ldp session a.b.c.d overload fec-type {services | prefixes | multicast}**]

- Clears the overload state for the specified session and FEC type.
- Clears the overload state for the local FECs.

7.27 LDP IPv6 Control and Data Planes

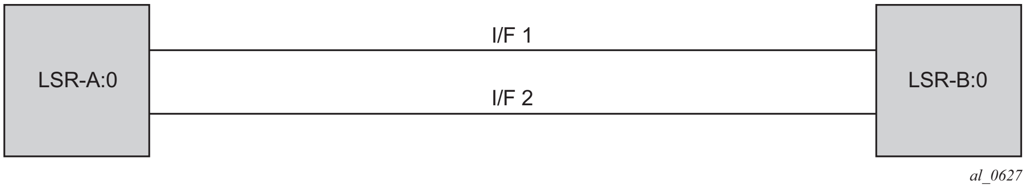
SR OS extends the LDP control plane and data plane to support LDP IPv6 adjacency and session using 128-bit LSR-ID.

The implementation allows for concurrent support of independent LDP IPv4 (32-bit LSR-ID) and IPv6 (128-bit LSR-ID) adjacencies and sessions between peer LSRs and over the same or different set of interfaces.

7.27.1 LDP Operation in an IPv6 Network

LDP IPv6 can be enabled on the SR OS interface. Figure 111 shows the LDP adjacency and session over an IPv6 interface.

Figure 111 **LDP Adjacency and Session over an IPv6 Interface**



LSR-A and LSR-B have the following IPv6 LDP identifiers respectively:

- <LSR Id=A/128> : <label space id=0>
- <LSR Id=B/128> : <label space id=0>

By default, A/128 and B/128 use the system interface IPv6 address.

Note: Although the LDP control plane can operate using only the IPv6 system address, the user must configure the IPv4-formatted router ID for OSPF, IS-IS, and BGP to operate properly.

The following sections describe the behavior when LDP IPv6 is enabled on the interface.

7.27.2 Link LDP

The SR OS LDP IPv6 implementation uses a 128-bit LSR-ID as defined in *draft-pdutta-mpls-ldp-v2-00*. See LDP Process Overview for more information about interoperability of this implementation with 32-bit LSR-ID, as defined in *RFC 7552*.

Hello adjacency will be brought up using link Hello packet with source IP address set to the interface link-local unicast address and a destination IP address set to the link-local multicast address FF02:0:0:0:0:0:0:2.

The transport address for the TCP connection, which is encoded in the Hello packet, will be set to the LSR-ID of the LSR by default. It will be set to the interface IPv6 address if the user enabled the interface option under one of the following contexts:

- **config>router>ldp>if-params>ipv6>transport-address**
- **config>router>ldp>if-params>if>ipv6>transport-address**

The interface global unicast address, meaning the primary IPv6 unicast address of the interface, is used.

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 to another interface name, loopback or not. The global unicast IPv6 address corresponding to the primary IPv6 address of the interface is used as the LSR-ID. If the user invokes an interface which does not have a global unicast IPv6 address in the configuration of the transport address or the configuration of the **local-lsr-id** option, the session will not come up and an error message will be displayed.

The LSR with the highest transport address will bootstrap the IPv6 TCP connection and IPv6 LDP session.

Source and destination addresses of LDP/TCP session packets are the IPv6 transport addresses.

7.27.3 Targeted LDP

Source and destination addresses of targeted Hello packet are the LDP IPv6 LSR-IDs of systems A and B.

The user can configure the **local-lsr-id** option on the targeted session and change the value of the LSR-ID to either the local interface or to some other interface name, loopback or not. The global unicast IPv6 address corresponding to the primary IPv6 address of the interface is used as the LSR-ID. If the user invokes an interface which does not have a global unicast IPv6 address in the configuration of the transport address or the configuration of the **local-lsr-id** option, the session will not come up and an error message will be displayed. In all cases, the transport address for the LDP session and the source IP address of targeted Hello message will be updated to the new LSR-ID value.

The LSR with the highest transport address (in this case, the LSR-ID) will bootstrap the IPv6 TCP connection and IPv6 LDP session.

Source and destination IP addresses of LDP/TCP session packets are the IPv6 transport addresses (in this case, LDP LSR-IDs of systems A and B).

7.27.4 FEC Resolution

LDP will advertise and withdraw all interface IPv6 addresses using the Address/Address-Withdraw message. Both the link-local unicast address and the configured global unicast addresses of an interface are advertised.

All LDP FEC types can be exchanged over a LDP IPv6 LDP session like in LDP IPv4 session.

The LSR does not advertise a FEC for a link-local address and, if received, the LSR will not resolve it.

A IPv4 or IPv6 prefix FEC can be resolved to an LDP IPv6 interface in the same way as it is resolved to an LDP IPv4 interface. The outgoing interface and next-hop are looked up in RTM cache. The next-hop can be the link-local unicast address of the other side of the link or a global unicast address. The FEC is resolved to the LDP IPv6 interface of the downstream LDP IPv6 LSR that advertised the IPv4 or IPv6 address of the next hop.

An mLDP P2MP FEC with an IPv4 root LSR address, and carrying one or more IPv4 or IPv6 multicast prefixes in the opaque element, can be resolved to an upstream LDP IPv6 LSR by checking if the LSR advertised the next-hop for the IPv4 root LSR address. The upstream LDP IPv6 LSR will then resolve the IPv4 P2MP FEC to one of the LDP IPV6 links to this LSR.

Note: Beginning in Release 13.0, a P2MP FEC with an IPv6 root LSR address, carrying one or more IPv4 or IPv6 multicast prefixes in the opaque element, is not supported. Manually configured mLDP P2MP LSP, NG-mVPN, and dynamic mLDP will not be able to operate in an IPv6-only network.

A PW FEC can be resolved to a targeted LDP IPv6 adjacency with an LDP IPv6 LSR if there is a context for the FEC with local spoke-SDP configuration or spoke-SDP auto-creation from a service such as BGP-AD VPLS, BGP-VPWS or dynamic MS-PW.

7.27.5 LDP Session Capabilities

LDP supports advertisement of all FEC types over an LDP IPv4 or an LDP IPv6 session. These FEC types are: IPv4 prefix FEC, IPv6 prefix FEC, IPv4 P2MP FEC, PW FEC 128, and PW FEC 129.

In addition, LDP supports signaling the enabling or disabling of the advertisement of the following subset of FEC types both during the LDP IPv4 or IPv6 session initialization phase, and subsequently when the session is already up.

- IPv4 prefix FEC—This is performed using the State Advertisement Control (SAC) capability TLV as specified in RFC 7473. The SAC capability TLV includes the IPv4 SAC element having the D-bit (Disable-bit) set or reset to disable or enable this FEC type respectively. The LSR can send this TLV in the LDP Initialization message and subsequently in a LDP Capability message.
- IPv6 prefix FEC—This is performed using the State Advertisement Control (SAC) capability TLV as specified in RFC 7473. The SAC capability TLV includes the IPv6 SAC element having the D-bit (Disable-bit) set or reset to disable or enable this FEC type respectively. The LSR can send this TLV in the LDP Initialization message and subsequently in a LDP Capability message to update the state of this FEC type.
- P2MP FEC—This is performed using the P2MP capability TLV as specified in RFC 6388. The P2MP capability TLV has the S-bit (State-bit) with a value of set or reset to enable or disable this FEC type respectively. Unlike the IPv4 SAC and IPv6 SAC capabilities, the P2MP capability does not distinguish between IPv4 and IPv6 P2MP FEC. The LSR can send this TLV in the LDP Initialization message and, subsequently, in a LDP Capability message to update the state of this FEC type.

During LDP session initialization, each LSR indicates to its peers which FEC type it supports by including the capability TLV for it in the LDP Initialization message. The SR OS implementation will enable the above FEC types by default and will thus send the corresponding capability TLVs in the LDP initialization message. If one or both peers advertise the disabling of a capability in the LDP Initialization message, no FECs of the corresponding FEC type will be exchanged between the two peers for the lifetime of the LDP session unless a Capability message is sent subsequently to explicitly enable it. The same behavior applies if no capability TLV for a FEC type is advertised in the LDP initialization message, except for the IPv4 prefix FEC which is assumed to be supported by all implementations by default.

Dynamic Capability, as defined in RFC 5561, allows all above FEC types to update the enabled or disabled state after the LDP session initialization phase. An LSR informs its peer that it supports the Dynamic Capability by including the Dynamic Capability Announcement TLV in the LDP Initialization message. If both LSRs advertise this capability, the user is allowed to enable or disable any of the above FEC types while the session is up and the change takes effect immediately. The LSR then sends a SAC Capability message with the IPv4 or IPv6 SAC element having the D-bit (Disable-bit) set or reset, or the P2MP capability TLV in a Capability message with the S-bit (State-bit) set or reset. Each LSR then takes the consequent action of withdrawing or advertising the FECs of that type to the peer LSR. If one or both LSRs did not advertise the Dynamic Capability Announcement TLV in the LDP Initialization message, any change to the enabled or disabled FEC types will only take effect at the next time the LDP session is restarted.

The user can enable or disable a specific FEC type for a given LDP session to a peer by using the following CLI commands:

- **config>router>ldp>session-params>peer>fec-type-capability p2mp**
- **config>router>ldp>session-params>peer>fec-type-capability prefix-ipv4**
- **config>router>ldp>session-params>peer>fec-type-capability prefix-ipv6**

7.27.6 LDP Adjacency Capabilities

Adjacency-level FEC-type capability advertisement is defined in *draft-pdutta-mpls-ldp-adj-capability*. By default, all FEC types supported by the LSR are advertised in the LDP IPv4 or IPv6 session initialization; see LDP Session Capabilities for more information. If a given FEC type is enabled at the session level, it can be disabled over a given LDP interface at the IPv4 or IPv6 adjacency level for all IPv4 or IPv6 peers over that interface. If a given FEC type is disabled at the session level, then FECs will not be advertised and enabling that FEC type at the adjacency level will not have any effect. The LDP adjacency capability can be configured on link Hello adjacency only and does not apply to targeted Hello adjacency.

The LDP adjacency capability TLV is advertised in the Hello message with the D-bit (Disable-bit) set or reset to disable or enable the resolution of this FEC type over the link of the Hello adjacency. It is used to restrict which FECs can be resolved over a given interface to a peer. This provides the ability to dedicate links and data path resources to specific FEC types. For IPv4 and IPv6 prefix FECs, a subset of ECMP links to a LSR peer may be each be configured to carry one of the two FEC types. An mLDP P2MP FEC can exclude specific links to a downstream LSR from being used to resolve this type of FEC.

Like the LDP session-level FEC-type capability, the adjacency FEC-type capability is negotiated for both directions of the adjacency. If one or both peers advertise the disabling of a capability in the LDP Hello message, no FECs of the corresponding FEC type will be resolved by either peer over the link of this adjacency for the lifetime of the LDP Hello adjacency, unless one or both peers sends the LDP adjacency capability TLV subsequently to explicitly enable it.

The user can enable or disable a specific FEC type for a given LDP interface to a peer by using the following CLI commands:

- **config>router>ldp>if-params>if>ipv4/ipv6>fec-type-capability p2mp-ipv4**
- **config>router>ldp>if-params>if>ipv4/ipv6>fec-type-capability p2mp-ipv6**
- **config>router>ldp>if-params>if>ipv4/ipv6>fec-type-capability prefix-ipv4**
- **config>router>ldp>if-params>if> ipv4/ipv6>fec-type-capability prefix-ipv6**

These commands, when applied for the P2MP FEC, deprecate the existing command **multicast-traffic** {**enable** | **disable**} under the interface. Unlike the session-level capability, these commands can disable multicast FEC for IPv4 and IPv6 separately.

The encoding of the adjacency capability TLV uses a PRIVATE Vendor TLV. It is used only in a hello message to negotiate a set of capabilities for a specific LDP IPv4 or IPv6 hello adjacency.

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|1|0| ADJ_CAPABILITY_TLV        |      Length                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           VENDOR_OUI                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|S|  Reserved   |                                               |
+-+-+-+-+-+-+-+-+                                               +
|                Adjacency capability elements                  |
+                                                               +
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The value of the U-bit for the TLV is set to 1 so that a receiver must silently ignore if the TLV is deemed unknown.

The value of the F-bit is 0. After being advertised, this capability cannot be withdrawn; thus, the S-bit is set to 1 in a hello message.

Adjacency capability elements are encoded as follows:

```
0 1 2 3 4 5 6 7
+-+-+-+-+-+-+-+-+
|D|  CapFlag    |
+-+-+-+-+-+-+-+-+
```

D bit: Controls the capability state.

1 : Disable capability

0 : Enable capability

CapFlag: The adjacency capability

1 : Prefix IPv4 forwarding

2 : Prefix IPv6 forwarding

3 : P2MP IPv4 forwarding

4 : P2MP IPv6 forwarding

5 : MP2MP IPv4 forwarding

6 : MP2MP IPv6 forwarding

Each CapFlag appears no more than once in the TLV. If duplicates are found, the D-bit of the first element is used. For forward compatibility, if the CapFlag is unknown, the receiver must silently discard the element and continue processing the rest of the TLV.

7.27.7 Address and FEC Distribution

After an LDP LSR initializes the LDP session to the peer LSR and the session comes up, local IPv4 and IPv6 interface addresses are exchanged using the Address and Address Withdraw messages. Similarly, FECs are exchanged using Label Mapping messages.

By default, IPv6 address distribution is determined by whether the Dual-stack capability TLV, which is defined in *RFC 7552*, is present in the Hello message from the peer. This coupling is introduced because of interoperability issues found with existing third-party LDP IPv4 implementations.

The following is the detailed behavior:

- If the peer sent the dual-stack capability TLV in the Hello message, then IPv6 local addresses will be sent to the peer. The user can configure a new address export policy to further restrict which local IPv6 interface addresses to send to the peer. If the peer explicitly stated enabling of LDP IPv6 FEC type by including the IPv6 SAC TLV with the D-bit (Disable-bit) set to 0 in the initialization message, then IPv6 FECs will be sent to the peer. FEC prefix export policies can be used to restrict which LDP IPv6 FEC can be sent to the peer.
- If the peer sent the dual-stack capability TLV in the Hello message, but explicitly stated disabling of LDP IPv6 FEC type by including the IPv6 SAC TLV with the D-bit (Disable-bit) set to 1 in the initialization message, then IPv6 FECs will not be sent but IPv6 local addresses will be sent to the peer. A CLI is provided to allow the configuration of an address export policy to further restrict which local IPv6 interface addresses to send to the peer. FEC prefix export policy has no effect because the peer explicitly requested disabling the IPv6 FEC type advertisement.
- If the peer did not send the dual-stack capability TLV in the Hello message, then no IPv6 addresses or IPv6 FECs will be sent to that peer, regardless of the presence or not of the IPv6 SAC TLV in the initialization message. This case is added to prevent interoperability issues with existing third-party LDP IPv4 implementations. The user can override this by explicitly configuring an address export policy and a FEC export policy to select which addresses and FECs to send to the peer.

The above behavior applies to LDP IPv4 and IPv6 addresses and FECs. The procedure is summarized in the flowchart diagrams in Figure 112 and Figure 113.

Figure 112 **LDP IPv6 Address and FEC Distribution Procedure**

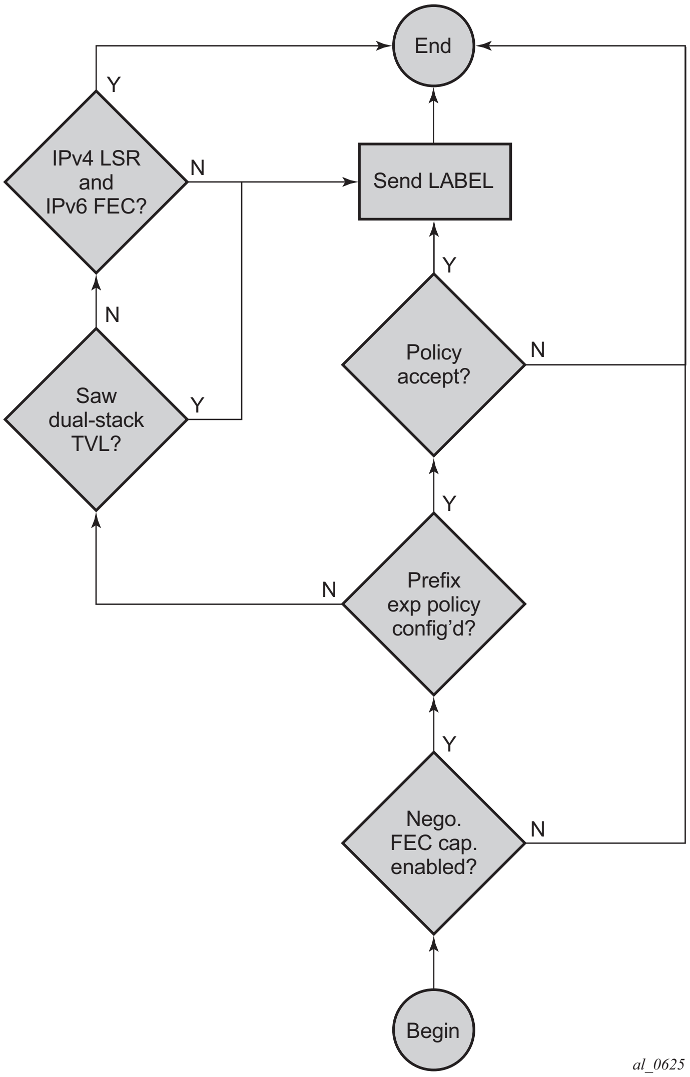
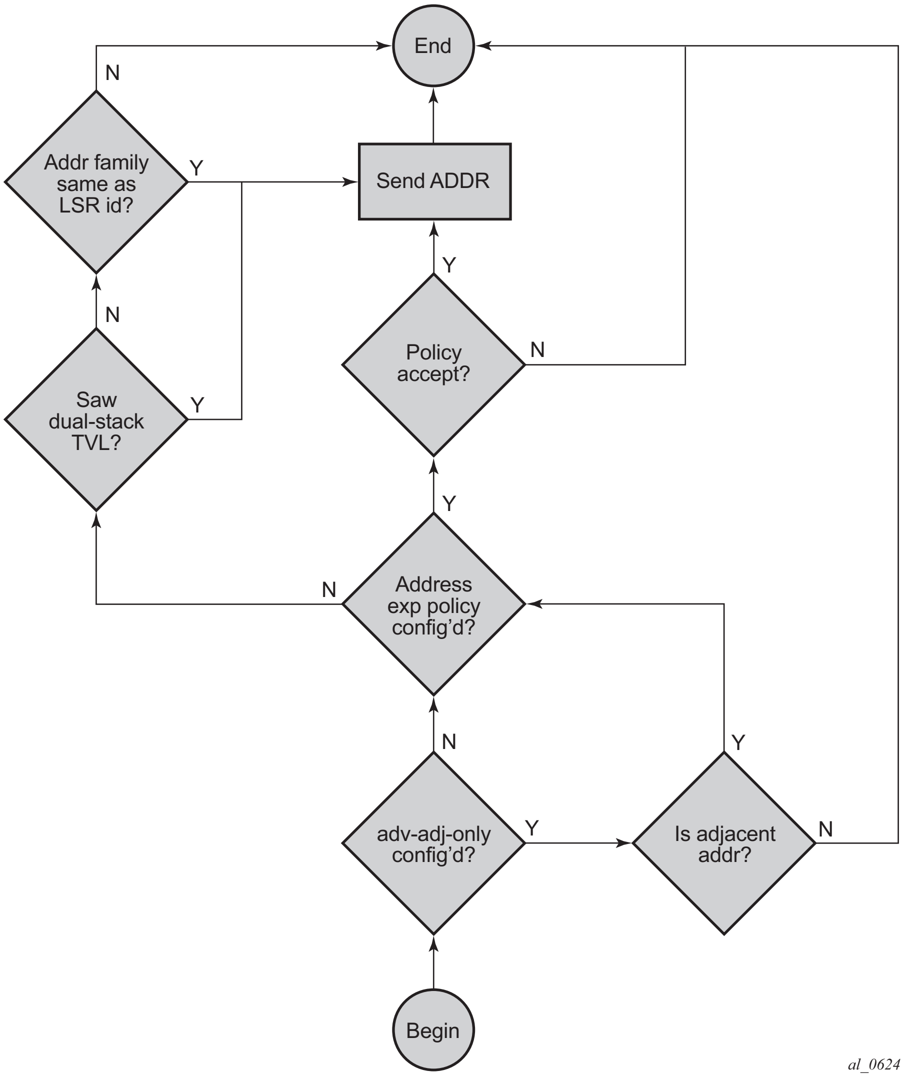


Figure 113 **LDP IPv6 Address and FEC Distribution Procedure**



7.27.8 Controlling IPv6 FEC Distribution During an Upgrade to SR OS Supporting LDP IPv6

A FEC for each of the IPv4 and IPv6 system interface addresses is advertised and resolved automatically by the LDP peers when the LDP session comes up, regardless of whether the session is IPv4 or IPv6.

To avoid the automatic advertisement and resolution of IPv6 system FEC when the LDP session is IPv4, the following procedure must be followed before and after the upgrade to the SR OS version which introduces support of LDP IPv6.

1. Before the upgrade, implement a global prefix policy which rejects prefix [::0/0 longer] to prevent IPv6 FECs from being installed after the upgrade.
2. In MISSU case:
   - If new IPv4 sessions are created on the node, the per-peer FEC-capabilities must be configured to filter out IPv6 FECs.
   - Until an existing IPv4 session is flapped, FEC-capabilities have no effect on filtering out IPv6 FECs, thus the import global policy must remain configured in place until the session flaps. Alternatively, a per-peer-import-policy [::0/0 longer] can be associated with this peer.
3. In cold upgrade case:
   - If new IPv4 sessions are created on the node, the per-peer FEC-capabilities must be configured to filter out IPv6 FECs.
   - On older, pre-existing IPv4 sessions, the per-peer FEC-capabilities must be configured to filter out IPv6 FECs.
4. When all LDP IPv4 sessions have dynamic capabilities enabled, with per-peer FEC-capabilities for IPv6 FECs disabled, then the GLOBAL IMPORT policy can be removed.

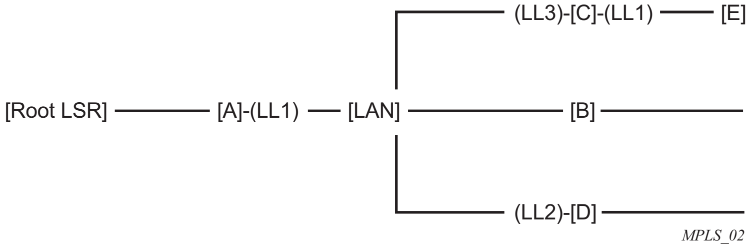
7.27.9 Handling of Duplicate Link-Local IPv6 Addresses in FEC Resolution

Link-local IPv6 addresses are scoped to a link and, as such, duplicate addresses can be used on different links to the same or different peer LSR. When the duplicate addresses exist on the same LAN, routing will detect them and block one of them. In all other cases, duplicate links are valid because they are scoped to the local link.

In this section, LLn refers to Link-Local address (n).

Figure 114 shows FEC resolution in a LAN.

Figure 114 **FEC Resolution in LAN**



LSR B resolves a mLDP FEC with the root node being Root LSR. The route lookup shows that best route to loopback of Root LSR is {interface if-B and next-hop LL1}.

However, LDP will find that both LSR A and LSR C advertised address LL1 and that there are hello adjacencies (IPv4 or IPv6) to both A and C. In this case, a change is made so that an LSR only advertises link-local IPv6 addresses to a peer for the links over which it established a Hello adjacency to that peer. In this case, LSR C will advertise LL1 to LSR E but not to LSRs A, B, and D. This behavior will apply with both P2P and broadcast interfaces.

Ambiguity also exists with prefix FEC (unicast FEC); the above solution also applies.

FEC Resolution over P2P links

```
                                   ---------(LL1)-[C]------

                                   |

[Root LSR]-------[A]-(LL1)-----[B] ------(LL4)-[D]------

                    |              |

                    |-(LL2)---------|

                    |              |

                    |-(LL3)---------|
```

LSR B resolves an mLDP FEC with root node being Root LSR. The route lookup shows that best route to loopback of Root LSR is {interface if-B and next-hop LL1}.

- Case 1—LDP is enabled on all links. This case has no ambiguity. LDP will only select LSR A because the address LL1 from LSR C is discovered over a different interface. This case also applies to prefix FEC (unicast FEC) and thus no ambiguity in the resolution.

- Case 2—LDP is disabled on link A-B with next-hop LL1; LSR B can still select one of the two other interfaces to upstream LSR A as long as LSR A advertised LL1 address in the LDP session.

7.27.10 IGP and Static Route Synchronization with LDP

The IGP-LDP synchronization and the static route to LDP synchronization features are modified to operate on a dual-stack IPv4/IPv6 LDP interface as follows:

1. If the router interface goes down or both LDP IPv4 and LDP IPv6 sessions go down, IGP sets the interface metric to maximum value and all static routes with the **ldp-sync** option enabled and resolved on this interface will be de-activated.
2. If the router interface is up and only one of the LDP IPv4 or LDP IPv6 interfaces goes down, no action is taken.
3. When the router interface comes up from a down state, and one of either the LDP IPv4 or LDP IPv6 sessions comes up, IGP starts the sync timer at the expiry of which the interface metric is restored to its configured value. All static routes with the **ldp-sync** option enabled are also activated at the expiry of the timer.

Given the above behavior, it is recommended that the user configures the sync timer to a value which allows enough time for both the LDP IPv4 and LDP IPv6 sessions to come up.

7.27.11 BFD Operation

The operation of BFD over a LDP interface tracks the next-hop of prefix IPv4 and prefix IPv6 in addition to tracking of the LDP peer address of the Hello adjacency over that link. This tracking is required as LDP can now resolve both IPv4 and IPv6 prefix FECs over a single IPv4 or IPv6 LDP session and, as such, the next-hop of a prefix will not necessarily match the LDP peer source address of the Hello adjacency. The failure of either or both of the BFD session tracking the FEC next-hop and the one tracking the Hello adjacency will cause the LFA backup NHLFE for the FEC to be activated, or the FEC to be re-resolved if there is no FRR backup.

The following CLI command allows the user to decide if they want to track only with an IPv4 BFD session, only with an IPv6 BFD session, or both:

config>router>ldp>if-params>if>bfd-enable [ipv4] [ipv6]

This command provides the flexibility required in case the user does not need to track both Hello adjacency and next-hops of FECs. For example, if the user configures **bfd-enable ipv6** only to save on the number of BFD sessions, then LDP will track the IPv6 Hello adjacency and the next-hops of IPv6 prefix FECs. LDP will not track next-hops of IPv4 prefix FECs resolved over the same LDP IPv6 adjacency. If the IPv4 data plane encounters errors and the IPv6 Hello adjacency is not affected and remains up, traffic for the IPv4 prefix FECs resolved over that IPv6 adjacency will be black-holed. If the BFD tracking the IPv6 Hello adjacency times out, then all IPv4 and IPv6 prefix FECs will be updated.

The tracking of a mLDP FEC has the following behavior:

- IPv4 and IPv6 mLDP FECs will only be tracked with the Hello adjacency because they do not have the concept of downstream next-hop.
- The upstream LSR peer for an mLDP FEC supports the multicast upstream FRR procedures, and the upstream peer will be tracked using the Hello adjacency on each link or the IPv6 transport address if there is a T-LDP session.
- The tracking of a targeted LDP peer with BFD does not change with the support of IPv6 peers. BFD tracks the transport address conveyed by the Hello adjacency which bootstrapped the LDP IPv6 session.

7.27.12 Services Using SDP with an LDP IPv6 FEC

The SDP of type **LDP** with **far-end** and **tunnel-farend** options using IPv6 addresses is supported. The addresses need not be of the same family (IPv6 or IPv4) for the SDP configuration to be allowed. The user can have an SDP with an IPv4 (or IPv6) control plane for the T-LDP session and an IPv6 (or IPv4) LDP FEC as the tunnel.

Because IPv6 LSP is only supported with LDP, the use of a **far-end** IPv6 address will not be allowed with a BGP or RSVP/MPLS LSP. In addition, the CLI will not allow an SDP with a combination of an IPv6 LDP LSP and an IPv4 LSP of a different control plane. As a result, the following commands are blocked within the SDP configuration context when the far-end is an IPv6 address:

- **bgp-tunnel**
- **lsp**
- **mixed-lsp-mode**

SDP admin groups are not supported with an SDP using an LDP IPv6 FEC, and the attempt to assign them is blocked in CLI.

Services which use LDP control plane (such as T-LDP VPLS and R-VPLS, VLL, and IES/VPRN spoke interface) will have the spoke-SDP (PW) signaled with an IPv6 T-LDP session when the **far-end** option is configured to an IPv6 address. The spoke-SDP for these services binds by default to an SDP that uses a LDP IPv6 FEC, which prefix matches the far end address. The spoke-SDP can use a different LDP IPv6 FEC or a LDP IPv4 FEC as the tunnel by configuring the **tunnel-far-end** option. In addition, the IPv6 PW control word is supported with both data plane packets and VCCV OAM packets. Hash label is also supported with the above services, including the signaling and negotiation of hash label support using T-LDP (Flow sub-TLV) with the LDP IPv6 control plane. Finally, network domains are supported in VPLS.

7.27.13 Mirror Services and Lawful Intercept

The user can configure a spoke-SDP bound to an LDP IPv6 LSP to forward mirrored packets from a mirror source to a remote mirror destination. In the configuration of the mirror destination service at the destination node, the remote-source command must use a spoke-SDP with a VC-ID that matches the one that is configured in the mirror destination service at the mirror source node. The **far-end** option will not be supported with an IPv6 address.

This also applies to the configuration of the mirror destination for a LI source.

7.27.13.1 Configuration at mirror source node

Use the following rules and syntax to configure at the mirror source node.

- The *sdp-id* must match an SDP which uses LDP IPv6 FEC
- Configuring *egress-vc-label* is optional.
  config mirror mirror-dest 10

CLI Syntax:

```
no spoke-sdp sdp-id:vc-id
spoke-sdp sdp-id:vc-id [create]
   egress
      vc-label egress-vc-label
```

7.27.13.2 Configuration at mirror destination node

Use the following rules and syntax to configure at the mirror destination node.

- The **far-end** *ip-address* command is not supported with LDP IPv6 transport tunnel. The user must reference a spoke-SDP using a LDP IPv6 SDP coming from mirror source node.
- In the **spoke-sdp** *sdp-id:vc-id* command, *vc-id* should match that of the **spoke-sdp** configured in the mirror-destination context at mirror source node.
- Configuring *ingress-vc-label* is optional; both static and t-ldp are supported.

  configure mirror mirror-dest 10 remote-source

CLI Syntax:

```
far-end ip-address [vc-id vc-id] [ing-svc-label ingress-
  vc-label | tldp] [icb]
no far-end ip-address
   spoke-sdp sdp-id:vc-id [create]
      ingress-vc-label ingress-vc-label
      exit
   no shutdown
   exit
exit
```

Mirroring and LI will also be supported with PW redundancy feature when the endpoint spoke-SDP, including the ICB, is using a LDP IPv6 tunnel.

7.27.14 Static Route Resolution to a LDP IPv6 FEC

An LDP IPv6 FEC can be used to resolve a static IPv6 route with an indirect next-hop matching the FEC prefix. The user configures a resolution filter to specify the LDP tunnel type to be selected from TTM:

config>router>static-route-entry *ip-prefix/prefix-length* **[mcast]**

```
indirect ip-address
   tunnel-next-hop
      [no] disallow-igp
      resolution {any | disabled | filter}
      resolution-filter
      [no] ldp
```

A static route of an IPv6 prefix cannot be resolved to an indirect next-hop using a LDP IPv4 FEC. An IPv6 prefix can only be resolved to an IPv4 next-hop using the 6-over-4 encapsulation by which the outer IPv4 header uses system IPv4 address as source and the next-hop as a destination. So the following example will return an error:

```
A:SRU4>config>router# static-route-entry 3ffe::30/128 indirect 192.168.1.1 tunnel-
next-hop
 resolution-filter ldp

MINOR: CLI LDP not allowed for 6over4.
```

7.27.15 IGP Route Resolution to a LDP IPv6 FEC

LDP IPv6 shortcut for IGP IPv6 prefix is supported. The following commands allow a user to select if shortcuts must be enabled for IPv4 prefixes only, for IPv6 prefixes only, or for both.

config>router>ldp-shortcut [ipv4][ipv6]

```
ldp-shortcut [ipv4][ipv6]
no ldp-shortcut
```

This CLI command has the following behaviors:

- When executing a pre-Release 13.0 config file, the existing command is converted as follows:
  config>**router**>**ldp-shortcut** changed to **config**>**router**>**ldp-shortcut ipv4**
- If the user enters the command without the optional arguments in the CLI, it defaults to enabling shortcuts for IPv4 IGP prefixes:
  config>**router**>**ldp-shortcut** changed to **config**>**router**>**ldp-shortcut ipv4**
- When the user enters both IPv4 and IPv6 arguments in the CLI, shortcuts for both IPv4 and IPv6 prefixes are enabled:
  config>**router**>**ldp-shortcut ipv4 ipv6**

7.27.16 OAM Support with LDP IPv6

MPLS OAM tools **lsp-ping** and **lsp-trace** are updated to operate with LDP IPv6 and support the following:

- use of IPv6 addresses in the echo request and echo reply messages, including in DSMAP TLV, as per RFC 8029
- use of LDP IPv6 prefix target FEC stack TLV as per RFC 8029
- use of IPv6 addresses in the DDMAP TLV and FEC stack change sub-TLV, as per RFC 6424
- use of 127/8 IPv4 mapped IPv6 address; that is, in the range ::ffff:127/104, as the destination address of the echo request message, as per RFC 8029.
- use of 127/8 IPv4 mapped IPv6 address; that is, in the range ::ffff:127/104, as the **path-destination** address when the user wants to exercise a specific LDP ECMP path.

The behavior at the sender and receiver nodes is updated to support both LDP IPv4 and IPv6 target FEC stack TLVs. Specifically:

1. The IP family (IPv4/IPv6) of the UDP/IP echo request message will always match the family of the LDP target FEC stack TLV as entered by the user in the **prefix** option.
2. The **src-ip-address** option is extended to accept IPv6 address of the sender node. If the user did not enter a source IP address, the system IPv6 address will be used. If the user entered a source IP address of a different family than the LDP target FEC stack TLV, an error is returned and the test command is aborted.
3. The IP family of the UDP/IP echo reply message must match that of the received echo request message.
4. For **lsp-trace**, the downstream information in DSMAP/DDMAP will be encoded as the same family as the LDP control plane of the link LDP or targeted LDP session to the downstream peer.
5. The sender node inserts the experimental value of 65503 in the Router Alert Option in the echo request packet's IPv6 header as per RFC 5350. Once a value is allocated by IANA for MPLS OAM as part of *draft-ietf-mpls-oam-ipv6-rao*, it will be updated.

Finally, **vccv-ping** and **vccv-trace** for a single-hop PW are updated to support IPv6 PW FEC 128 and FEC 129 as per RFC 6829. In addition, the PW OAM control word is supported with VCCV packets when the **control-word** option is enabled on the spoke-SDP configuration. The value of the Channel Type field is set to 0x57, which indicates that the Associated Channel carries an IPv6 packet, as per RFC 4385.

7.27.17 LDP IPv6 Interoperability Considerations

7.27.17.1 Interoperability with Implementations Compliant with RFC 7552

The SR OS implementation uses a 128-bit LSR-ID, as defined in RFC 7552, to establish an LDP IPv6 Hello adjacency and session with a peer LSR. This allows a routable system IPv6 address to be used by default to bring up the LDP task on the router and establish link LDP and T-LDP sessions to other LSRs, as is the common practice with LDP IPv4 in existing customer deployments. More importantly, this allows for the establishment of control plane independent LDP IPv4 and LDP IPv6 sessions between two LSRs over the same interface or set of interfaces. The SR OS implementation allows for multiple separate LDP IPv4 and LDP IPv6 sessions between two routers over the same interface or a set of interfaces, as long as each session uses a unique LSR-ID (32-bit for IPv4 and 128-bit for IPv6).

The SR OS LDP IPv6 implementation complies with the control plane procedures defined in RFC 7552 for establishing an LDP IPv6 Hello adjacency and LDP session. However, the implementation does not interoperate, by default, with third-party implementations of this standard since the latter encode a 32-bit LSR-ID in the IPv6 Hello message while SR OS encodes a 128-bit LSR-ID.

To assure interoperability in deployments strictly adhering to RFC 7552, SR OS provides the option for configuring and encoding a 32-bit LSR-ID in the LDP IPv6 Hello message. When this option is enabled, an SR OS LSR establishes an LDP IPv6 Hello adjacency and an LDP IPv6 session with an RFC 7552 compliant peer or targeted peer LSR, using a 32-bit LSR-ID and a 128-bit transport address. See LDP IPv6 32-bit LSR-ID for more information.

In a dual-stack IPv4/IPV6 interface environment, the SR OS based LSR will not originate both IPv6 and IPv4 Hello messages with the configured 32-bit LSR-ID value when both IPv4 and IPv6 contexts are enabled on the same LDP interface. This behavior is allowed in RFC 7552 for migration purposes. However, the SR OS implements separate IPv4 and IPv6 Hello adjacencies and LDP sessions with different LSR-ID values for the LDP IPv4 (32-bit value) and LDP IPv6 (32-bit or 128-bit value) Hello adjacencies. Therefore, the LDP IPv4 and LDP IPv6 sessions are independent in the control plane.

However, if the peer LSR sends both IPv4 and IPv6 Hello messages using the same 32-bit LSR-ID value, as allowed in RFC 7552, only a single LDP session with the local 32-bit LSR-ID will come up toward that peer LSR-ID, depending on which of the IPv4 or IPv6 adjacencies came up first.

The dual-stack capability TLV, in the Hello message, is used by an LSR to inform its peer that it is capable of establishing either an LDP IPv4 or LDP IPv6 session, and the IP family preference for the LDP Hello adjacency for the resulting LDP session.

Finally, the SR OS LDP implementation inter-operates with an implementation using a 32-bit LSR-ID, as defined in RFC 7552, to establish an IPv4 LDP session and to resolve both IPv4 and IPv6 prefix FECs. In this case, the dual-stack capability TLV indicates implicitly the LSR support for resolving IPv6 FECs over an IPv4 LDP session.

7.27.17.2 LDP IPv6 32-bit LSR-ID

The SR OS implementation provides the option for configuring and encoding a 32-bit LSR-ID in the LDP IPv6 Hello message to achieve interoperability in deployments strictly adhering to RFC 7552.

The LSR-ID of an LDP Label Switched Router (LSR) is a 32-bit integer used to uniquely identify it in a network. SR OS also supports LDP IPv6 in both the control plane and data plane. However, the implementation uses a 128-bit LSR-ID, as defined in *draft-pdutta-mpls-ldp-v2* to establish an LDP IPv6 Hello adjacency and session with a peer LSR.

The SR OS LDP IPv6 implementation complies with the control plane procedures defined in RFC 7552 for establishing an LDP IPv6 Hello adjacency and LDP session. However, the SR OS LDP IPv6 implementation does not interoperate with third-party implementations of this standard, since the latter encode a 32-bit LSR-ID in the IPv6 Hello message while SR OS encodes a 128-bit LSR-ID.

When this feature is enabled, an SR OS LSR will be able to establish an LDP IPv6 Hello adjacency and an LDP IPv6 session with an RFC 7552 compliant peer or targeted peer LSR, using a 32-bit LSR-ID and a 128-bit transport address.

7.27.17.2.1 Feature Configuration

This user configures the 32-bit LSR-ID on a LDP peer or targeted peer using the following CLI:

config>**router**>**ldp**>**interface-parameters**>**interface**>**ipv6**>**local-lsr-id interface** [**32bit-format**]

config>**router**>**ldp**>**interface-parameters**>**interface**>**ipv6**>**local-lsr-id interface-name** *interface-name* [**32bit-format**]

config>**router**>**ldp**>**targeted-session**>**peer**>**local-lsr-id** *interface-name* [**32bit-format**]

When the **local-lsr-id** command is enabled with the **32bit-format** option, an SR OS LSR will be able to establish a LDP IPv6 Hello adjacency and a LDP IPv6 session with a RFC 7552 compliant peer or targeted peer LSR using a 32-bit LSR-ID set to the value of the IPv4 address of the specified local LSR-ID interface and a 128-bit transport address set to the value of the IPv6 address of the specified local LSR-ID interface.

➡ **Note:** The system interface cannot be used as a local LSR-ID with the **32bit-format** option enabled as it is the default LSR-ID and transport address for all LDP sessions to peers and targeted peers on this LSR. This configuration is blocked in CLI.

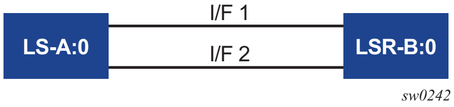
If the user enables the **32bit-format** option in the IPv6 context of a running LDP interface or in the targeted session peer context of a running IPv6 peer, the already established LDP IPv6 Hello adjacency and LDP IPv6 session will be brought down and re-established with the new 32-bit LSR-ID value.

The detailed control plane procedures are provided in LDP LSR IPv6 Operation with 32-bit LSR-ID.

7.27.17.2.2 LDP LSR IPv6 Operation with 32-bit LSR-ID

Consider the setup shown in Figure 115.

Figure 115 **LDP Adjacency and Session over IPv6 Interface**



LSR A and LSR B have the following LDP parameters.

LSR A:

- Interface I/F1 : link local address = fe80::a1
- Interface I/F2 : link local address = fe80::a2
- Interface LoA1: IPv4 address = <A1/32>; primary IPv6 unicast address = <A2/128>
- Interface LoA2: IPv4 address = <A3/32>; primary IPv6 unicast address = <A4/128>
- local-lsr-id (**config**>**router**>**ldp**>**interface-parameters**>**interface**>**ipv6**) = interface LoA1; option **32bit-format** enabled
  - LDP identifier = {<LSR Id=A1/32> : <label space id=0>}; transport address = <A2/128>
- local-lsr-id (**config**>>**router**>**ldp**>**targeted-session**>**peer**) = interface LoA2; option **32bit-format** enabled
  - LDP identifier = {<LSR Id=A3/32> : <label space id=0>}; transport address = <A4/128>

LSR B:

- Interface I/F1 : link local address = fe80::b1
- Interface I/F2 : link local address = fe80::b2

- Interface LoB1: IPv4 address = <B1/32>; primary IPv6 unicast address = <B2/128>
- Interface LoB2: IPv4 address = <B3/32>; primary IPv6 unicast address = <B4/128>
- local-lsr-id (**config**>**router**>**ldp**>**interface-parameters**>**interface**>**ipv6**) = interface LoB1; option **32bit-format** enabled
  - LDP identifier = {<LSR Id=B1/32> : <label space id=0>}; transport address = <B2/128>
- local-lsr-id (**config**>**router**>**ldp**>**targeted-session**>**peer**) = interface LoB2; option **32bit-format** enabled
  - LDP identifier = {<LSR Id=B3/32> : <label space id=0>}; transport address = <B4/128>

Link LDP

When the IPv6 context of interfaces I/F1 and I/F2 are brought up, the following procedures are performed.

- LSR A (LSR B) sends a IPv6 Hello message with source IP address set to the link-local unicast address of the specified local LSR ID interface, for example, fe80::a1 (fe80::a2), and a destination IP address set to the link-local multicast address ff02:0:0:0:0:0:0:2.
- LSR A (LSR B) sets the LSR-ID in LDP identifier field of the common LDP PDU header to the 32-bit IPv4 address of the specified local LSR-ID interface LoA1 (LoB1), for example, A1/32 (B1/32).

  If the specified local LSR-ID interface is unnumbered or does not have an IPv4 address configured, the adjacency will not come up and an error will be returned (lsrInterfaceNoValidIp (17) in output of '**show router ldp interface detail**').
- LSR A (LSR B) sets the transport address TLV in the Hello message to the IPv6 address of the specified local LSR-ID interface LoA1 (LoB1), for example, A2/128 (B2/128).

  If the specified local LSR-ID interface is unnumbered or does not have an IPv6 address configured, the adjacency will not come up and an error will be returned (interfaceNoValidIp (16) in output of '**show router ldp interface detail**'.
- LSR A (LSR B) includes in each IPv6 Hello message the dual-stack TLV with the transport connection preference set to IPv6 family.
  - If the peer is a third-party LDP IPv6 implementation and does not include the dual-stack TLV, then LSR A (LSR B) resolves IPv6 FECs only because IPv6 addresses will not be advertised in Address messages as per RFC 7552 [ldp-ipv6-rfc].

  - If the peer is a third-party LDP IPv6 implementation and includes the dual-stack TLV with transport connection preference set to IPv4, LSR A (LSR B) will not bring up the Hello adjacency and discard the Hello message. If the LDP session was already established, then LSRA(B) will send a fatal Notification message with status code of 'Transport Connection Mismatch' (0x00000032)' and restart the LDP session [ldp-ipv6-rfc]. In both cases, a new counter for the transport connection mismatches will be incremented in the output of '**show router ldp statistics**'.
- The LSR with highest transport address takes on the active role and initiates the TCP connection for the LDP IPv6 session using the corresponding source and destination IPv6 transport addresses.

Targeted LDP

Similarly, when the new option is invoked on a targeted IPv6 peer, the router sends a IPv6 targeted Hello message with source IP address set to the global unicast IPv6 address corresponding to the primary IPv6 address of the specified interface and a destination IP address set to configured IPv6 address of the peer. The LSR-ID field in the LDP identifier in the common LDP PDU header is set the 32-bit address of the specified interface. If the specified interface does not have an IPv4 address configured the adjacency will not come up. Any subsequent adjacency or session level messages will be sent with the common LDP PDU header set as above.

When the targeted IPv6 peer contexts are brought up, the following procedures are performed.

- LSR A (LSR B) sends a IPv6 Hello message with source IP address set to the primary IPv6 unicast address of the specified local LSR ID interface LoA2(LoB2), for example, A4/128 (B4/128), and a destination IP address set to the peer address B4/128(A4/128).
- LSR A (LSR B) sets the LSR-ID in LDP identifier field of the common LDP PDU header to the 32-bit IPv4 address of the specified local LSR-ID interface LoA2(LoB2), for example, A3/32 (B3/32).

  If the specified local LSR-ID interface is unnumbered or does not have an IPv4 address configured, the adjacency will not come up and an error will be returned.
- LSR A (LSR B) sets the transport address TLV in the Hello message to the IPv6 address of the specified local LSR-ID interface LoA2 (LoB2), for example, A4/128 (B4/128).

  If the specified local LSR-ID interface is unnumbered or does not have an IPv6 address configured, the adjacency will not come up and an error will be returned.
- LSR A (LSR B) includes in each IPv6 Hello message the dual-stack TLV with the preference set to IPv6 family.

  - If the peer is a third-party LDP IPv6 implementation and does not include the dual-stack TLV, then LSR A (LSR B) resolves IPv6 FECs only since IPv6 addresses will not be advertised in Address messages as per RFC 7552 [ldp-ipv6-rfc].
  - If the peer is a third-party LDP IPv6 implementation and includes the dual-stack TLV with transport connection preference set to IPv4, LSR A (LSR B) will not bring up the Hello adjacency and discard the Hello message. If the LDP session was already established, then LSRA(B) will send a fatal Notification message with status code of 'Transport Connection Mismatch' (0x00000032)' and restart the LDP session [ldp-ipv6-rfc]. In both cases, a new counter for the transport connection mismatches will be incremented in the output of '**show router ldp statistics**'.
- The LSR with highest transport address takes on the active role and initiates the TCP connection for the LDP IPv6 session using the corresponding source and destination IPv6 transport addresses.

Link and Targeted LDP Feature Interaction

The following describes feature interactions.

- LSR A (LSR B) will not originate both IPv6 and IPv4 Hello messages with the configured 32-bit LSR-ID value when both IPv4 and IPv6 contexts are enabled on the same LDP interface (dual-stack LDP IPv4/IPv6). This behavior is allowed in RFC 7552 for migration purposes but SR OS implements separate IPv4 and IPv6 Hello adjacencies and LDP sessions with different LSR-ID values. Therefore, an IPv6 context which uses a 32-bit LSR-ID address matching that of the IPv4 context on the same interface will not be allowed to be brought up (**no shutdown** will fail) and vice-versa.

  Furthermore, an IPv6 context of any interface or targeted peer which uses a 32-bit LSR-ID address matching that of an IPv4 context of any other interface, an IPv6 context of any other interface using 32-bit LSR-ID, a targeted IPv4 peer, a targeted IPv6 peer using 32-bit LSR-ID, or an auto T-LDP IPv4 template on the same router will not be allowed to be brought up (**no shutdown** will fail) and vice-versa.
- With the introduction of a 32-bit LSR-ID for a IPv6 LDP interface or peer, it is possible to configure the same IPv6 transport address for an IPv4 LSR-ID and an IPv6 LSR-ID on the same node. For instance, assume the following configuration:
  - Interface I/F1:
    - local-lsr-id (**config**>**router**>**ldp**>**interface-parameters**>**interface**>**ipv6**) = interface LoA1; option **32bit-format** enabled.

    - LDP identifier = {<LSR Id=A1/32> : <label space id=0>}; transport address = <A2/128>
- Interface I/F2:
    - local-lsr-id (**config**>**router**>**ldp**>**interface-parameters**>**interface**>**ipv6**) = interface LoA1;
    - LDP identifier = {<LSR Id=A2/128> : <label space id=0>}; transport address = <A2/128>
- Targeted Session:
    - local-lsr-id (**config**>**router**>**ldp**> **targeted-session**>**peer**) = interface LoA1;
    - LDP identifier = {<LSR Id=A2/128> : <label space id=0>}; transport address = <A2/128>

The above configuration will result in two interfaces and a targeted session with the same local end transport IPv6 address but the local LSR-ID for interface I/F1 is different.

If an IPv6 Hello adjacency over interface I/F1 towards a given peer comes up first and initiates an IPv6 LDP session, then the other two Hello adjacencies to the same peer will not come up.

If one of the IPv6 Hello adjacencies of interface I/F2 or Targeted Session 1 comes up first to a peer, it will trigger an IPv6 LDP session shared by both these adjacencies and the Hello adjacency over interface I/F1 to the same peer will not come up.

7.27.17.2.3 Migration Considerations

Migrating Services from LDP IPv4 Session to 32-bit LSR-ID LDP IPv6 Session

Assume the user deploys on a SR OS based LSR a service bound to a SDP which auto-creates the IPv4 targeted LDP session to a peer LSR running a third party LDP implementation. In this case, the auto-created T-LDP session uses the system interface IPv4 address as the local LSR-ID and as the local transport address because there is no targeted session configured in LDP to set these parameters away from default values.

When both LSR nodes are being migrated to using LDP IPv6 with a 32-bit LSR-ID, the user must configure the IPv6 context of the local LDP interfaces to use a local LSR-ID interface different than the system interface and with the **32bit-format** option enabled. Similarly, the user must configure a new Targeted session in LDP with that same local LSR-ID interface and with the **32bit-format** option enabled. This will result in a LDP IPv6 session triggered by the link LDP IPv6 Hello adjacency or the targeted IPv6 Hello adjacency which came up first. This LDP IPv6 session uses the IPv4 address and the IPv6 address of the configured local LSR-ID interface as the LSR-ID and transport address respectively.

The user must then modify the service configuration on both ends to use a far-end address matching the far-end IPv6 transport address of the LDP IPv6 session. On the SR OS based LSR, this can be done by creating a new IPv6 SDP of type LDP with the far-end address matching the far-end IPv6 transport address.

If the service enabled PW redundancy, the migration may be eased by creating a standby backup PW bound to the IPv6 SDP and adding it to the same VLL or VPLS endpoint the spoke-sdp bound to the IPv4 SDP belongs to. Then, activate the backup PW using the command '**tools**>**perform**>**service**>**id**>**endpoint**>**force-switchover sdp-id**:**vc-id**'. This make the spoke-sdp bound to the IPv6 SDP the primary PW. Finally, the spoke-sdp bound to the IPv4 SDP can be deleted.

7.27.17.3 Interoperability with Implementations Compliant with RFC 5036 for IPv4 LDP Control Plane Only

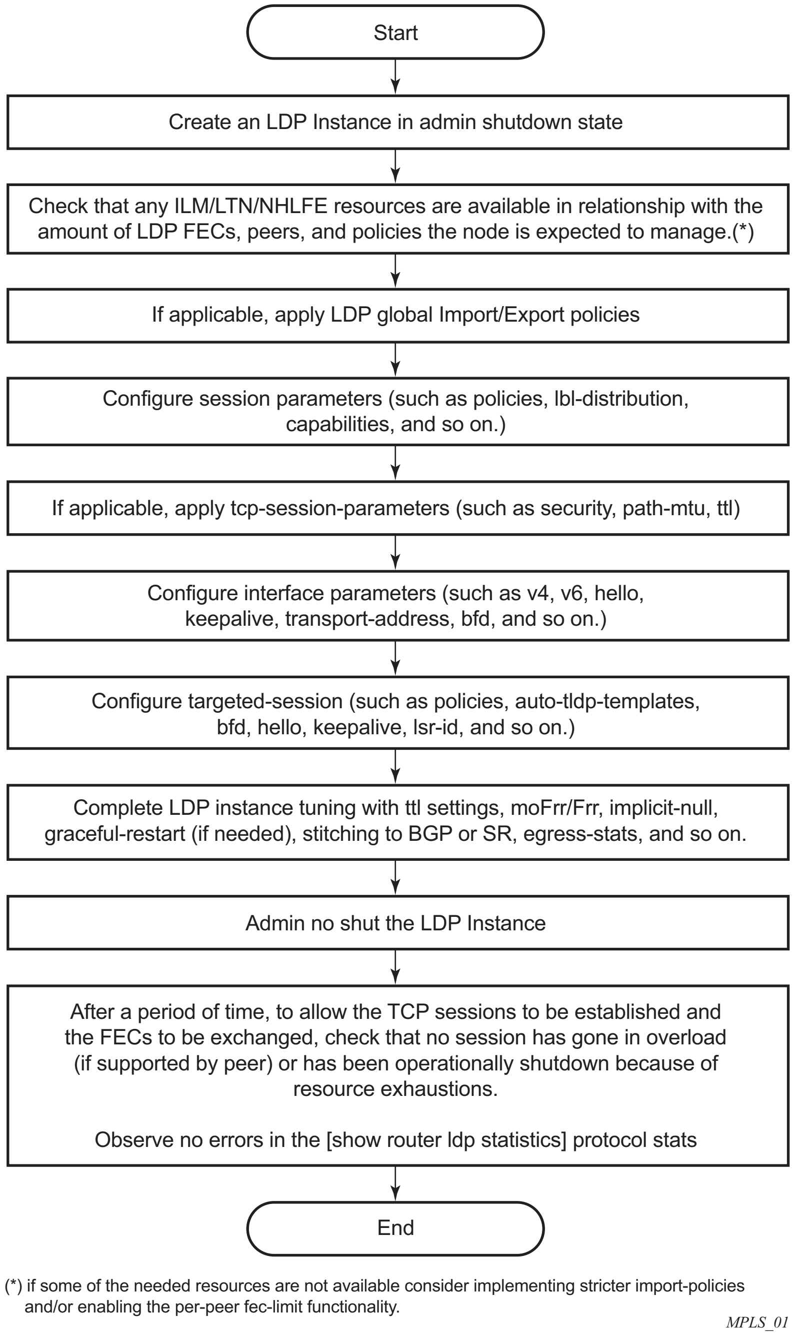
The SR OS implementation supports advertising and resolving IPv6 prefix FECs over an LDP IPv4 session using a 32-bit LSR-ID, in compliance with RFC 7552. When introducing an LSR based on the SR OS in a LAN with a broadcast interface, it can peer with third-party LSR implementations that support RFC 7552 and LSRs that do not. When it peers, using an IPv4 LDP control plane, with a third-party LSR implementation that does not support it, the advertisement of IPv6 addresses or IPv6 FECs to that peer may cause it to bring down the IPv4 LDP session.

That is, there are deployed third-party LDP implementations that are compliant with RFC 5036 for LDP IPv4, but that are not compliant with RFC 5036 for handling IPv6 address or IPv6 FECs over an LDP IPv4 session. To resolve this issue, RFC 7552 modifies RFC 5036 by requiring implementations complying with RFC 7552 to check for the dual-stack capability TLV in the IPv4 Hello message from the peer. Without the peer advertising this TLV, an LSR must not send IPv6 addresses and FECs to that peer. The SR OS implementation supports this requirement.

7.28 LDP Process Overview

Figure 116 displays the process to provision basic LDP parameters.

Figure 116 **LDP Configuration and Implementation**


Start
Create an LDP Instance in admin shutdown state
Check that any ILM/LTN/NHLFE resources are available in relationship with the amount of LDP FECs, peers, and policies the node is expected to manage.(*)
If applicable, apply LDP global Import/Export policies
Configure session parameters (such as policies, lbl-distribution, capabilities, and so on.)
If applicable, apply tcp-session-parameters (such as security, path-mtu, ttl)
Configure interface parameters (such as v4, v6, hello, keepalive, transport-address, bfd, and so on.)
Configure targeted-session (such as policies, auto-tldp-templates, bfd, hello, keepalive, lsr-id, and so on.)
Complete LDP instance tuning with ttl settings, moFrr/Frr, implicit-null, graceful-restart (if needed), stitching to BGP or SR, egress-stats, and so on.
Admin no shut the LDP Instance
After a period of time, to allow the TCP sessions to be established and the FECs to be exchanged, check that no session has gone in overload (if supported by peer) or has been operationally shutdown because of resource exhaustions.
Observe no errors in the [show router ldp statistics] protocol stats
End
(*) if some of the needed resources are not available consider implementing stricter import-policies and/or enabling the per-peer fec-limit functionality.
MPLS_01

7.29 Configuring LDP with CLI

This section provides information to configure LDP using the command line interface.

7.29.1 LDP Configuration Overview

When the implementation of LDP is instantiated, the protocol is in the no shutdown state. In addition, targeted sessions are then enabled. The default parameters for LDP are set to the documented values for targeted sessions in *draft-ietf-mpls-ldp-mib-09.txt*.

LDP must be enabled in order for signaling to be used to obtain the ingress and egress labels in frames transmitted and received on the service distribution path (SDP). When signaling is *off*, labels must be manually configured when the SDP is bound to a service.

7.29.2 Basic LDP Configuration

This section provides information to configure LDP and remove configuration examples of common configuration tasks.

The LDP protocol instance is created in the no shutdown (enabled) state.

The following displays the default LDP configuration.

```
A:ALA-1>config>router>ldp# info
----------------------------------------------
            session-parameters
            exit
            interface-parameters
            exit
            targeted-session
            exit
            no shutdown
----------------------------------------------
A:ALA-1>config>router>ldp#
```

7.29.3 Common Configuration Tasks

This section provides an overview of the tasks to configure LDP and provides the CLI commands.

7.29.3.1 Enabling LDP

LDP must be enabled in order for the protocol to be active. MPLS does not need to be enabled on the router except if the network interface uses the Packet over Sonet (POS) encapsulation (Sonet path encapsulation type set to ppp-auto). In this case, MPLS must be enabled and the interface name added into MPLS to allow for the MPLSCP to come up on the PPP link between the two peers and for MPLS to be used on the interface. MPLS is enabled in the config>router>mpls context.

Use the following syntax to enable LDP on a router:

CLI Syntax: `ldp`

Example: `config>router#` **`ldp`**

The following displays the enabled LDP configuration.

```
A:ALA-1>config>router# info
----------------------------------------------
...
#------------------------------------------
echo "LDP Configuration"
#------------------------------------------
        ldp
            session-parameters
            exit
            interface-parameters
            exit
            targeted-session
            exit
        exit
----------------------------------------------
...
A:ALA-1>config>router#
```

7.29.3.2 Configuring FEC Originate Parameters

A FEC can be added to the LDP IP prefix database with a specific label operation on the node. Permitted operations are pop or swap. For a swap operation, an incoming label can be swapped with a label in the range of 16 to 1048575. If a swap- label is not configured then the default value is 3.

A route table entry is required for a FEC with a pop operation to be advertised. For a FEC with a swap operation, a route-table entry must exist and user configured next-hop for swap operation must match one of the next-hops in route-table entry.

Use the following syntax to configure FEC originate parameters:

CLI Syntax:

```
config>router>ldp
fec-originate ip-prefix/mask [advertised-label in-label]
 next-hop ip-address [swap-label out-label]
fec-originate ip-prefix/mask [advertised-label in-label]
 pop
```

The following displays a FEC originate configuration example.

```
A:ALA-5>config>router# info
----------------------------------------------
            fec-originate 10.1.1.1/32 pop
            fec-originate 10.1.2.1/32 advertised-label 1000 next-hop 10.10.1.2
            fec-originate 10.1.3.1/32 advertised-label 1001 next-hop 10.10.2.3
swap-label 131071
            session-parameters
            exit
            interface-parameters
            exit
            targeted-session
            exit
        exit
----------------------------------------------
A:ALA-5>config>router>ldp#
```

7.29.3.3 Configuring Graceful-Restart Helper Parameters

Graceful-restart helper advertises to its LDP neighbors by carrying the fault tolerant (FT) session TLV in the LDP initialization message, assisting the LDP in preserving its IP forwarding state across the restart. Nokia's recovery is self-contained and relies on information stored internally to self-heal. This feature is only used to help third-party routers without a self-healing capability to recover.

Maximum recovery time is the time (in seconds) the sender of the TLV would like the receiver to wait, after detecting the failure of LDP communication with the sender.

Neighbor liveness time is the time (in seconds) the LSR is willing to retain its MPLS forwarding state. The time should be long enough to allow the neighboring LSRs to re-sync all the LSPs in a graceful manner, without creating congestion in the LDP control plane.

Use the following syntax to configure graceful-restart parameters:

CLI Syntax:

```
config>router>ldp
   [no] graceful-restart
```

7.29.3.4 Applying Export and Import Policies

Both inbound and outbound label binding filtering are supported. Inbound filtering allows a route policy to control the label bindings an LSR accepts from its peers. An import policy can accept or reject label bindings received from LDP peers.

Label bindings can be filtered based on:

- Neighbor — Match on bindings received from the specified peer.
- Prefix-list — Match on bindings with the specified prefix/prefixes.

Outbound filtering allows a route policy to control the set of LDP label bindings advertised by the LSR. An export policy can control the set of LDP label bindings advertised by the router. By default, label bindings for only the system address are advertised and propagate all FECs that are received. All other local interface FECs can be advertised using policies.

Note: Static FECs cannot be blocked using an export policy.

Matches can be based on:

- All — all local subnets.
- Match — match on bindings with the specified prefix/prefixes.

Use the following syntax to apply import and export policies:

CLI Syntax:

```
config>router>ldp
export policy-name [policy-name...(upto 32 max)]
import policy-name [policy-name...(upto 32 max)]
```

The following displays export and import policy configuration examples.

```
A:ALA-1>config>router# info
----------------------------------------------
            export "LDP-export"
            fec-originate 192.168.1.1/32 pop
            fec-originate 192.168.2.1/32 advertised-label 1000 next-hop 10.10.1.2
            import "LDP-import"
            session-parameters
            exit
            interface-parameters
            exit
            targeted-session
            exit
----------------------------------------------
A:ALA-1>config>router#
```

7.29.3.5 Targeted Session Parameters

Use the following syntax to specify **targeted-session** parameters:

CLI Syntax:

```
config>router# ldp
targeted-session
  disable-targeted-session
  export-prefixes policy-name [policy-name...(up to 5
    max)]
  ipv4
    hello timeout factor
    keepalive timeout factor
  import-prefixes policy-name [policy-name...(up to 5
    max)]
  peer ip-address
    hello timeout factor
    keepalive timeout factor
    no shutdown
    tunneling
      lsp lsp-name
```

The following example displays an LDP configuration example:

```
A:ALA-1>config>router>ldp# info
----------------------------------------------
...
            targeted-session
                ipv4
                    hello 120 3
                    keepalive 120 3
                exit
                peer 10.10.10.104
                    hello 240 3
                    keepalive 240 3
                exit
            exit
----------------------------------------------
A:ALA-1>config>router>ldp#
```

7.29.3.6 Interface Parameters

Use the following syntax to configure interface parameters:

CLI Syntax:

```
config>router# ldp
interface-parameters
  interface ip-int-name [dual stack]
    bfd-enable [ipv4][ipv6]
    ipv4/ipv6
```

```
            hello timeout factor
            keepalive timeout factor
            transport-address {system | interface}
          no shutdown
        ipv4/ipv6
          hello timeout factor
          keepalive timeout factor
          transport-address {system | interface}
```

The following example displays an interface parameter configuration example:

```
A:ALA-1>config>router>ldp# info
----------------------------------------------
...
            interface-parameters
                interface "to-DUT1" dual-stack
                    ipv4
                        hello 240 3
                        keepalive 240 3
                    exit
                exit
            exit
----------------------------------------------
A:ALA-1>config>router>ldp#
```

7.29.3.7 Session Parameters

Use the following syntax to specify session parameters:

CLI Syntax:

```
config>router# ldp
session-parameters
  peer ip-address
tcp-session-parameters
  peer transport ip-address
    auth-keychain name
    authentication-key [authentication-key | hash-key]
      [hash | hash2 | custom]
    ttl-security min-ttl-value [log log-id]
```

The following example displays an LDP configuration example:

```
A:ALA-1>config>router>ldp# info
----------------------------------------------
            export "LDP-export"
            import "LDP-import"
            session-parameters
                peer 10.1.1.1
                exit
                peer 10.10.10.104
                exit
            exit
            tcp-session-parameters
                peer-transport 10.10.10.104
                    authentication-key "E7GtYNZHTAaQqVMRDbfNIZpLtHg4ECOk" hash2
                exit
            exit
            interface-parameters
                interface "to-DUT1" dual-stack
                    ipv4
                        hello 240 3
                        keepalive 240 3
                    exit
                exit
            exit
            targeted-session
                ipv4
                    hello 120 3
                    keepalive 120 3
                exit
                peer 10.10.10.104
                    hello 240 3
                    keepalive 240 3
                exit
            exit
----------------------------------------------
A:ALA-1>config>router>ldp#
```

7.29.3.8 LDP Signaling and Services

When LDP is enabled, targeted sessions can be established to create remote adjacencies with nodes that are not directly connected. When service distribution paths (SDPs) are configured, extended discovery mechanisms enable LDP to send periodic targeted hello messages to the SDP far-end point. The exchange of LDP hellos trigger session establishment. The SDP signaling default enables **tldp**. The service SDP uses the targeted-session parameters configured in the **config>router>ldp>targeted-session** context.

The SDP LDP and LSP commands are mutually exclusive; either one LSP can be specified or LDP can be enabled. If LDP is already enabled on an MPLS SDP, then an LSP cannot be specified on the SDP. If an LSP is specified on an MPLS SDP, then LDP cannot be enabled on the SDP.

To enable LDP on the SDP when an LSP is already specified, the LSP must be removed from the configuration using the no lsp lsp-name command. For more information about configuring SDPs, refer to the *7450 ESS, 7750 SR, 7950 XRS, and VSR Services Overview Guide*.

The following example displays the command syntax usage to configure enable LDP on an MPLS SDP:

CLI Syntax:

```
config>service>sdp#
ldp
signaling {off | tldp}
```

The following displays an example of an SDP configuration showing the signaling default tldp enabled.

```
A:ALA-1>config>service>sdp# info detail
----------------------------------------------
            description "MPLS: to-99"
            far-end 10.10.10.99
            signaling tldp
            path-mtu 4462
            keep-alive
                hello-time 10
                hold-down-time 10
                max-drop-count 3
                timeout 5
                no message-length
                no shutdown
            exit
            no shutdown
----------------------------------------------
A:ALA-1>config>service>sdp#
```

The following displays an example of an SDP configuration for the 7750 SR, showing the signaling default tldp enabled.

```
A:ALA-1>config>service>sdp# info detail
----------------------------------------------
            description "MPLS: to-99"
            far-end 10.10.10.99
            ldp
            signaling tldp
            path-mtu 4462
            keep-alive
                hello-time 10
                hold-down-time 10
                max-drop-count 3
                timeout 5
                no message-length
                no shutdown
            exit
            no shutdown
----------------------------------------------
A:ALA-1>config>service>sdp#
```

The following shows a working configuration of LDP over RSVP-TE (1) where tunnels look like the second example (2):

Example 1: LDP over RSVP-TE

```
*A:ALA-1>config>router>ldp# info
----------------------------------------------
            prefer-tunnel-in-tunnel
            interface-parameters
                interface "port-1/1/3"
                exit
                interface "port-lag-1"
                exit
            exit
            targeted-session
                peer 10.51.0.1
                    shutdown
                    tunneling
                        lsp "to_P_1"
                    exit
                exit
                peer 10.51.0.17
                    shutdown
                    tunneling
                        lsp "to_P_6"
                    exit
                exit
            exit
----------------------------------------------
*A:ALA-1>config>router>ldp#
```

Example 2: Tunnels

```
ALA-1>config>router>if-attr# info
----------------------------------------------

admin-group "lower" value 2
admin-group "upper" value 1
----------------------------------------------
*A:ALA-1>config>router>mpls# info
----------------------------------------------
            resignal-timer 30
            interface "system"
            exit
            interface "port-1/1/3"
            exit
            interface "port-lag-1"
            exit
            path "dyn"
                no shutdown
            exit
            lsp "to_P_1"
                to 10.51.0.1
                cspf
                fast-reroute facility
                exit
                primary "dyn"
                exit
                no shutdown
            exit
            lsp "to_P_6"
                to 10.51.0.17
                cspf
                fast-reroute facility
                exit
                primary "dyn"
                exit
                no shutdown
            exit
            no shutdown
----------------------------------------------
*A:ALA-1>config>router>mpls#
```

7.30 LDP Configuration Management Tasks

This section discusses LDP configuration management tasks.

7.30.1 Disabling LDP

The **no ldp** command disables the LDP protocol on the router. All parameters revert to the default settings. LDP must be shut down before it can be disabled.

Use the following command syntax to disable LDP:

CLI Syntax:

```
no ldp
shutdown
```

7.30.2 Modifying Targeted Session Parameters

The modification of LDP targeted session parameters does not take effect until the next time the session goes down and is re-establishes. Individual parameters cannot be deleted. The no form of a **targeted-session** parameter command reverts modified values back to the default. Different default parameters can be configured for IPv4 and IPv6 LDP targeted hello adjacencies.

The following example displays the command syntax usage to revert targeted session parameters back to the default values:

Example:

```
config>router# ldp
config>router>ldp# targeted-session
config>router>ldp>tcp-session-params>peer# no
 authentication-key
config>router>ldp>targ-session# no disable-targeted-
 session
config>router>ldp>targ-session>ipv4# no hello
config>router>ldp>targ-session>ipv4# no keepalive
config>router>ldp>targ-session# no peer 10.10.10.104
```

The following output displays the default values:

```
A:ALA-1>config>router>ldp>targeted# info detail
----------------------------------------------
               no disable-targeted-session
               no import-prefixes
               no export-prefixes
```

```
                ipv4
                    no hello
                    no keepalive
                    no hello-reduction
                exit
                ipv6
                    no hello
                    no keepalive
                    no hello-reduction
                exit
----------------------------------------------
A:ALA-1>config>router>ldp>targeted#
```

7.30.3 Modifying Interface Parameters

Individual parameters cannot be deleted. The **no** form of an **interface-parameter** command reverts modified values back to the defaults. The modification of LDP targeted session parameters does not take effect until the next time the session goes down and is re-establishes.

The following output displays the default values:

```
A:ALA-1>config>router>ldp>if-params>if# info detail
----------------------------------------------
                    no bfd-enable
                    ipv4
                        no hello
                        no keepalive
                        no local-lsr-id
                        fec-type-capability
                            p2mp-ipv4 enable
                        exit
                        no transport-address
                        no shutdown
                    exit
                    no shutdown
----------------------------------------------
```

8 Standards and Protocol Support

Note: The information provided in this chapter is subject to change without notice and may not apply to all platforms.

Nokia assumes no responsibility for inaccuracies.

Access Node Control Protocol (ANCP)

draft-ietf-ancp-protocol-02, *Protocol for Access Node Control Mechanism in Broadband Networks*

RFC 5851, *Framework and Requirements for an Access Node Control Mechanism in Broadband Multi-Service Networks*

Application Assurance (AA)

3GPP Release 12, *ADC rules over Gx interfaces*

RFC 3507, *Internet Content Adaptation Protocol (ICAP)*

Asynchronous Transfer Mode (ATM)

AF-ILMI-0065.000 Version 4.0, *Integrated Local Management Interface (ILMI)*

AF-PHY-0086.001 Version 1.1, *Inverse Multiplexing for ATM (IMA) Specification*

AF-TM-0121.000 Version 4.1, *Traffic Management Specification*

GR-1113-CORE Issue 1, *Asynchronous Transfer Mode (ATM) and ATM Adaptation Layer (AAL) Protocols Generic Requirements*

GR-1248-CORE Issue 3, *Generic Requirements for Operations of ATM Network Elements (NEs)*

RFC 1626, *Default IP MTU for use over ATM AAL5*

RFC 2684, *Multiprotocol Encapsulation over ATM Adaptation Layer 5*

Bidirectional Forwarding Detection (BFD)

draft-ietf-idr-bgp-ls-sbfd-extensions-01, *BGP Link-State Extensions for Seamless BFD*

RFC 5880, *Bidirectional Forwarding Detection (BFD)*

RFC 5881, *Bidirectional Forwarding Detection (BFD) IPv4 and IPv6 (Single Hop)*

RFC 5882, *Generic Application of Bidirectional Forwarding Detection (BFD)*

RFC 5883, *Bidirectional Forwarding Detection (BFD) for Multihop Paths*

RFC 7130, *Bidirectional Forwarding Detection (BFD) on Link Aggregation Group (LAG) Interfaces*

RFC 7880, *Seamless Bidirectional Forwarding Detection (S-BFD)*

RFC 7881, *Seamless Bidirectional Forwarding Detection (S-BFD) for IPv4, IPv6, and MPLS*

RFC 7883, *Advertising Seamless Bidirectional Forwarding Detection (S-BFD) Discriminators in IS-IS*

RFC 7884, *OSPF Extensions to Advertise Seamless Bidirectional Forwarding Detection (S-BFD) Target Discriminators*

Border Gateway Protocol (BGP)

draft-hares-idr-update-attrib-low-bits-fix-01, *Update Attribute Flag Low Bits Clarification*

draft-ietf-idr-add-paths-guidelines-08, *Best Practices for Advertisement of Multiple Paths in IBGP*

draft-ietf-idr-best-external-03, *Advertisement of the best external route in BGP*

draft-ietf-idr-bgp-flowspec-oid-03, *Revised Validation Procedure for BGP Flow Specifications*

draft-ietf-idr-bgp-gr-notification-01, *Notification Message support for BGP Graceful Restart*

draft-ietf-idr-bgp-ls-app-specific-attr-01, *Application Specific Attributes Advertisement with BGP Link-State*

draft-ietf-idr-bgp-optimal-route-reflection-10, *BGP Optimal Route Reflection (BGP-ORR)*

draft-ietf-idr-error-handling-03, *Revised Error Handling for BGP UPDATE Messages*

draft-ietf-idr-flowspec-interfaceset-03, *Applying BGP flowspec rules on a specific interface set*

draft-ietf-idr-flowspec-path-redirect-05, *Flowspec Indirection-id Redirect* - localised ID

draft-ietf-idr-flowspec-redirect-ip-02, *BGP Flow-Spec Redirect to IP Action*

draft-ietf-idr-link-bandwidth-03, *BGP Link Bandwidth Extended Community*

draft-ietf-idr-long-lived-gr-00, *Support for Long-lived BGP Graceful Restart*

draft-ietf-sidr-origin-validation-signaling-04, *BGP Prefix Origin Validation State Extended Community*

RFC 1772, *Application of the Border Gateway Protocol in the Internet*

RFC 1997, *BGP Communities Attribute*

RFC 2385, *Protection of BGP Sessions via the TCP MD5 Signature Option*

RFC 2439, *BGP Route Flap Damping*

RFC 2545, *Use of BGP-4 Multiprotocol Extensions for IPv6 Inter-Domain Routing*

RFC 2858, *Multiprotocol Extensions for BGP-4*

RFC 2918, *Route Refresh Capability for BGP-4*

RFC 3107, *Carrying Label Information in BGP-4*

RFC 4271, *A Border Gateway Protocol 4 (BGP-4)*

RFC 4360, *BGP Extended Communities Attribute*

RFC 4364, *BGP/MPLS IP Virtual Private Networks (VPNs)*

RFC 4456, *BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP)*

RFC 4486, *Subcodes for BGP Cease Notification Message*

RFC 4659, *BGP-MPLS IP Virtual Private Network (VPN) Extension for IPv6 VPN*

RFC 4684, *Constrained Route Distribution for Border Gateway Protocol/ MultiProtocol Label Switching (BGP/MPLS) Internet Protocol (IP) Virtual Private Networks (VPNs)*

RFC 4724, *Graceful Restart Mechanism for BGP* - helper mode

RFC 4760, *Multiprotocol Extensions for BGP-4*

RFC 4798, *Connecting IPv6 Islands over IPv4 MPLS Using IPv6 Provider Edge Routers (6PE)*

RFC 5004, *Avoid BGP Best Path Transitions from One External to Another*

RFC 5065, *Autonomous System Confederations for BGP*

RFC 5291, *Outbound Route Filtering Capability for BGP-4*

RFC 5396, *Textual Representation of Autonomous System (AS) Numbers* - asplain

RFC 5492, *Capabilities Advertisement with BGP-4*

RFC 5549, *Advertising IPv4 Network Layer Reachability Information with an IPv6 Next Hop*

RFC 5575, *Dissemination of Flow Specification Rules*

RFC 5668, *4-Octet AS Specific BGP Extended Community*

RFC 6286, *Autonomous-System-Wide Unique BGP Identifier for BGP-4*

RFC 6793, *BGP Support for Four-Octet Autonomous System (AS) Number Space*

RFC 6810, *The Resource Public Key Infrastructure (RPKI) to Router Protocol*

RFC 6811, *Prefix Origin Validation*

RFC 6996, *Autonomous System (AS) Reservation for Private Use*

RFC 7311, *The Accumulated IGP Metric Attribute for BGP*

RFC 7607, *Codification of AS 0 Processing*

RFC 7674, *Clarification of the Flowspec Redirect Extended Community*

RFC 7752, *North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP*

RFC 7854, *BGP Monitoring Protocol (BMP)*

RFC 7911, *Advertisement of Multiple Paths in BGP*

RFC 7999, *BLACKHOLE Community*

RFC 8092, *BGP Large Communities Attribute*

RFC 8212, *Default External BGP (EBGP) Route Propagation Behavior without Policies*

RFC 8571, *BGP - Link State (BGP-LS) Advertisement of IGP Traffic Engineering Performance Metric Extensions*

Broadband Network Gateway (BNG) - Control and User Plane Separation (CUPS)

3GPP 23.007, *Restoration procedures*

3GPP 29.244, *Interface between the Control Plane and the User Plane nodes*

3GPP 29.281, *General Packet Radio System (GPRS) Tunnelling Protocol User Plane (GTPv1-U)*

BBF TR-459, *Control and User Plane Separation for a Disaggregated BNG*

RFC 8300, *Network Service Header (NSH)*

Circuit Emulation

RFC 4553, *Structure-Agnostic Time Division Multiplexing (TDM) over Packet (SAToP)*

RFC 5086, *Structure-Aware Time Division Multiplexed (TDM) Circuit Emulation Service over Packet Switched Network (CESoPSN)*

RFC 5287, *Control Protocol Extensions for the Setup of Time-Division Multiplexing (TDM) Pseudowires in MPLS Networks*

Ethernet

IEEE 802.1AB, *Station and Media Access Control Connectivity Discovery*

IEEE 802.1ad, *Provider Bridges*

IEEE 802.1ag, *Connectivity Fault Management*

IEEE 802.1ah, *Provider Backbone Bridges*

IEEE 802.1ak, *Multiple Registration Protocol*

IEEE 802.1aq, *Shortest Path Bridging*

IEEE 802.1ax, *Link Aggregation*

IEEE 802.1D, *MAC Bridges*

IEEE 802.1p, *Traffic Class Expediting*

IEEE 802.1Q, *Virtual LANs*

IEEE 802.1s, *Multiple Spanning Trees*

IEEE 802.1w, *Rapid Reconfiguration of Spanning Tree*

IEEE 802.1X, *Port Based Network Access Control*

IEEE 802.3ac, *VLAN Tag*

IEEE 802.3ad, *Link Aggregation*

IEEE 802.3ah, *Ethernet in the First Mile*

IEEE 802.3x, *Ethernet Flow Control*

ITU-T G.8031/Y.1342, *Ethernet Linear Protection Switching*

ITU-T G.8032/Y.1344, *Ethernet Ring Protection Switching*

ITU-T Y.1731, *OAM functions and mechanisms for Ethernet based networks*

Ethernet VPN (EVPN)

draft-ietf-bess-evpn-igmp-mld-proxy-05, *IGMP and MLD Proxy for EVPN*

draft-ietf-bess-evpn-irb-mcast-04, *EVPN Optimized Inter-Subnet Multicast (OISM) Forwarding* - ingress replication

draft-ietf-bess-evpn-pref-df-06, *Preference-based EVPN DF Election*

draft-ietf-bess-evpn-prefix-advertisement-11, *IP Prefix Advertisement in EVPN*

draft-ietf-bess-evpn-proxy-arp-nd-08, *Operational Aspects of Proxy-ARP/ND in EVPN Networks*

draft-ietf-bess-pbb-evpn-isid-cmacflush-00, *PBB-EVPN ISID-based CMAC-Flush*

RFC 7432, *BGP MPLS-Based Ethernet VPN*

RFC 7623, *Provider Backbone Bridging Combined with Ethernet VPN (PBB-EVPN)*

RFC 8214, *Virtual Private Wire Service Support in Ethernet VPN*

RFC 8317, *Ethernet-Tree (E-Tree) Support in Ethernet VPN (EVPN) an Provider Backbone Bridging EVPN (PBB-EVPN)*

RFC 8365, *A Network Virtualization Overlay Solution Using Ethernet VPN (EVPN)*

RFC 8560, *Seamless Integration of Ethernet VPN (EVPN) with Virtual Private LAN Service (VPLS) and Their Provider Backbone Bridge (PBB) Equivalents*

RFC 8584, *DF Election and AC-influenced DF Election*

Frame Relay

ANSI T1.617 Annex D, *DSS1 - Signalling Specification For Frame Relay Bearer Service*

FRF.1.2, *PVC User-to-Network Interface (UNI) Implementation Agreement*

FRF.12, *Frame Relay Fragmentation Implementation Agreement*

FRF.16.1, *Multilink Frame Relay UNI/NNI Implementation Agreement*

FRF.5, *Frame Relay/ATM PVC Network Interworking Implementation*

FRF2.2, *PVC Network-to-Network Interface (NNI) Implementation Agreement*

ITU-T Q.933 Annex A, *Additional procedures for Permanent Virtual Connection (PVC) status management*

Generalized Multiprotocol Label Switching (GMPLS)

draft-ietf-ccamp-rsvp-te-srlg-collect-04, *RSVP-TE Extensions for Collecting SRLG Information*

RFC 3471, *Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description*

RFC 3473, *Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions*

RFC 4204, *Link Management Protocol (LMP)*

RFC 4208, *Generalized Multiprotocol Label Switching (GMPLS) User-Network Interface (UNI): Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Support for the Overlay Model*

RFC 4872, *RSVP-TE Extensions in Support of End-to-End Generalized Multi-Protocol Label Switching (GMPLS) Recovery*

RFC 5063, *Extensions to GMPLS Resource Reservation Protocol (RSVP) Graceful Restart* - helper mode

gRPC Remote Procedure Calls (gRPC)

cert.proto Version 0.1.0, *gRPC Network Operations Interface (gNOI) Certificate Management Service*

file.proto Version 0.1.0, *gRPC Network Operations Interface (gNOI) File Service*

gnmi.proto Version 0.7.0, *gRPC Network Management Interface (gNMI) Service Specification*

PROTOCOL-HTTP2, *gRPC over HTTP2*

system.proto Version 1.0.0, *gRPC Network Operations Interface (gNOI) System Service*

Intermediate System to Intermediate System (IS-IS)

draft-ietf-isis-mi-02, *IS-IS Multi-Instance*

draft-kaplan-isis-ext-eth-02, *Extended Ethernet Frame Size Support*

ISO/IEC 10589:2002 Second Edition, *Intermediate system to Intermediate system intra-domain routeing information exchange protocol for use in conjunction with the protocol for providing the connectionless-mode Network Service (ISO 8473)*

RFC 1195, *Use of OSI IS-IS for Routing in TCP/IP and Dual Environments*

RFC 2973, *IS-IS Mesh Groups*

RFC 3359, *Reserved Type, Length and Value (TLV) Codepoints in Intermediate System to Intermediate System*

RFC 3719, *Recommendations for Interoperable Networks using Intermediate System to Intermediate System (IS-IS)*

RFC 3787, *Recommendations for Interoperable IP Networks using Intermediate System to Intermediate System (IS-IS)*

RFC 4971, *Intermediate System to Intermediate System (IS-IS) Extensions for Advertising Router Information*

RFC 5120, *M-ISIS: Multi Topology (MT) Routing in IS-IS*

RFC 5130, *A Policy Control Mechanism in IS-IS Using Administrative Tags*

RFC 5301, *Dynamic Hostname Exchange Mechanism for IS-IS*

RFC 5302, *Domain-wide Prefix Distribution with Two-Level IS-IS*

RFC 5303, *Three-Way Handshake for IS-IS Point-to-Point Adjacencies*

RFC 5304, *IS-IS Cryptographic Authentication*

RFC 5305, *IS-IS Extensions for Traffic Engineering TE*

RFC 5306, *Restart Signaling for IS-IS* - helper mode

RFC 5307, *IS-IS Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)*

RFC 5308, *Routing IPv6 with IS-IS*

RFC 5309, *Point-to-Point Operation over LAN in Link State Routing Protocols*

RFC 5310, *IS-IS Generic Cryptographic Authentication*

RFC 6119, *IPv6 Traffic Engineering in IS-IS*

RFC 6213, *IS-IS BFD-Enabled TLV*

RFC 6232, *Purge Originator Identification TLV for IS-IS*

RFC 6233, *IS-IS Registry Extension for Purges*

RFC 6329, *IS-IS Extensions Supporting IEEE 802.1aq Shortest Path Bridging*

RFC 7775, *IS-IS Route Preference for Extended IP and IPv6 Reachability*

RFC 7794, *IS-IS Prefix Attributes for Extended IPv4 and IPv6 Reachability*

RFC 7987, *IS-IS Minimum Remaining Lifetime*

RFC 8202, *IS-IS Multi-Instance* - single topology

RFC 8570, *IS-IS Traffic Engineering (TE) Metric Extensions* - delay metric

RFC 8919, *IS-IS Application-Specific Link Attributes*

Internet Protocol (IP) — Fast Reroute

draft-ietf-rtgwg-lfa-manageability-08, *Operational management of Loop Free Alternates*

RFC 5286, *Basic Specification for IP Fast Reroute: Loop-Free Alternates*

RFC 7431, *Multicast-Only Fast Reroute*

RFC 7490, *Remote Loop-Free Alternate (LFA) Fast Reroute (FRR)*

Internet Protocol (IP) — General

draft-grant-tacacs-02, *The TACACS+ Protocol*

RFC 768, *User Datagram Protocol*

RFC 793, *Transmission Control Protocol*

RFC 854, *Telnet Protocol Specifications*

RFC 1350, *The TFTP Protocol (revision 2)*

RFC 2347, *TFTP Option Extension*

RFC 2348, *TFTP Blocksize Option*

RFC 2349, *TFTP Timeout Interval and Transfer Size Options*

RFC 2428, *FTP Extensions for IPv6 and NATs*

RFC 2784, *Generic Routing Encapsulation (GRE)*

RFC 2818, *HTTP Over TLS*

RFC 2890, *Key and Sequence Number Extensions to GRE*

RFC 3164, *The BSD syslog Protocol*

RFC 4250, *The Secure Shell (SSH) Protocol Assigned Numbers*

RFC 4251, *The Secure Shell (SSH) Protocol Architecture*

RFC 4252, *The Secure Shell (SSH) Authentication Protocol* - publickey, password

RFC 4253, *The Secure Shell (SSH) Transport Layer Protocol*

RFC 4254, *The Secure Shell (SSH) Connection Protocol*

RFC 4511, *Lightweight Directory Access Protocol (LDAP): The Protocol*

RFC 4513, *Lightweight Directory Access Protocol (LDAP): Authentication Methods and Security Mechanisms* - TLS

RFC 4632, *Classless Inter-domain Routing (CIDR): The Internet Address Assignment and Aggregation Plan*

RFC 5082, *The Generalized TTL Security Mechanism (GTSM)*

RFC 5246, *The Transport Layer Security (TLS) Protocol Version 1.2* - TLS client, RSA public key

RFC 5656, *Elliptic Curve Algorithm Integration in the Secure Shell Transport Layer* - ECDSA

RFC 5925, *The TCP Authentication Option*

RFC 5926, *Cryptographic Algorithms for the TCP Authentication Option (TCP-AO)*

RFC 6398, *IP Router Alert Considerations and Usage* - MLD

RFC 6528, *Defending against Sequence Number Attacks*

RFC 7011, *Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information*

RFC 7012, *Information Model for IP Flow Information Export*

RFC 7230, *Hypertext Transfer Protocol (HTTP/1.1): Message Syntax and Routing*

RFC 7231, *Hypertext Transfer Protocol (HTTP/1.1): Semantics and Content*

RFC 7232, *Hypertext Transfer Protocol (HTTP/1.1): Conditional Requests*

RFC 7301, *Transport Layer Security (TLS) Application Layer Protocol Negotiation Extension*

Internet Protocol (IP) — Multicast

cisco-ipmulticast/pim-autorp-spec01, *Auto-RP: Automatic discovery of Group-to-RP mappings for IP multicast* - version 1

draft-ietf-bier-pim-signaling-08, *PIM Signaling Through BIER Core*

draft-ietf-idmr-traceroute-ipm-07, *A "traceroute" facility for IP Multicast*

draft-ietf-l2vpn-vpls-pim-snooping-07, *Protocol Independent Multicast (PIM) over Virtual Private LAN Service (VPLS)*

RFC 1112, *Host Extensions for IP Multicasting*

RFC 2236, *Internet Group Management Protocol, Version 2*

RFC 2365, *Administratively Scoped IP Multicast*

RFC 2375, *IPv6 Multicast Address Assignments*

RFC 2710, *Multicast Listener Discovery (MLD) for IPv6*

RFC 3306, *Unicast-Prefix-based IPv6 Multicast Addresses*

RFC 3376, *Internet Group Management Protocol, Version 3*

RFC 3446, *Anycast Rendevous Point (RP) mechanism using Protocol Independent Multicast (PIM) and Multicast Source Discovery Protocol (MSDP)*

RFC 3590, *Source Address Selection for the Multicast Listener Discovery (MLD) Protocol*

RFC 3618, *Multicast Source Discovery Protocol (MSDP)*

RFC 3810, *Multicast Listener Discovery Version 2 (MLDv2) for IPv6*

RFC 3956, *Embedding the Rendezvous Point (RP) Address in an IPv6 Multicast Address*

RFC 3973, *Protocol Independent Multicast - Dense Mode (PIM-DM): Protocol Specification (Revised)* - auto-RP groups

RFC 4541, *Considerations for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Snooping Switches*

RFC 4604, *Using Internet Group Management Protocol Version 3 (IGMPv3) and Multicast Listener Discovery Protocol Version 2 (MLDv2) for Source-Specific Multicast*

RFC 4607, *Source-Specific Multicast for IP*

RFC 4608, *Source-Specific Protocol Independent Multicast in 232/8*

RFC 4610, *Anycast-RP Using Protocol Independent Multicast (PIM)*

RFC 4611, *Multicast Source Discovery Protocol (MSDP) Deployment Scenarios*

RFC 5059, *Bootstrap Router (BSR) Mechanism for Protocol Independent Multicast (PIM)*

RFC 5186, *Internet Group Management Protocol Version 3 (IGMPv3) / Multicast Listener Discovery Version 2 (MLDv2) and Multicast Routing Protocol Interaction*

RFC 5384, *The Protocol Independent Multicast (PIM) Join Attribute Format*

RFC 5496, *The Reverse Path Forwarding (RPF) Vector TLV*

RFC 6037, *Cisco Systems' Solution for Multicast in MPLS/BGP IP VPNs*

RFC 6512, *Using Multipoint LDP When the Backbone Has No Route to the Root*

RFC 6513, *Multicast in MPLS/BGP IP VPNs*

RFC 6514, *BGP Encodings and Procedures for Multicast in MPLS/IP VPNs*

RFC 6515, *IPv4 and IPv6 Infrastructure Addresses in BGP Updates for Multicast VPNs*

RFC 6516, *IPv6 Multicast VPN (MVPN) Support Using PIM Control Plane and Selective Provider Multicast Service Interface (S-PMSI) Join Messages*

RFC 6625, *Wildcards in Multicast VPN Auto-Discover Routes*

RFC 6826, *Multipoint LDP In-Band Signaling for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Path*

RFC 7246, *Multipoint Label Distribution Protocol In-Band Signaling in a Virtual Routing and Forwarding (VRF) Table Context*

RFC 7385, *IANA Registry for P-Multicast Service Interface (PMSI) Tunnel Type Code Points*

RFC 7716, *Global Table Multicast with BGP Multicast VPN (BGP-MVPN) Procedures*

RFC 7761, *Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)*

RFC 8279, *Multicast Using Bit Index Explicit Replication (BIER)*

RFC 8296, *Encapsulation for Bit Index Explicit Replication (BIER) in MPLS and Non-MPLS Networks* - MPLS encapsulation

RFC 8401, *Bit Index Explicit Replication (BIER) Support via IS-IS*

RFC 8444, *OSPFv2 Extensions for Bit Index Explicit Replication (BIER)*

RFC 8487, *Mtrace Version 2: Traceroute Facility for IP Multicast*

RFC 8534, *Explicit Tracking with Wildcard Routes in Multicast VPN* - (C-*,C-*) wildcard

RFC 8556, *Multicast VPN Using Bit Index Explicit Replication (BIER)*

Internet Protocol (IP) — Version 4

RFC 791, *Internet Protocol*

RFC 792, *Internet Control Message Protocol*

RFC 826, *An Ethernet Address Resolution Protocol*

RFC 951, *Bootstrap Protocol (BOOTP)* - relay

RFC 1034, *Domain Names - Concepts and Facilities*

RFC 1035, *Domain Names - Implementation and Specification*

RFC 1191, *Path MTU Discovery* - router specification

RFC 1519, *Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy*

RFC 1534, *Interoperation between DHCP and BOOTP*

RFC 1542, *Clarifications and Extensions for the Bootstrap Protocol*

RFC 1812, *Requirements for IPv4 Routers*

RFC 1918, *Address Allocation for Private Internets*

RFC 2003, *IP Encapsulation within IP*

RFC 2131, *Dynamic Host Configuration Protocol*

RFC 2132, *DHCP Options and BOOTP Vendor Extensions*

RFC 2401, *Security Architecture for Internet Protocol*

RFC 3021, *Using 31-Bit Prefixes on IPv4 Point-to-Point Links*

RFC 3046, *DHCP Relay Agent Information Option (Option 82)*

RFC 3768, *Virtual Router Redundancy Protocol (VRRP)*

RFC 4884, *Extended ICMP to Support Multi-Part Messages* - ICMPv4 and ICMPv6 Time Exceeded

Internet Protocol (IP) — Version 6

RFC 2464, *Transmission of IPv6 Packets over Ethernet Networks*

RFC 2529, *Transmission of IPv6 over IPv4 Domains without Explicit Tunnels*

RFC 3122, *Extensions to IPv6 Neighbor Discovery for Inverse Discovery Specification*

RFC 3315, *Dynamic Host Configuration Protocol for IPv6 (DHCPv6)*

RFC 3587, *IPv6 Global Unicast Address Format*

RFC 3596, *DNS Extensions to Support IP version 6*

RFC 3633, *IPv6 Prefix Options for Dynamic Host Configuration Protocol (DHCP) version 6*

RFC 3646, *DNS Configuration options for Dynamic Host Configuration Protocol for IPv6 (DHCPv6)*

RFC 3736, *Stateless Dynamic Host Configuration Protocol (DHCP) Service for IPv6*

RFC 3971, *SEcure Neighbor Discovery (SEND)*

RFC 3972, *Cryptographically Generated Addresses (CGA)*

RFC 4007, *IPv6 Scoped Address Architecture*

RFC 4193, *Unique Local IPv6 Unicast Addresses*

RFC 4291, *Internet Protocol Version 6 (IPv6) Addressing Architecture*

RFC 4443, *Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification*

RFC 4861, *Neighbor Discovery for IP version 6 (IPv6)*

RFC 4862, *IPv6 Stateless Address Autoconfiguration* - router functions

RFC 4890, *Recommendations for Filtering ICMPv6 Messages in Firewalls*

RFC 4941, *Privacy Extensions for Stateless Address Autoconfiguration in IPv6*

RFC 5007, *DHCPv6 Leasequery*

RFC 5095, *Deprecation of Type 0 Routing Headers in IPv6*

RFC 5722, *Handling of Overlapping IPv6 Fragments*

RFC 5798, *Virtual Router Redundancy Protocol (VRRP) Version 3 for IPv4 and IPv6* - IPv6

RFC 5952, *A Recommendation for IPv6 Address Text Representation*

RFC 6092, *Recommended Simple Security Capabilities in Customer Premises Equipment (CPE) for Providing Residential IPv6 Internet Service* - Internet Control and Management, Upper-Layer Transport Protocols, UDP Filters, IPsec and Internet Key Exchange (IKE), TCP Filters

RFC 6106, *IPv6 Router Advertisement Options for DNS Configuration*

RFC 6164, *Using 127-Bit IPv6 Prefixes on Inter-Router Links*

RFC 8021, *Generation of IPv6 Atomic Fragments Considered Harmful*

RFC 8200, *Internet Protocol, Version 6 (IPv6) Specification*

RFC 8201, *Path MTU Discovery for IP version 6*

Internet Protocol Security (IPsec)

draft-ietf-ipsec-isakmp-mode-cfg-05, *The ISAKMP Configuration Method*

draft-ietf-ipsec-isakmp-xauth-06, *Extended Authentication within ISAKMP/Oakley (XAUTH)*

RFC 2401, *Security Architecture for the Internet Protocol*

RFC 2403, *The Use of HMAC-MD5-96 within ESP and AH*

RFC 2404, *The Use of HMAC-SHA-1-96 within ESP and AH*

RFC 2405, *The ESP DES-CBC Cipher Algorithm With Explicit IV*

RFC 2406, *IP Encapsulating Security Payload (ESP)*

RFC 2407, *IPsec Domain of Interpretation for ISAKMP (IPsec DoI)*

RFC 2408, *Internet Security Association and Key Management Protocol (ISAKMP)*

RFC 2409, *The Internet Key Exchange (IKE)*

RFC 2410, *The NULL Encryption Algorithm and Its Use With IPsec*

RFC 3526, *More Modular Exponential (MODP) Diffie-Hellman group for Internet Key Exchange (IKE)*

RFC 3566, *The AES-XCBC-MAC-96 Algorithm and Its Use With IPsec*

RFC 3602, *The AES-CBC Cipher Algorithm and Its Use with IPsec*

RFC 3706, *A Traffic-Based Method of Detecting Dead Internet Key Exchange (IKE) Peers*

RFC 3947, *Negotiation of NAT-Traversal in the IKE*

RFC 3948, *UDP Encapsulation of IPsec ESP Packets*

RFC 4106, *The Use of Galois/Counter Mode (GCM) in IPsec ESP*

RFC 4210, *Internet X.509 Public Key Infrastructure Certificate Management Protocol (CMP)*

RFC 4211, *Internet X.509 Public Key Infrastructure Certificate Request Message Format (CRMF)*

RFC 4301, *Security Architecture for the Internet Protocol*

RFC 4303, *IP Encapsulating Security Payload*

RFC 4307, *Cryptographic Algorithms for Use in the Internet Key Exchange Version 2 (IKEv2)*

RFC 4308, *Cryptographic Suites for IPsec*

RFC 4434, *The AES-XCBC-PRF-128 Algorithm for the Internet Key Exchange Protocol (IKE)*

RFC 4543, *The Use of Galois Message Authentication Code (GMAC) in IPsec ESP and AH*

RFC 4868, *Using HMAC-SHA-256, HMAC-SHA-384, and HMAC-SHA-512 with IPSec*

RFC 4945, *The Internet IP Security PKI Profile of IKEv1/ISAKMP, IKEv2 and PKIX*

RFC 5019, *The Lightweight Online Certificate Status Protocol (OCSP) Profile for High-Volume Environments*

RFC 5280, *Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile*

RFC 5282, *Using Authenticated Encryption Algorithms with the Encrypted Payload of the IKEv2 Protocol*

RFC 5903, *ECP Groups for IKE and IKEv2*

RFC 5998, *An Extension for EAP-Only Authentication in IKEv2*

RFC 6379, *Suite B Cryptographic Suites for IPsec*

RFC 6380, *Suite B Profile for Internet Protocol Security (IPsec)*

RFC 6712, *Internet X.509 Public Key Infrastructure -- HTTP Transfer for the Certificate Management Protocol (CMP)*

RFC 6960, *X.509 Internet Public Key Infrastructure Online Certificate Status Protocol - OCSP*

RFC 7296, *Internet Key Exchange Protocol Version 2 (IKEv2)*

RFC 7321, *Cryptographic Algorithm Implementation Requirements and Usage Guidance for Encapsulating Security Payload (ESP) and Authentication Header (AH)*

RFC 7383, *Internet Key Exchange Protocol Version 2 (IKEv2) Message Fragmentation*

RFC 7427, *Signature Authentication in the Internet Key Exchange Version 2 (IKEv2)*

RFC 7468, *Textual Encodings of PKIX, PKCS, and CMS Structures*

Label Distribution Protocol (LDP)

draft-pdutta-mpls-ldp-adj-capability-00, *LDP Adjacency Capabilities*

draft-pdutta-mpls-ldp-v2-00, *LDP Version 2*

draft-pdutta-mpls-mldp-up-redundancy-00, *Upstream LSR Redundancy for Multi-point LDP Tunnels*

draft-pdutta-mpls-multi-ldp-instance-00, *Multiple LDP Instances*

draft-pdutta-mpls-tldp-hello-reduce-04, *Targeted LDP Hello Reduction*

RFC 3037, *LDP Applicability*

RFC 3478, *Graceful Restart Mechanism for Label Distribution Protocol* - helper mode

RFC 5036, *LDP Specification*

RFC 5283, *LDP Extension for Inter-Area Label Switched Paths (LSPs)*

RFC 5443, *LDP IGP Synchronization*

RFC 5561, *LDP Capabilities*

RFC 5919, *Signaling LDP Label Advertisement Completion*

RFC 6388, *Label Distribution Protocol Extensions for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths*

RFC 6512, *Using Multipoint LDP When the Backbone Has No Route to the Root*

RFC 6826, *Multipoint LDP in-band signaling for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths*

RFC 7032, *LDP Downstream-on-Demand in Seamless MPLS*

RFC 7473, *Controlling State Advertisements of Non-negotiated LDP Applications*

RFC 7552, *Updates to LDP for IPv6*

Layer Two Tunneling Protocol (L2TP) Network Server (LNS)

draft-mammoliti-l2tp-accessline-avp-04, *Layer 2 Tunneling Protocol (L2TP) Access Line Information Attribute Value Pair (AVP) Extensions*

RFC 2661, *Layer Two Tunneling Protocol "L2TP"*

RFC 2809, *Implementation of L2TP Compulsory Tunneling via RADIUS*

RFC 3438, *Layer Two Tunneling Protocol (L2TP) Internet Assigned Numbers: Internet Assigned Numbers Authority (IANA) Considerations Update*

RFC 3931, *Layer Two Tunneling Protocol - Version 3 (L2TPv3)*

RFC 4719, *Transport of Ethernet Frames over Layer 2 Tunneling Protocol Version 3 (L2TPv3)*

RFC 4951, *Fail Over Extensions for Layer 2 Tunneling Protocol (L2TP) "failover"*

Multiprotocol Label Switching (MPLS)

draft-ietf-mpls-lsp-ping-ospfv3-codepoint-02, *OSPFv3 CodePoint for MPLS LSP Ping*

RFC 3031, *Multiprotocol Label Switching Architecture*

RFC 3032, *MPLS Label Stack Encoding*

RFC 3270, *Multi-Protocol Label Switching (MPLS) Support of Differentiated Services* - E-LSP

RFC 3443, *Time To Live (TTL) Processing in Multi-Protocol Label Switching (MPLS) Networks*

RFC 4023, *Encapsulating MPLS in IP or Generic Routing Encapsulation (GRE)*

RFC 4182, *Removing a Restriction on the use of MPLS Explicit NULL*

RFC 5332, *MPLS Multicast Encapsulations*

RFC 5884, *Bidirectional Forwarding Detection (BFD) for MPLS Label Switched Paths (LSPs)*

RFC 6374, *Packet Loss and Delay Measurement for MPLS Networks* - Delay Measurement, Channel Type 0x000C

RFC 6424, *Mechanism for Performing Label Switched Path Ping (LSP Ping) over MPLS Tunnels*

RFC 6425, *Detecting Data Plane Failures in Point-to-Multipoint Multiprotocol Label Switching (MPLS) - Extensions to LSP Ping*

RFC 6790, *The Use of Entropy Labels in MPLS Forwarding*

RFC 7510, *Encapsulating MPLS in UDP*

RFC 7746, *Label Switched Path (LSP) Self-Ping*

RFC 7876, *UDP Return Path for Packet Loss and Delay Measurement for MPLS Networks* - Delay Measurement

RFC 8029, *Detecting Multiprotocol Label Switched (MPLS) Data-Plane Failures*

Multiprotocol Label Switching — Transport Profile (MPLS-TP)

RFC 5586, *MPLS Generic Associated Channel*

RFC 5921, *A Framework for MPLS in Transport Networks*

RFC 5960, *MPLS Transport Profile Data Plane Architecture*

RFC 6370, *MPLS Transport Profile (MPLS-TP) Identifiers*

RFC 6378, *MPLS Transport Profile (MPLS-TP) Linear Protection*

RFC 6426, *MPLS On-Demand Connectivity and Route Tracing*

RFC 6427, *MPLS Fault Management Operations, Administration, and Maintenance (OAM)*

RFC 6428, *Proactive Connectivity Verification, Continuity Check and Remote Defect indication for MPLS Transport Profile*

RFC 6478, *Pseudowire Status for Static Pseudowires*

RFC 7213, *MPLS Transport Profile (MPLS-TP) Next-Hop Ethernet Addressing*

Network Address Translation (NAT)

draft-ietf-behave-address-format-10, *IPv6 Addressing of IPv4/IPv6 Translators*

draft-ietf-behave-v6v4-xlate-23, *IP/ICMP Translation Algorithm*

draft-miles-behave-l2nat-00, *Layer2-Aware NAT*

draft-nishitani-cgn-02, *Common Functions of Large Scale NAT (LSN)*

RFC 4787, *Network Address Translation (NAT) Behavioral Requirements for Unicast UDP*

RFC 5382, *NAT Behavioral Requirements for TCP*

RFC 5508, *NAT Behavioral Requirements for ICMP*

RFC 6146, *Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers*

RFC 6333, *Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion*

RFC 6334, *Dynamic Host Configuration Protocol for IPv6 (DHCPv6) Option for Dual-Stack Lite*

RFC 6887, *Port Control Protocol (PCP)*

RFC 6888, *Common Requirements For Carrier-Grade NATs (CGNs)*

RFC 7753, *Port Control Protocol (PCP) Extension for Port-Set Allocation*

RFC 7915, *IP/ICMP Translation Algorithm*

Network Configuration Protocol (NETCONF)

RFC 5277, *NETCONF Event Notifications*

RFC 6020, *YANG - A Data Modeling Language for the Network Configuration Protocol (NETCONF)*

RFC 6022, *YANG Module for NETCONF Monitoring*

RFC 6241, *Network Configuration Protocol (NETCONF)*

RFC 6242, *Using the NETCONF Protocol over Secure Shell (SSH)*

RFC 6243, *With-defaults Capability for NETCONF*

RFC 8342, *Network Management Datastore Architecture (NMDA)* - Startup, Candidate, Running and Intended datastores

RFC 8525, *YANG Library*

RFC 8526, *NETCONF Extensions to Support the Network Management Datastore Architecture* - <get-data> operation

Open Shortest Path First (OSPF)

RFC 1586, *Guidelines for Running OSPF Over Frame Relay Networks*

RFC 1765, *OSPF Database Overflow*

RFC 2328, *OSPF Version 2*

RFC 3101, *The OSPF Not-So-Stubby Area (NSSA) Option*

RFC 3509, *Alternative Implementations of OSPF Area Border Routers*

RFC 3623, *Graceful OSPF Restart Graceful OSPF Restart* - helper mode

RFC 3630, *Traffic Engineering (TE) Extensions to OSPF Version 2*

RFC 4203, *OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)*

RFC 4222, *Prioritized Treatment of Specific OSPF Version 2 Packets and Congestion Avoidance*

RFC 4552, *Authentication/Confidentiality for OSPFv3*

RFC 4576, *Using a Link State Advertisement (LSA) Options Bit to Prevent Looping in BGP/MPLS IP Virtual Private Networks (VPNs)*

RFC 4577, *OSPF as the Provider/Customer Edge Protocol for BGP/MPLS IP Virtual Private Networks (VPNs)*

RFC 5185, *OSPF Multi-Area Adjacency*

RFC 5187, *OSPFv3 Graceful Restart* - helper mode

RFC 5243, *OSPF Database Exchange Summary List Optimization*

RFC 5250, *The OSPF Opaque LSA Option*

RFC 5309, *Point-to-Point Operation over LAN in Link State Routing Protocols*

RFC 5340, *OSPF for IPv6*

RFC 5642, *Dynamic Hostname Exchange Mechanism for OSPF*

RFC 5709, *OSPFv2 HMAC-SHA Cryptographic Authentication*

RFC 5838, *Support of Address Families in OSPFv3*

RFC 6549, *OSPFv2 Multi-Instance Extensions*

RFC 6987, *OSPF Stub Router Advertisement*

RFC 7684, *OSPFv2 Prefix/Link Attribute Advertisement*

RFC 7770, *Extensions to OSPF for Advertising Optional Router Capabilities*

RFC 8362, *OSPFv3 Link State Advertisement (LSA) Extensibility*

RFC 8920, *OSPF Application-Specific Link Attributes*

OpenFlow

TS-007 Version 1.3.1, *OpenFlow Switch Specification* - OpenFlow-hybrid switches

Path Computation Element Protocol (PCEP)

draft-alvarez-pce-path-profiles-04, *PCE Path Profiles*

draft-dhs-spring-pce-sr-p2mp-policy-00, *PCEP extensions for p2mp sr policy*

draft-ietf-pce-segment-routing-08, *PCEP Extensions for Segment Routing*

RFC 5440, *Path Computation Element (PCE) Communication Protocol (PCEP)*

RFC 8281, *PCEP Extensions for PCE-initiated LSP Setup in a Stateful PCE Model*

RFC 8321, *Path Computation Element Communication Protocol (PCEP) Extensions for Stateful PCE*

Point-to-Point Protocol (PPP)

RFC 1332, *The PPP Internet Protocol Control Protocol (IPCP)*

RFC 1377, *The PPP OSI Network Layer Control Protocol (OSINLCP)*

RFC 1661, *The Point-to-Point Protocol (PPP)*

RFC 1662, *PPP in HDLC-like Framing*

RFC 1877, *PPP Internet Protocol Control Protocol Extensions for Name Server Addresses*

RFC 1989, *PPP Link Quality Monitoring*

RFC 1990, *The PPP Multilink Protocol (MP)*

RFC 1994, *PPP Challenge Handshake Authentication Protocol (CHAP)*

RFC 2153, *PPP Vendor Extensions*

RFC 2516, *A Method for Transmitting PPP Over Ethernet (PPPoE)*

RFC 2615, *PPP over SONET/SDH*

RFC 2686, *The Multi-Class Extension to Multi-Link PPP*

RFC 2878, *PPP Bridging Control Protocol (BCP)*

RFC 4638, *Accommodating a Maximum Transit Unit/Maximum Receive Unit (MTU/MRU) Greater Than 1492 in the Point-to-Point Protocol over Ethernet (PPPoE)*

RFC 5072, *IP Version 6 over PPP*

Policy Management and Credit Control

3GPP TS 29.212 Release 11, *Policy and Charging Control (PCC); Reference points* - Gx support as it applies to wireline environment (BNG)

RFC 4006, *Diameter Credit-Control Application*

RFC 6733, *Diameter Base Protocol*

Pseudowire

draft-ietf-l2vpn-vpws-iw-oam-04, *OAM Procedures for VPWS Interworking*

MFA Forum 9.0.0, *The Use of Virtual trunks for ATM/MPLS Control Plane Interworking*

MFA Forum 12.0.0, *Multiservice Interworking - Ethernet over MPLS*

MFA Forum 13.0.0, *Fault Management for Multiservice Interworking v1.0*

MFA Forum 16.0.0, *Multiservice Interworking - IP over MPLS*

RFC 3916, *Requirements for Pseudo-Wire Emulation Edge-to-Edge (PWE3)*

RFC 3985, *Pseudo Wire Emulation Edge-to-Edge (PWE3)*

RFC 4385, *Pseudo Wire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN*

RFC 4446, *IANA Allocations for Pseudowire Edge to Edge Emulation (PWE3)*

RFC 4447, *Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)*

RFC 4448, *Encapsulation Methods for Transport of Ethernet over MPLS Networks*

RFC 4619, *Encapsulation Methods for Transport of Frame Relay over Multiprotocol Label Switching (MPLS) Networks*

RFC 4717, *Encapsulation Methods for Transport Asynchronous Transfer Mode (ATM) over MPLS Networks*

RFC 4816, *Pseudowire Emulation Edge-to-Edge (PWE3) Asynchronous Transfer Mode (ATM) Transparent Cell Transport Service*

RFC 5085, *Pseudowire Virtual Circuit Connectivity Verification (VCCV): A Control Channel for Pseudowires*

RFC 5659, *An Architecture for Multi-Segment Pseudowire Emulation Edge-to-Edge*

RFC 5885, *Bidirectional Forwarding Detection (BFD) for the Pseudowire Virtual Circuit Connectivity Verification (VCCV)*

RFC 6073, *Segmented Pseudowire*

RFC 6310, *Pseudowire (PW) Operations, Administration, and Maintenance (OAM) Message Mapping*

RFC 6391, *Flow-Aware Transport of Pseudowires over an MPLS Packet Switched Network*

RFC 6575, *Address Resolution Protocol (ARP) Mediation for IP Interworking of Layer 2 VPNs*

RFC 6718, *Pseudowire Redundancy*

RFC 6829, *Label Switched Path (LSP) Ping for Pseudowire Forwarding Equivalence Classes (FECs) Advertised over IPv6*

RFC 6870, *Pseudowire Preferential Forwarding Status bit*

RFC 7023, *MPLS and Ethernet Operations, Administration, and Maintenance (OAM) Interworking*

RFC 7267, *Dynamic Placement of Multi-Segment Pseudowires*

RFC 7392, *Explicit Path Routing for Dynamic Multi-Segment Pseudowires* - ER-TLV and ER-HOP IPv4 Prefix

Quality of Service (QoS)

RFC 2430, *A Provider Architecture for Differentiated Services and Traffic Engineering (PASTE)*

RFC 2474, *Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers*

RFC 2597, *Assured Forwarding PHB Group*

RFC 3140, *Per Hop Behavior Identification Codes*

RFC 3246, *An Expedited Forwarding PHB (Per-Hop Behavior)*

Remote Authentication Dial In User Service (RADIUS)

RFC 2865, *Remote Authentication Dial In User Service (RADIUS)*

RFC 2866, *RADIUS Accounting*

RFC 2867, *RADIUS Accounting Modifications for Tunnel Protocol Support*

RFC 2868, *RADIUS Attributes for Tunnel Protocol Support*

RFC 2869, *RADIUS Extensions*

RFC 3162, *RADIUS and IPv6*

RFC 4818, *RADIUS Delegated-IPv6-Prefix Attribute*

RFC 5176, *Dynamic Authorization Extensions to RADIUS*

RFC 6911, *RADIUS attributes for IPv6 Access Networks*

RFC 6929, *Remote Authentication Dial-In User Service (RADIUS) Protocol Extensions*

Resource Reservation Protocol — Traffic Engineering (RSVP-TE)

draft-newton-mpls-te-dynamic-overbooking-00, *A Diffserv-TE Implementation Model to dynamically change booking factors during failure events*

RFC 2702, *Requirements for Traffic Engineering over MPLS*

RFC 2747, *RSVP Cryptographic Authentication*

RFC 2961, *RSVP Refresh Overhead Reduction Extensions*

RFC 3097, *RSVP Cryptographic Authentication -- Updated Message Type Value*

RFC 3209, *RSVP-TE: Extensions to RSVP for LSP Tunnels*

RFC 3473, *Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions* - IF_ID RSVP_HOP object with unnumbered interfaces and RSVP-TE graceful restart helper procedures

RFC 3477, *Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE)*

RFC 3564, *Requirements for Support of Differentiated Services-aware MPLS Traffic Engineering*

RFC 3906, *Calculating Interior Gateway Protocol (IGP) Routes Over Traffic Engineering Tunnels*

RFC 4090, *Fast Reroute Extensions to RSVP-TE for LSP Tunnels*

RFC 4124, *Protocol Extensions for Support of Diffserv-aware MPLS Traffic Engineering*

RFC 4125, *Maximum Allocation Bandwidth Constraints Model for Diffserv-aware MPLS Traffic Engineering*

RFC 4127, *Russian Dolls Bandwidth Constraints Model for Diffserv-aware MPLS Traffic Engineering*

RFC 4561, *Definition of a Record Route Object (RRO) Node-Id Sub-Object*

RFC 4875, *Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)*

RFC 4950, *ICMP Extensions for Multiprotocol Label Switching*

RFC 5151, *Inter-Domain MPLS and GMPLS Traffic Engineering -- Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Extensions*

RFC 5712, *MPLS Traffic Engineering Soft Preemption*

RFC 5817, *Graceful Shutdown in MPLS and Generalized MPLS Traffic Engineering Networks*

Routing Information Protocol (RIP)

RFC 1058, *Routing Information Protocol*

RFC 2080, *RIPng for IPv6*

RFC 2082, *RIP-2 MD5 Authentication*

RFC 2453, *RIP Version 2*

Segment Routing (SR)

draft-bashandy-rtgwg-segment-routing-uloop-06, *Loop avoidance using Segment Routing*

draft-ietf-idr-bgp-ls-segment-routing-ext-16, *BGP Link-State extensions for Segment Routing* draft-ietf-idr-bgp-ls-segment-routing-msd-09, *Signaling MSD (Maximum SID Depth) using Border Gateway Protocol Link-State*

draft-ietf-idr-segment-routing-te-policy-09, *Advertising Segment Routing Policies in BGP*

draft-ietf-isis-mpls-elc-10, *Signaling Entropy Label Capability and Entropy Readable Label Depth Using IS-IS* - advertising ELC

draft-ietf-lsr-flex-algo-08, *IGP Flexible Algorithm*

draft-ietf-ospf-mpls-elc-12, *Signaling Entropy Label Capability and Entropy Readable Label-stack Depth Using OSPF* - advertising ELC

draft-ietf-rtgwg-segment-routing-ti-lfa-01, *Topology Independent Fast Reroute using Segment Routing*

draft-ietf-spring-conflict-resolution-05, *Segment Routing MPLS Conflict Resolution*

draft-ietf-spring-segment-routing-policy-08, *Segment Routing Policy Architecture*

draft-ietf-teas-sr-rsvp-coexistence-rec-02, *Recommendations for RSVP-TE and Segment Routing LSP co-existence*

draft-voyer-pim-sr-p2mp-policy-02, *Segment Routing Point-to-Multipoint Policy*

draft-voyer-spring-sr-p2mp-policy-03, *SR Replication Policy for P2MP Service Delivery*

RFC 8287, *Label Switched Path (LSP) Ping/Traceroute for Segment Routing (SR) IGP-Prefix and IGP-Adjacency Segment Identifiers (SIDs) with MPLS Data Planes*

RFC 8476, *Signaling Maximum SID Depth (MSD) Using OSPF* - node MSD

RFC 8491, *Signaling Maximum SID Depth (MSD) Using IS-IS* - node MSD

RFC 8660, *Segment Routing with the MPLS Data Plane*

RFC 8661, *Segment Routing MPLS Interworking with LDP*

RFC 8663, *MPLS Segment Routing over IP* - BGP SR with SR-MPLS-over-UDP/IP

RFC 8665, *OSPF Extensions for Segment Routing*

RFC 8666, *OSPFv3 Extensions for Segment Routing*

RFC 8667, *IS-IS Extensions for Segment Routing*

RFC 8669, *Segment Routing Prefix Segment Identifier Extensions for BGP*

Simple Network Management Protocol (SNMP)

RFC 1157, *A Simple Network Management Protocol (SNMP)*

RFC 1215, *A Convention for Defining Traps for use with the SNMP*

RFC 1901, *Introduction to Community-based SNMPv2*

RFC 3410, *Introduction and Applicability Statements for Internet Standard Management Framework*

RFC 3411, *An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks*

RFC 3412, *Message Processing and Dispatching for the Simple Network Management Protocol (SNMP)*

RFC 3413, *Simple Network Management Protocol (SNMP) Applications*

RFC 3414, *User-based Security Model (USM) for version 3 of the Simple Network Management Protocol (SNMPv3)*

RFC 3415, *View-based Access Control Model (VACM) for the Simple Network Management Protocol (SNMP)*

RFC 3416, *Version 2 of the Protocol Operations for the Simple Network Management Protocol (SNMP)*

RFC 3417, *Transport Mappings for the Simple Network Management Protocol (SNMP)* - SNMP over UDP over IPv4

RFC 3584, *Coexistence between Version 1, Version 2, and Version 3 of the Internet-standard Network Management Framework*

RFC 3826, *The Advanced Encryption Standard (AES) Cipher Algorithm in the SNMP User-based Security Model*

Simple Network Management Protocol (SNMP) - Management Information Base (MIB)

draft-ietf-snmpv3-update-mib-05, *Management Information Base (MIB) for the Simple Network Management Protocol (SNMP)*

draft-ietf-isis-wg-mib-06, *Management Information Base for Intermediate System to Intermediate System (IS-IS)*

draft-ietf-mboned-msdp-mib-01, *Multicast Source Discovery protocol MIB*

draft-ietf-mpls-ldp-mib-07, *Definitions of Managed Objects for the Multiprotocol Label Switching, Label Distribution Protocol (LDP)*

draft-ietf-mpls-lsr-mib-06, *Multiprotocol Label Switching (MPLS) Label Switching Router (LSR) Management Information Base Using SMIv2*

draft-ietf-mpls-te-mib-04, *Multiprotocol Label Switching (MPLS) Traffic Engineering Management Information Base*

draft-ietf-ospf-mib-update-08, *OSPF Version 2 Management Information Base*

draft-ietf-vrrp-unified-mib-06, *Definitions of Managed Objects for the VRRP over IPv4 and IPv6* - IPv6

ianaaddressfamilynumbers-mib, *IANA-ADDRESS-FAMILY-NUMBERS-MIB*

ianagmplstc-mib, *IANA-GMPLS-TC-MIB*

ianaiftype-mib, *IANAifType-MIB*

ianaiprouteprotocol-mib, *IANA-RTPROTO-MIB*

IEEE8021-CFM-MIB, *IEEE P802.1ag(TM) CFM MIB*

IEEE8021-PAE-MIB, *IEEE 802.1X MIB*

IEEE8023-LAG-MIB, *IEEE 802.3ad MIB*

LLDP-MIB, *IEEE P802.1AB(TM) LLDP MIB*

RFC 1212, *Concise MIB Definitions*

RFC 1213, *Management Information Base for Network Management of TCP/IP-based Internets: MIB-II*

RFC 1724, *RIP Version 2 MIB Extension*

RFC 2021, *Remote Network Monitoring Management Information Base Version 2 using SMIv2*

RFC 2115, *Management Information Base for Frame Relay DTEs Using SMIv2*

RFC 2206, *RSVP Management Information Base using SMIv2*

RFC 2213, *Integrated Services Management Information Base using SMIv2*

RFC 2494, *Definitions of Managed Objects for the DS0 and DS0 Bundle Interface Type*

RFC 2514, *Definitions of Textual Conventions and OBJECT-IDENTITIES for ATM Management*

RFC 2515, *Definitions of Managed Objects for ATM Management*

RFC 2578, *Structure of Management Information Version 2 (SMIv2)*

RFC 2579, *Textual Conventions for SMIv2*

RFC 2580, *Conformance Statements for SMIv2*

RFC 2787, *Definitions of Managed Objects for the Virtual Router Redundancy Protocol*

RFC 2819, *Remote Network Monitoring Management Information Base*

RFC 2856, *Textual Conventions for Additional High Capacity Data Types*

RFC 2863, *The Interfaces Group MIB*

RFC 2864, *The Inverted Stack Table Extension to the Interfaces Group MIB*

RFC 2933, *Internet Group Management Protocol MIB*

RFC 3014, *Notification Log MIB*

RFC 3165, *Definitions of Managed Objects for the Delegation of Management Scripts*

RFC 3231, *Definitions of Managed Objects for Scheduling Management Operations*

RFC 3273, *Remote Network Monitoring Management Information Base for High Capacity Networks*

RFC 3419, *Textual Conventions for Transport Addresses*

RFC 3498, *Definitions of Managed Objects for Synchronous Optical Network (SONET) Linear Automatic Protection Switching (APS) Architectures*

RFC 3592, *Definitions of Managed Objects for the Synchronous Optical Network/ Synchronous Digital Hierarchy (SONET/SDH) Interface Type*

RFC 3593, *Textual Conventions for MIB Modules Using Performance History Based on 15 Minute Intervals*

RFC 3635, *Definitions of Managed Objects for the Ethernet-like Interface Types*

RFC 3637, *Definitions of Managed Objects for the Ethernet WAN Interface Sublayer*

RFC 3877, *Alarm Management Information Base (MIB)*

RFC 3895, *Definitions of Managed Objects for the DS1, E1, DS2, and E2 Interface Types*

RFC 3896, *Definitions of Managed Objects for the DS3/E3 Interface Type*

RFC 4001, *Textual Conventions for Internet Network Addresses*

RFC 4022, *Management Information Base for the Transmission Control Protocol (TCP)*

RFC 4113, *Management Information Base for the User Datagram Protocol (UDP)*

RFC 4220, *Traffic Engineering Link Management Information Base*

RFC 4273, *Definitions of Managed Objects for BGP-4*

RFC 4292, *IP Forwarding Table MIB*

RFC 4293, *Management Information Base for the Internet Protocol (IP)*

RFC 4631, *Link Management Protocol (LMP) Management Information Base (MIB)*

RFC 4878, *Definitions and Managed Objects for Operations, Administration, and Maintenance (OAM) Functions on Ethernet-Like Interfaces*

RFC 7420, *Path Computation Element Communication Protocol (PCEP) Management Information Base (MIB) Module*

SFLOW-MIB Version 1.3 (Draft 5), *sFlow MIB*

Timing

GR-1244-CORE Issue 3, *Clocks for the Synchronized Network: Common Generic Criteria*

GR-253-CORE Issue 3, *SONET Transport Systems: Common Generic Criteria*

IEEE 1588-2008, *IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems*

ITU-T G.781, *Synchronization layer functions*

ITU-T G.813, *Timing characteristics of SDH equipment slave clocks (SEC)*

ITU-T G.8261, *Timing and synchronization aspects in packet networks*

ITU-T G.8262, *Timing characteristics of synchronous Ethernet equipment slave clock (EEC)*

ITU-T G.8264, *Distribution of timing information through packet networks*

ITU-T G.8265.1, *Precision time protocol telecom profile for frequency synchronization*

ITU-T G.8275.1, *Precision time protocol telecom profile for phase/time synchronization with full timing support from the network*

RFC 3339, *Date and Time on the Internet: Timestamps*

RFC 5905, *Network Time Protocol Version 4: Protocol and Algorithms Specification*

Two-Way Active Measurement Protocol (TWAMP)

RFC 5357, *A Two-Way Active Measurement Protocol (TWAMP)* - server, unauthenticated mode

RFC 5938, *Individual Session Control Feature for the Two-Way Active Measurement Protocol (TWAMP)*

RFC 6038, *Two-Way Active Measurement Protocol (TWAMP) Reflect Octets and Symmetrical Size Features*

RFC 8545, *Well-Known Port Assignments for the One-Way Active Measurement Protocol (OWAMP) and the Two-Way Active Measurement Protocol (TWAMP)* - TWAMP

RFC 8762, *Simple Two-Way Active Measurement Protocol* - Unauthenticated

Virtual Private LAN Service (VPLS)

RFC 4761, *Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling*

RFC 4762, *Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling*

RFC 5501, *Requirements for Multicast Support in Virtual Private LAN Services*

RFC 6074, *Provisioning, Auto-Discovery, and Signaling in Layer 2 Virtual Private Networks (L2VPNs)*

RFC 7041, *Extensions to the Virtual Private LAN Service (VPLS) Provider Edge (PE) Model for Provider Backbone Bridging*

RFC 7117, *Multicast in Virtual Private LAN Service (VPLS)*

Voice and Video

DVB BlueBook A86, *Transport of MPEG-2 TS Based DVB Services over IP Based Networks*

ETSI TS 101 329-5 Annex E, *QoS Measurement for VoIP - Method for determining an Equipment Impairment Factor using Passive Monitoring*

ITU-T G.1020 Appendix I, *Performance Parameter Definitions for Quality of Speech and other Voiceband Applications Utilizing IP Networks - Mean Absolute Packet Delay Variation & Markov Models*

ITU-T G.107, *The E Model - A computational model for use in planning*

ITU-T P.564, *Conformance testing for voice over IP transmission quality assessment models*

RFC 3550, *RTP: A Transport Protocol for Real-Time Applications* - Appendix A.8

RFC 4585, *Extended RTP Profile for Real-time Transport Control Protocol (RTCP)-Based Feedback (RTP/AVPF)*

RFC 4588, *RTP Retransmission Payload Format*

Wireless Local Area Network (WLAN) Gateway

3GPP TS 23.402, *Architecture enhancements for non-3GPP accesses* - S2a roaming based on GPRS

Yet Another Next Generation (YANG)

RFC 6991, *Common YANG Data Types*

RFC 7950, *The YANG 1.1 Data Modeling Language*

RFC 7951, *JSON Encoding of Data Modeled with YANG*

Yet Another Next Generation (YANG) - OpenConfig Modules

openconfig-aaa.yang version 0.4.0, *OpenConfig AAA Module*

openconfig-aaa-radius.yang version 0.3.0, *OpenConfig AAA RADIUS Module*

openconfig-aaa-tacacs.yang version 0.3.0, *OpenConfig AAA TACACS+ Module*

openconfig-acl.yang version 1.0.0, *OpenConfig ACL Module*

openconfig-bfd.yang version 0.1.0, *OpenConfig BFD Module*

openconfig-bgp.yang version 3.0.1, *OpenConfig BGP Module*

openconfig-bgp-common.yang version 3.0.1, *OpenConfig BGP Common Module*

openconfig-bgp-common-multiprotocol.yang version 3.0.1, *OpenConfig BGP Common Multiprotocol Module*

openconfig-bgp-common-structure.yang version 3.0.1, *OpenConfig BGP Common Structure Module*

openconfig-bgp-global.yang version 3.0.1, *OpenConfig BGP Global Module*

openconfig-bgp-neighbor.yang version 3.0.1, *OpenConfig BGP Neighbor Module*

openconfig-bgp-peer-group.yang version 3.0.1, *OpenConfig BGP Peer Group Module*

openconfig-bgp-policy.yang version 4.0.1, *OpenConfig BGP Policy Module*

openconfig-if-aggregate.yang version 2.0.0, *OpenConfig Interfaces Aggregated Module*

openconfig-if-ethernet.yang version 2.0.0, *OpenConfig Interfaces Ethernet Module*

openconfig-if-ip.yang version 2.0.0, *OpenConfig Interfaces IP Module*

openconfig-if-ip-ext.yang version 2.0.0, *OpenConfig Interfaces IP Extensions Module*

openconfig-interfaces.yang version 2.0.0, *OpenConfig Interfaces Module*

openconfig-isis.yang version 0.3.0, *OpenConfig IS-IS Module*

openconfig-isis-policy.yang version 0.3.0, *OpenConfig IS-IS Policy Module* openconfig-isis-routing.yang version 0.3.0, *OpenConfig IS-IS Routing Module*

openconfig-lacp.yang version 1.1.0, *OpenConfig LACP Module*

openconfig-lldp.yang version 0.1.0, *OpenConfig LLDP Module*

openconfig-local-routing.yang version 1.0.1, *OpenConfig Local Routing Module*

openconfig-network-instance.yang version 0.8.0, *OpenConfig Network Instance Module*

openconfig-mpls.yang version 2.3.0, *OpenConfig MPLS Module*

openconfig-mpls-rsvp.yang version 2.3.0, *OpenConfig MPLS RSVP Module*

openconfig-mpls-te.yang version 2.3.0, *OpenConfig MPLS TE Module*

openconfig-packet-match.yang version 1.0.0, *OpenConfig Packet Match Module*

openconfig-platform.yang version 0.12.2, *OpenConfig Platform Module*

openconfig-platform-fan.yang version 0.1.1, *OpenConfig Platform Fan Module*

openconfig-platform-linecard.yang version 0.1.2, *OpenConfig Platform Linecard Module*

openconfig-relay-agent.yang version 0.1.0, *OpenConfig Relay Agent Module*

openconfig-routing-policy.yang version 3.0.0, *OpenConfig Routing Policy Module*

openconfig-system-logging.yang version 0.3.1, *OpenConfig System Logging Module*

openconfig-system-terminal.yang version 0.3.0, *OpenConfig System Terminal Module*

openconfig-telemetry.yang version 0.5.0, *OpenConfig Telemetry Module*

openconfig-vlan.yang version 2.0.0, *OpenConfig VLAN Module*

Customer Document and Product Support

Customer Documentation

Customer Documentation Welcome Page

Technical Support

Product Support Portal

Documentation Feedback

Customer Documentation Feedback