NOKIA

7705 Service Aggregation Router Gen 2
Release 25.3.R2

MPLS Guide

3HE 21575 AAAA TQZZA 01
Edition: 01
April 2025

Table of contents

List of tables

List of figures

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.

Unless otherwise indicated, the topics and commands described in this guide apply only to the 7705 SAR Gen 2 platforms listed in Platforms and terminology.

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

> **Note:** Unless otherwise indicated, CLI commands, contexts, and configuration examples in this guide apply for both the classic CLI and the MD-CLI.

The SR OS CLI trees and command descriptions can be found in the following guides:

- *7705 SAR Gen 2 Classic CLI Command Reference Guide*
- *7705 SAR Gen 2 Clear, Monitor, Show, Tools CLI Command Reference Guide* (for both the MD-CLI and classic CLI)
- *7705 SAR Gen 2 MD-CLI Command Reference Guide*

> **Note:** This guide generically covers Release 25.*x*.R*x* content and may contain some content that will be released in later maintenance loads. See the *SR OS R25.x.Rx Software Release Notes*, part number 3HE 21562 000*x* TQZZA, for information about features supported in each load of the Release 25.*x*.R*x* software. For a list of features and CLI commands that are present in SR OS but not supported on the 7705 SAR Gen 2 platforms, see "SR OS Features not Supported on SAR Gen 2" in the *SR OS R25.x.Rx Software Release Notes*.

1.2 Platforms and terminology

> **Note:**
> Unless explicitly noted otherwise, this guide uses the terminology defined in the following table to collectively designate the specified platforms.

Table 1: Platforms and terminology

Platform	Collective platform designation
7705 SAR-1	7705 SAR Gen 2

1.3 Conventions

This section describes the general conventions used in this guide.

1.3.1 Precautionary and information messages

The following information symbols are used in the documentation.

DANGER: Danger warns that the described activity or situation may result in serious personal injury or death. An electric shock hazard could exist. Before you begin work on this equipment, be aware of hazards involving electrical circuitry, be familiar with networking environments, and implement accident prevention procedures.

WARNING: Warning indicates that the described activity or situation may, or will, cause equipment damage, serious performance problems, or loss of data.

Caution: Caution indicates that the described activity or situation may reduce your component or system performance.

Note: Note provides additional operational information.

Tip: Tip provides suggestions for use or best practices.

1.3.2 Options or substeps in procedures and sequential workflows

Options in a procedure or a sequential workflow are indicated by a bulleted list. In the following example, at step 1, the user must perform the described action. At step 2, the user must perform one of the listed options to complete the step.

Example: Options in a procedure

1. User must perform this step.

2. This step offers three options. User must perform one option to complete this step.

    •   This is one option.

    •   This is another option.

    •   This is yet another option.

Substeps in a procedure or a sequential workflow are indicated by letters. In the following example, at step 1, the user must perform the described action. At step 2, the user must perform two substeps (a. and b.) to complete the step.

Example: Substeps in a procedure

1. User must perform this step.

2. User must perform all substeps to complete this action.

    a. This is one substep.

b. This is another substep.

2 MPLS and RSVP

This chapter provides information about how to configure 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 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 then turns them into labeled packets. Routers that support MPLS are known as Label Switching Routers (LSRs). To transmit a labeled packet on a particular data link, an LSR must support the encoding technique, which, when a label stack and a network layer packet are specified, produces a labeled packet.

In MPLS, packets can carry not only 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: Label placement shows the label placement in a packet. Table 2: Packet/label field description describes the label fields.

Figure 1: Label placement
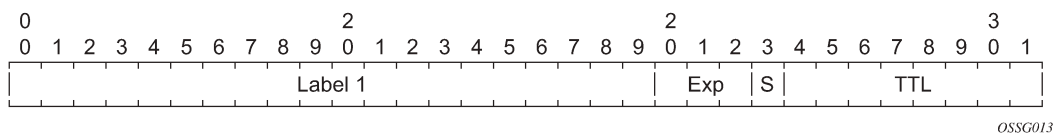
Table 2: Packet/label field description

Field	Description
Label	This 20-bit field carries the actual value (unstructured) of the label

Field	Description
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 the following figure.

Figure 2: Label packet placement

Layer 2 Header	Top Label	…	Bottom Label	Data Packet
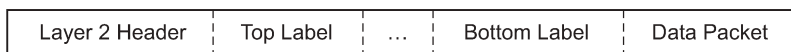
OSSG014

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

- 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 correctly 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 label stack of the packet.

2.1.1.1 Label values

A packet traveling along an LSP (see LSRs) is 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 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. The 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. 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. Because this label cannot typically occur at the bottom of the stack, it is not associated with a specific network layer protocol.

- 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 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, or later FP generation, profile B) are assigned dynamically by RSVP, LDP, and BGP control planes for both MPLS LSP and service labels.

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 IS-IS.

• Reserved Label Block for applications such as SR policy, MPLS forwarding policy, and the assignment of a static label to the SID of a IS-IS 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 range can be configured up to the maximum supported MPLS label value on the system. On the 7705 SAR Gen 2, the SRLB maximum limit is 524,287.

Example: Reserved label block configuration (MD-CLI)

```
[ex:/configure router "Base" mpls-labels]
A:admin@node-2# info
    reserved-label-block "test" {
        start-label 18432
        end-label 20000
    }
```

Example: Reserved label block configuration (classic CLI)

```
A:node-2>config>router>mpls-labels# info
----------------------------------------------
            reserved-label-block "test"
                start-label 18432 end-label 20000
            exit
----------------------------------------------
```

2.1.2 MPLS and hash label

The router supports the flow-aware transport (FAT) label (also known as the hash label), as specified in RFC 6391, on spoke SDPs bound to a VPLS EVPN service as well as on EVPN unicast destinations (in Epipe and VPLS services), if enabled by the **hash-label** command. Routers in the LSR role typically use label stacks for load-balancing traffic and, with ingress hashing that uses ingress flow information to generate the hash label (FAT label), LSRs can perform load balancing in a more effective manner. The use of labels also removes the need for an LSR to inspect the payload below the label stack to check for an IPv4 or IPv6 header, and determine how load balancing is applied.

The hash label is primarily applicable to Layer 2 services such as VLL and VPLS.

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 the **auto-bind-tunnel** command with the **resolution-filter** command set to any MPLS tunnel type. When enabled, the ingress datapath is modified such that the result of the hash on the payload packet header is communicated to the egress datapath for use as the value of the label field of the hash label. The egress datapath appends the hash label to the bottom of the stack (BoS) and sets the S-bit to 1. The TTL of the hash label is set to a value of 0. 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.

2.1.3 LSRs

LSRs perform different label switching functions based on their 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.

- An 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 up 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 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 the supported 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.

  The following are the 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, which means 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. These LSPs 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 that 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.

  When 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, which can cause some of the detours to fail. Otherwise, the LSP is not impacted.

Hop-limit parameters specify 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 MPLS Fast Reroute

The MPLS facility bypass method of MPLS Fast Reroute (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 iLER 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. Thus, the detour at the ingress does not protect the standby. 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, the LSP switches to standby after the expiration of the hold timer configured for the MPLS router in the following command.

    ```
    configure router mpls hold-timer
    ```

    If the **hold-timer** is disabled, a switchover to standby occurs immediately. On the successful global revert of the primary path, the LSP switches back to the primary path.

> **Note:** Admin groups are not taken into account when creating detours for LSPs.

2.1.5 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 or the FRR object in the path message indicates that facility backup is needed, 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. Increase or decrease the number of associations with the following command.

```
configure router mpls max-bypass-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.

See Configuring manual bypass tunnels for configuration information.

2.1.5.1 PLR bypass LSP selection rules

The following figure shows a sample bypass tunnel node arrangement.

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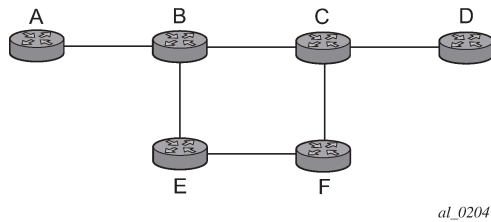
The PLR uses the following sequence of rules to select a bypass LSP among multiple manual and dynamic bypass LSPs at the establishment of the primary LSP path, or when searching for a bypass for a protected LSP that does not have an association with a bypass tunnel.

1. The MPLS task in the PLR node checks whether an existing manual bypass satisfies the constraints. If the path message for the primary LSP path indicates node protection is needed, 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 indicates link protection is needed, 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 satisfy the constraints, and dynamic bypass tunnels have not been disabled on PLR node, the MPLS task in the PLR checks whether any of the already established dynamic bypasses of the requested type satisfy the constraints.

4. If no existing dynamic bypasses of the requested type satisfy the constraints, 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 indicates that node protection is needed, and no manual bypass was found after step 1, or no dynamic bypass LSP was found after 3 attempts of performing step 3, the MPLS task repeats steps 1 to 3 looking for a suitable link-protect bypass LSP. If none is 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.

6. If the path message for the primary LSP path indicates that link protection is needed, and no manual bypass was found after step 1, and 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 does 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 LSPs 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 occurs on the protected LSP. Also, if no manual bypass exists that satisfies the constraints of the protected LSP, the LSP remains without protection.

If the user configures a bypass tunnel on node B and dynamic bypass tunnels have been disabled, LSPs that have been previously signaled and were not associated with any manual bypass tunnel (for example, if 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 that have been previously signaled over dynamic bypass tunnels are not automatically switched into the manual bypass tunnel, even if the manual bypass is a more optimized path. The user must 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) clears the "protection available" flag in the RRO IPv4 sub-object in the next RESV refresh message for each affected LSP. It then attempts to associate each of these LSPs with one of the manual bypass tunnels that are still up.

If it finds one, it makes the association and again set the "protection available" flag in the next RESV refresh message for each of these LSPs. If it cannot find one, it continues to check 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 that 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 automatically signals 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 only signals 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 does not switch the protected LSPs from the dynamic bypass tunnel into the manual bypass tunnel.

2.1.5.2 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 enables PLRs to be aware of the missing node protection and allows them to probe regularly for a node-bypass. The following figure shows an example of FRR node protection.

Figure 4: FRR node-protection example

al_0205
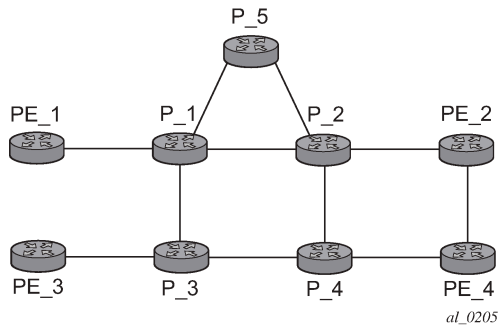
The following describes an LSP scenario where:

- 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 protects 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 because there is no 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. When 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.

2.1.6 MPLS LSP history

The router can store the 100 most recent events for each configured point-to-point RSVP or SR-TE LSP. This is independent of any other system log functionality.

Use the commands in the following context to enable the ability to store LSP state history.

```
configure router mpls lsp-history
```

When enabled, the router stores up to 100 of the most recent significant events for each LSP as a sliding window of events. When new events occur on an LSP and the record of 100 is fully consumed, new events are added and the oldest events are removed. The recording of LSP events is paused when the context is administratively disabled. The stored history for the LSPs is deleted when the context is deleted, and the memory allocated to store these events becomes available.

The history for a named RSVP or SR-TE LSP can be displayed for all LSPs, or for a single named LSP. Use the following command to display a specific RSVP LSP or all LSPs.

```
tools dump router mpls lsp-history [lsp-name]
```

If the LSP name is not specified, the output displays the LSP history for all RSVP and SR-TE LSPs, in sequence.

The history for a single named RSVP LSP or all LSPs can be cleared. Use the following command to clear the history for a specific named RSVP LSP or all LSPs.

```
clear router mpls lsp-history [lsp-name]
```

2.1.7 LSP failure codes

The table below lists the MPLS LSP path failure codes and their meanings. The failure codes are indicated in the FailureCode output field in the TiMetra MPLS MIB and for specific CLI commands. Use the following commands to display the FailureCode output field.

```
show router mpls lsp path detail
tools dump router mpls lsp-history
```

Table 3: LSP failure codes

LSP failure code (value)	Meaning
noError (0)	Indicates no errors for this LSP.
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 because of an unsupported traffic parameter, for example, a bad FLOWSPEC, TSPEC, or ADSPEC value.
routingError (4)	Indicates 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

LSP failure code (value)	Meaning
	"Unsupported L3PID." 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 defined in RFC 2205, *Resource Re SerVation Protocol (RSVP) - Version 1 Functional Specification*.
unknownCType (12)	A downstream LSR rejected an RSVP message because of an Unknown object C-type – Error code 14 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.
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 or standby path is not an 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.

LSP failure code (value)	Meaning
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.
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 because of maintenance. The LSR that is shut down 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 five sub-objects then it is rejected.

LSP failure code (value)	Meaning
rsvpNotSupported (41)	Maps to SubErrorCode 8 for ErrorCode 24 (Routing error) from RFC 3209. An LSR sends 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 each other, for example, the same group is both included and excluded.
nodeInIgpOverload (43)	An LSR along the path of the LSP has advertised the IS-IS 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 maximum SR labels.
adminDown (53)	A related MPLS path is disabled.
sidHopsInRsvpLsp (54)	SID hops in the path for RSVP-TE LSP.
ipv6HopsInRsvpLsp (55)	IPv6 hops in the path for RSVP-TE LSP.
ipv4HopsInIpv6Lsp (56)	IPv4 hops in the path for SR-TE LSP with IPv6 'to' address.
ipv6HopsInIpv4Lsp(57)	IPv6 hops in the path for SR-TE LSP with IPv4 'to' address.
sidHopsInIpv6Lsp (58)	SID hops in the path for SR-TE LSP with IPv4 'to' address.
srlgPathWithSidHops (59)	LSP path is SRLG enabled but has SID hops in the path.
mplsV4Down (60)	MPLS IPv4 operational state is down.
mplsV6Down (61)	MPLS IPv6 operational state is down.
lspAdminDown (62)	LSP is admin down.
pathAdminDown (63)	Path or LSP path is admin down.
templateAdminDown (64)	LSP template is admin down.
pceAssocConflict (65)	PCE association is conflicting.
pathRetried (66)	LSP path was brought down and retried.

LSP failure code (value)	Meaning
clearCommand (67)	LSP path was brought down because of a **clear** command.[1]
nonActiveSecondary (68)	The secondary path is down because the LSP has an alternate active path.
autoBandwidthAdjustment (69)	For an auto-bandwidth LSP, operational bandwidth for a non-active LSP path does not match the bandwidth configured for that path and the bandwidth was not adjusted using MBB. The path is brought down and retried to adjust its bandwidth back to configured bandwidth.
pathDegraded (71)	Non-active path was brought down because it was in a degraded state, for example, FRR active or soft-preempted, and the MBB could not be used to resignal the path.
lspSelfPingTimeout (72)	LSP self-ping timed out.
rsvpError (73)	RSVP signaling errors, such as, ResvTear received or egress neighbor down.
p2mpInstanceAdminDown (74)	P2MP instance is admin down.
bfdDown (75)	BFD session is down on the SR-TE LSP path.[1]
rsvpNeighborDown (76)	RSVP neighbor is down for the RSVP-TE LSP.
resvTimeout (77)	RESV state timed out for RSVP-TE LSP.
resvTear (78)	ResvTear message received for RSVP-TE LSP.
frrPathDown (79)	Fast-reroute enabled RSVP-TE LSP was actively using backup path and backup path went down.
lspInitRetryTimeout (80)	Initial retry timer expired for RSVP-TE LSP.
iomProgrammingFailure (81)	IOM programming failed for RSVP-TE LSP.
delayMetricLimitExceeded (82)	CSPF could not find a path for which the end-to-end delay metric was less than or equal to the delay metric limit configured for that path.

2.1.8 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.

---

[1] This code is not available in the **show** command output.

2.1.8.1 Interface statistics

By default, the system continuously collects statistics (packet and octet counts) of MPLS traffic on ingress and egress of MPLS interfaces. Use the following command to view these statistics.

```
show router mpls interface statistics
```

The implicit null on ingress is not regarded as labeled traffic and octet counts include Layer 2 headers and trailers.

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. Use the following command to access auxiliary statistics and display the types of labeled traffic that should be counted.

```
configure router mpls aux-stats
```

The **sr** command option 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 Layer 2 headers and trailers.

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

2.1.8.2 Traffic statistics for stacked tunnels

The nature of MPLS allows for LSPs, owned by a specific 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 toward 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.

Use the following command to control statistics collection on one or two labels.

```
configure system ip mpls label-stack-statistics-count
```

This command does not enable statistics collection. It only performs controls on a specific number of labels, and 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).

2.1.8.3 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 indexes (on ingress and on egress independently). Statistics with FC and QoS profile consume 16 indexes.

2.1.8.4 RSVP-TE traffic statistics

See RSVP-TE LSP statistics for information about RSVP-TE traffic statistics.

2.1.9 Monitoring MPLS resource consumption

SR OS supports the display of MPLS system resources on an egress-operation basis for NHLFEs, labels, and LTNs. Users can access resource consumption information directly using a **tools** or **state** command, or remotely through SNMP and NETCONF.

Use the following command to display all MPLS resource usage information.

```
tools dump mpls-resources
```

Use commands in the following MD-CLI context to display YANG state information for MPLS resources.

```
state system mpls-resource-usage
```

Output example: Global MPLS resource usage

```
Global MPLS Resource Usage
------------------------+------------+------------+------------+
                        |   Total    | Allocated  |    Free    |
------------------------+------------+------------+------------+
 mpls   NHLFE           |   262125   |       1    |   262124   |
     RSVP               |            |       1    |            |
     LDP                |            |       0    |            |
     BGP                |            |       0    |            |
     MPLS-TP            |            |       0    |            |
     SR                 |            |       0    |            |
     BIER               |            |       0    |            |
     TREE-SID           |            |       0    |            |
 mpls labels            |   524256   |       0    |   524256   |
     RSVP               |            |       0    |            |
     LDP                |            |       0    |            |
     BGP                |            |       0    |            |
     MPLS-TP            |            |       0    |            |
     STATIC-SVC         |            |       0    |            |
     SR                 |            |       0    |            |
     BIER               |            |       0    |            |
     RESERVED-BLK       |            |       0    |            |
 mpls   LTN (FTN)       |   131071   |       0    |   131071   |
```

```
        RSVP                  |              |        0  |               |
        LDP                   |              |        0  |               |
        BGP                   |              |        0  |               |
        MPLS-TP               |              |        0  |               |
        SR                    |              |        0  |               |
        BIER                  |              |        0  |               |
        TREE-SID              |              |        0  |               |
```

Example: YANG state MPLS resource usage

```
[/state system mpls-resource-usage]
A:admin@node-2# info
    nhlfe {
        total 262125
        allocated 1
        free 262124
        by-owner rsvp {
            allocated 1
        }
        by-owner ldp {
            allocated 0
        }
        by-owner bgp {
            allocated 0
        }
        by-owner mpls-tp {
            allocated 0
        }
        by-owner static-service {
            allocated 0
        }
        by-owner sr-mpls {
            allocated 0
        }
        by-owner bier {
            allocated 0
        }
        by-owner tree-sid {
            allocated 0
        }
        by-owner reserved-blk {
            allocated 0
        }
    }
    ltn {
        total 131071
        allocated 0
        free 131071
        by-owner rsvp {
            allocated 0
        }
        by-owner ldp {
            allocated 0
        }
        by-owner bgp {
            allocated 0
        }
        by-owner mpls-tp {
            allocated 0
        }
        by-owner static-service {
            allocated 0
        }
```

```
            by-owner sr-mpls {
                allocated 0
            }
            by-owner bier {
                allocated 0
            }
            by-owner tree-sid {
                allocated 0
            }
            by-owner reserved-blk {
                allocated 0
            }
        }
        label {
            total 524256
            allocated 0
            free 524256
            by-owner rsvp {
                allocated 0
            }
            by-owner ldp {
                allocated 0
            }
            by-owner bgp {
                allocated 0
            }
            by-owner mpls-tp {
                allocated 0
            }
            by-owner static-service {
                allocated 0
            }
            by-owner sr-mpls {
                allocated 0
            }
            by-owner bier {
                allocated 0
            }
            by-owner tree-sid {
                allocated 0
            }
            by-owner reserved-blk {
                allocated 0
            }
        }
```

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 application data streams or flows. RSVP is also used by routers to deliver quality of service (QoS) requests to all nodes along the paths of the flows, and to establish and maintain the state to provide the requested service. RSVP requests generally result in resources reserved in each node along the datapath. 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 (unidirectional) flows. 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. The following figure shows the process to establish an LSP.

Figure 5: Establishing LSPs
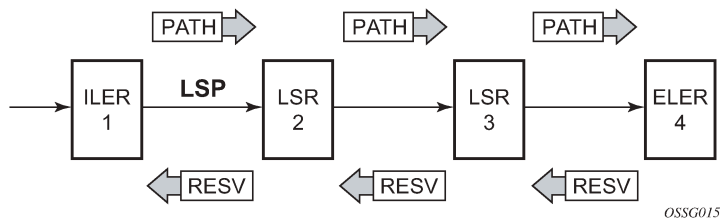
Establishing an LSP involves the following.

*   The sender (the iLER), sends PATH messages toward the receiver (the egress LER (eLER)) to indicate the FEC for which label bindings are needed. 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.

The following figure shows an example of an LSP path set up using RSVP.

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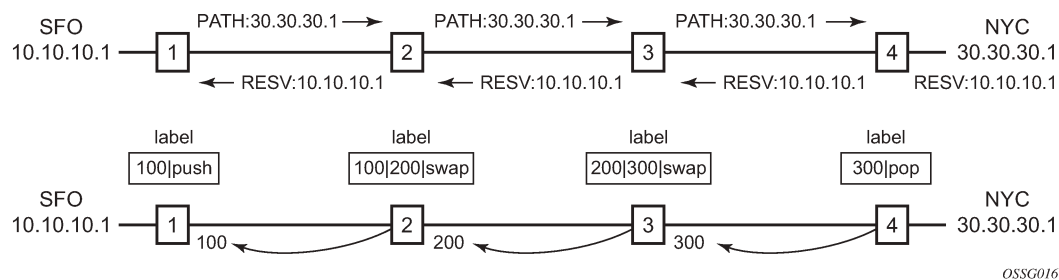
The ingress label edge router (iLER 1) transmits an RSVP path message (path: 30.30.30.1) downstream to the egress label edge router (ELER 4). The path message contains a label request object that requests intermediate LSRs and the eLER to provide a label binding for this path.

In addition to the label request object, an RSVP PATH message can also contain the following 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 traverses.

- **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 toward 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. After 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 is extended for MPLS to support automatic signaling of LSPs. To enhance the scalability, latency, and reliability of RSVP signaling, several extensions are 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 RSVP state information. In standard RSVP, neighbor monitoring occurs as part of RSVP 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 the following command is enabled on an RSVP interface, authentication of RSVP messages operates in both directions of the interface.

```
configure router rsvp interface authentication-key
```

A router maintains a security association using one authentication key for each interface to an RSVP neighbor.

An RSVP neighbor transmits an authenticating digest of the RSVP message that is 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. An RSVP neighbor uses the key together with the authentication algorithm to process received RSVP messages. The RSVP MD5 authentication complies to the procedures for RSVP message generation in RFC 2747, *RSVP Cryptographic Authentication*.

When a Point of Local Repair (PLR) activates a bypass LSP toward a Merge Point (MP), by default, the INTEGRITY object corresponding to the bypass LSP interface is not added to a transmitted RSVP message except for packets of routed RSVP messages (Resv, Srefresh, and ACK) and only when the packet is intended for a bypass LSP endpoint (PLR or MP) that is a directly connected neighbor.

When the following command is enabled, the INTEGRITY object of the interface corresponding to the bypass LSP is added to a transmitted RSVP message whether the bypass LSP endpoint (PLR or MP) is a directly connected RSVP neighbor.

```
configure router rsvp authentication-over-bypass
```

The INTEGRITY object is included with the following RSVP messages: Path, PathTear, PathErr, Resv, ResvTear, ResvErr, Srefresh, and ACK.

In all cases, an RSVP message received from a PLR or an MP (sender address in the SenderTemplate or FilterSpec is different from an Extended Tunnel ID in a Session Object), and which includes the INTEGRITY object, is authenticated against the bypass LSP interface. An RSVP message received from a PLR or MP without the INTEGRITY object is also accepted.

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 under the following context. The configured keychain must include at least on valid key entry, using a valid authentication algorithm for the RSVP protocol.

- **MD-CLI**

```
configure system security keychains
```

- **classic CLI**

```
configure system security keychain
```

2. Use the following command to associate the configured authentication keychain with RSVP at the interface level of the CLI.

- **MD-CLI**

```
configure router rsvp interface authentication-keychain
```

- **classic CLI**

```
configure router rsvp interface auth-keychain
```

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:

- cleartext 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 are not 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, when 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 several supported parameters. The style information is part of the LSP configuration. The SR OS supports the following reservation styles:

- **Fixed Filter (FF)**

The 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 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 the 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. If a PLR node receives a path message with fast-reroute requested with facility method and the FF reservation style, it rejects the reservation. The one-to-one detour method supports both FF and SE styles.

2.2.2.1 RSVP message pacing

When signaling message flooding occurs 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 because 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 an 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

Use the following command to enable RSVP Graceful Restart helper:

- **MD-CLI**

```
configure router rsvp interface graceful-restart-helper-mode
```

- **classic CLI**

```
configure router rsvp interface gr-helper
```

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 continues 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 required Restart Time and Recovery Time.

The Graceful Restart helper consists of a couple of phases. When 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.

When 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 activates 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 after the Recovery Phase time is over is considered to have expired and is deleted by the helper node causing the correct 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** command options under the following context:

- **MD-CLI**

```
configure router rsvp graceful-restart
```

- **classic CLI**

```
configure router rsvp gr-helper-time
```

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 doubles 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 because of 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. After 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 the number of refreshes of the pending PATH state that 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 continues 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 decreases the probability of overwhelming a congested LSR with new state while the previous states installed by the same LSP are lingering and is only 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 further decreases 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 options: outstanding CSPF requests, Resv timeouts and Path timeouts.

2.2.6 RSVP-TE LSP statistics

SR OS provides the following statistics:

- 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 RSVP LSPs, including template-based (mesh or one-hop, see Automatic creation of RSVP-TE LSP mesh), and for MPLS-TP LSPs at the egress datapath of an ingress LER and the ingress datapath of an egress LER. No LSR statistics are provided.

2.2.6.1 Rate statistics

SR OS also provides traffic rate statistics. For RSVP-TE LSPs, including template-based LSPs and MPLS-TP LSPs, perform one of the following options to enable that capability:

- configure an accounting policy that uses the following command with the **combined-mpls-lsp-egress** record name

```
configure log accounting-policy record
```

- assign that accounting policy to a specific LSP (or template)
- enable stats collection

The frequency at which the rate is determined is defined using the **collection-interval** command in the accounting policy. The minimum interval is 5 minutes.

Rate statistics are provided in packets per second and Mb/s. Rate statistics are provided as an aggregate across all paths of the LSP, which have a statistical index assigned, and for all forwarding classes in or out-of-profile.

Rate statistics are only available on the egress of the ingress LER. At least two samples are needed to determine a rate.

2.2.7 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 the 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 counts once as long as at least one LSP branch forwarded it.

- At BUD LSR and egress LER, per ILM statistics are provided. These counters include all packets received on the ILM, whether they match a Layer 2/Layer 3 MFIB record or not. ILM stats work the same way as for a P2P LSP. In other words, they 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 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 indexes the P2P LSP share on the node. Each P2P/P2MP RSVP LSP or LDP FEC consumes one statistics index for egress stats and one stat index for ingress statistics.

- The user can retrieve the above counters in four different ways:
  - In the CLI display of the output of the show command applied to a specific instance, or a specific template instance, of an RSVP P2MP.
  - In the CLI display of the output of the monitor command applied to a specific instance, or a specific template instance, of an 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 specific 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 Layer 2 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).

2.2.7.1 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.

Example: Egress statistics configuration (MD-CLI)

```
[ex:/configure router "Base" mpls]
A:admin@node-2# info
...
    lsp "test7" {
        type p2mp-rsvp
        egress-statistics {
            admin-state enable
            collect-stats true
            accounting-policy 99
        }
    }
```

Example: Egress statistics configuration (classic CLI)

```
A:node-2>config>router>mpls# info
----------------------------------------------
...
            lsp "test7" p2mp-lsp
                shutdown
                egress-statistics
                    collect-stats
                    accounting-policy 99
                    no shutdown
                exit
            exit
```

If there are no statistic indexes available when the user administratively enables 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.

Example: P2MP LSP template configuration (MD-CLI)

```
[ex:/configure router "Base" mpls]
A:admin@node-2# info
...
    lsp-template "test8" {
        type p2mp-rsvp
        egress-statistics {
            collect-stats true
            accounting-policy 99
        }
    }
```

Example: P2MP LSP template configuration (classic CLI)

```
A:node-2>config>router>mpls# info
---------------------------------------------
...
            lsp-template "test8" p2mp
                shutdown
                egress-statistics
                    collect-stats
                    accounting-policy 99
                exit
            exit
```

If there are no statistic indexes 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 statistics in the show command output of the P2MP LSP S2L path.

2.2.7.2 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 the *lsp-name* component is encoded as follows.

1. P2MP LSP via user configuration for Layer 3 multicast in global routing instance: "LspNameFromConfig"

2. P2MP LSP as I-PMSI or S-PMSI in Layer 3 mVPN: templateName-SvcId-mTTmIndex

3. P2MP LSP as I-PMSI in VPLS/B-VPLS: templateName-SvcId-mTTmIndex

The ingress statistics 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.

Example: RSVP P2MP LSP ingress statistics configuration (MD-CLI)

```
[ex:/configure router "Base" mpls ingress-statistics]
A:admin@node-2# info
```

```
    lsp sender 192.0.0.2 lsp-name "test9" {
        admin-state enable
        collect-stats true
        accounting-policy 88
    }
    p2mp-template-lsp sender 192.0.0.2 rsvp-session-name "test9" {
        admin-state enable
        collect-stats true
        accounting-policy 88
        max-stats 1000
    }
```

Example: RSVP P2MP LSP ingress statistics configuration (classic CLI)

```
A:node-2>config>router>mpls>ingr-stats# info
-----------------------------------------------
            lsp "test9" sender 192.0.0.2
                collect-stats
                accounting-policy 88
                no shutdown
            exit
            p2mp-template-lsp rsvp-session-name "test9" sender 192.0.0.2
                collect-stats
                accounting-policy 88
                max-stats 1000
                no shutdown
            exit
-----------------------------------------------
```

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 the **max-stats** command to limit the maximum number of stat indexes which can be assigned to this context. If the context matches more than this value, the additional request for stat indexes from this context is rejected.

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

- In the classic CLI, allocated **max-stats** can be increased but not decreased unless the entire ingress statistics context matching a template name is deleted.

- In the classic CLI, to delete ingress statistics context matching a template name, a shutdown is required.

- In the classic CLI, an accounting policy cannot be configured or de-configured until the ingress statistics context matching a template name is disabled.

- After deleting an accounting policy from an ingress statistics context matching a template name, the policy is not removed from the log until the ingress statistics context is enabled.

If there are no statistic indexes 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 indexes to the LSP names that match the context is not deterministic. The latter is because 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 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.

Use the following command to configure the 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.

```
configure router rsvp implicit-null-label
```

In the classic CLI, the user must administratively disable RSVP before being able to change the implicit null configuration.

Use the following commands to override the RSVP level configuration for a specific RSVP interface with the following command:

- **MD-CLI**

```
configure router rsvp interface implicit-null-label [true|false]
```

- **classic CLI**

```
configure router rsvp interface 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 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 uses 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. In the classic CLI, the user must administratively disable the RSVP interface before being able to change the implicit null configuration option.

> **Note:** In the classic CLI, the RSVP interface must be disabled 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. A user can push or swap an implicit null label on the MPLS packet using the **implicit-null-label** option and configuring next hop in the following contexts:

```
configure router mpls static-lsp push
configure router mpls interface label-map swap
```

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 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.

Use the following command to configure a borrowed IP address for an unnumbered interface. The default value is set to the system interface address:

- **MD-CLI**

```
configure router interface ipv4 unnumbered ip-address
```

- **classic CLI**

```
configure router interface unnumbered
```

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 specific 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 does not come up at the ingress LER because 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 remains in the 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 or 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 following command allows the user to manually enter the SRLG membership of any link in the network in a local database at the ingress LER

```
configure router mpls user-srlg-db
```

The user cannot enter an unnumbered interface into this database; and therefore, all unnumbered interfaces are considered as having no SRLG membership if the user enabled **user-srlg-db**.

This feature also extends the support of LSP ping, P2MP LSP ping, LSP trace, and P2MP LSP trace 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 does not come up.

In addition, the PLR node includes 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 creates 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

- C-Type is IF_ID RSVP_HOP and did not change but IF_ID TLV is different

- Change in IPv4 Next or Previous Hop Address in RSVP_HOP object regardless of the C-Type value.

These procedures at the PLR and MP nodes are followed in both the link-protect and the node-protect FRR. If the MP node is running a pre-Release 11.0 implementation, it rejects the new IF_ID C-Type and drops the PATH over bypass. This results in the protected LSP state expiring at the MP node, which tears 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 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.

2.3.1 LSP path computation with CSPF first algorithm

The following commands configure the path computation method for a RSVP-TE LSP.

```
configure router mpls lsp path-computation-method local-cspf
```

Users can select the computation method for the RSVP-TE LSP and set it to local Constrained Shortest Path First (CSPF). The default value sets the computation method to the local CSPF and is described in this section.

The CSPF algorithm implements the following steps:

1. The CSPF algorithm prunes the TE database with the links which do not satisfy the constraints of the RSVP-TE LSP path: bandwidth, SRLG, and admin-group constraints.

2. The CSPF algorithm computes all possible equal least-cost shortest paths based on cumulating the IGP link metric (default) or the TE link metric (if enabled for the LSP).

3. The CSPF algorithm selects the subset of paths which meet the maximum hop count constraint (**hop-limit** option) from among the equal least-cost shortest candidate paths.

4. The CSPF algorithm applies the random rule (default) or least-fill rule (if enabled for the LSP) to select a path from among the remaining equal least-cost candidate paths.

If a user configures strict or loose hops in the path definition for the LSP, CSPF applies the above procedures for each segment of the end-to-end path.

2.3.2 TE metric (IS-IS and OSPF)

When the use of the TE metric is selected for an LSP, by using the following command, the shortest path computation after the TE constraints are applied selects an LSP path based on the TE metric instead of the IGP metric.

- **MD-CLI**

```
configure router mpls lsp metric-type te
```

- **classic CLI**

```
configure router mpls lsp metric te
```

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 traffic engineering extensions of both IGP protocols.

A typical application of the TE metric is to allow CSPF to represent a dual TE topology for 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 and jitter figure, of the link. In this case, the shortest path satisfying the constraints of the LSP path effectively represents 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 prunes all links in the network topology that do not meet the constraints specified for the LSP path. The shortest path among all the SPF paths is selected based on the TE metric instead of the IGP metric, which is used by default.

2.3.3 LSP path reoptimization

MPLS performs a periodic reoptimization of the active path of each RSVP-TE LSP at the expiry of the global resignal timer if enabled using the following command.

```
configure router mpls resignal-timer
```

MPLS provides the current active path of the RSVP-TE LSP and the TE-DB updates the total IGP or TE metric of the path, checking the validity of the hops according to the 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 from the updated metric of the current path, or if one or more hops of the current path are invalid. Otherwise, the current path is considered one of the most optimal ECMP paths and is retained.

A new RSVP-TE LSP path is always instantiated using the Make-Before-Break (MBB) procedure. That is the new path is first signaled in the network and programmed into the datapath of the LER and LSR nodes before traffic is switched to it. When traffic is switched, the older path is torn down via the RSVP-TE control plane and its datapath resources are freed up at LER and LSR nodes.

The periodic reoptimization is referred to as the timer-based resignal MBB. The user can also perform a manual reoptimization of the active path of a specific RSVP-TE LSP or of the active paths of all RSVP-TE LSPs by using the following commands.

```
tools perform router mpls resignal lsp path
tools perform router mpls resignal delay
```

This is referred to as manual resignal MBB. The difference with the timer based resignal MBB is that the new computed manual resignal path is always signaled and programmed in the datapath regardless of its metric value.

SR OS implements the following additional reoptimization MBB types which are described later in this document. When multiple MBBs are scheduled at the same time, the highest priority MBB wins. CSPF still

uses the latest information from the TE-DB and any additional information associated with the lower priority MBBs to compute the new path. SR OS supports the following types of MBB that are ordered from highest to lowest priority.

1. delayed retry MBB

2. preemption MBB

3. global revertive MBB

4. configuration change MBB

5. TE Graceful Shutdown MBB

6. auto-bandwidth MBB

7. timer-based resignal MBB, manual resignal MBB, and ad hoc reoptimization MBB on receipt of IGP link event — these have the same priority. If one MBB is in progress, the other MBB requests are ignored.

2.3.4 Ad hoc RSVP-TE LSP reoptimization on receipt of IGP link events

The following command enables the ad hoc reoptimization of the active CSPF path of all RSVP-TE LSPs at the receipt of an IGP link event.

```
configure router mpls resignal-on-igp-event
```

The following link events are supported:

- link down

- link up

- IGP or TE metric change

- SRLG change

- admin group change

The ad hoc reoptimization follows the same behavior as in the timer-based resignal MBB feature. MPLS reevaluates the active paths of all RSVP-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 computing a new CSPF path. MPLS signals and programs the new path only if its total metric is different from the updated metric of the current path, or if one or more hops of the current path are invalid. Otherwise, the current path is considered to be the most optimal and retained.

This feature does not require enabling of the timer-based resignal command. If enabled, using the following command, it ends the resignal timer and performs the ad hoc reoptimization.

```
configure router mpls resignal-timer
```

2.3.5 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

2.3.5.1 Actions 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 command.

```
configure router mpls lsp fast-reroute propagate-admin-group
```

When this command is enabled at the ingress LER, the administrative 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 administrative 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 administrative group constraints in the selection of the FRR backup LSP for protecting the LSP primary path.

The ingress LER inserts the FAST_REROUTE object by default in a primary LSP path message. If the user disables the object using the following command, the administrative group constraints are not propagated:

- **MD-CLI**

```
configure router mpls frr-object false
```

- **classic CLI**

```
configure 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 following command.

```
configure 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.

2.3.5.2 Actions 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.

```
configure 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 reads 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 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 4: Bypass LSP admin-group constraint behavior.

Table 4: 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
	• **MD-CLI**
	
	`lsp propagate-admin-group false`
	
	`lsp fast-reroute propagate-admin-group true`
	
	• **classic CLI**
	
	`lsp no propagate-admin-group`
	
	`lsp fast-reroute propagate-admin-group`
		Admin color constraints not sent	Admin color constraints sent	Yes

2	frr-object
	• **MD-CLI**
	
	`lsp propagate-admin-group true`
	
	`lsp fast-reroute propagate-admin-group true`
	
	• **classic CLI**
	
	`lsp propagate-admin-group`
	
	`lsp fast-reroute propagate-admin-group`
		Admin color constraints sent	Admin color constraints sent	Yes

	Ingress LER configuration	Session attribute	FRR object	Bypass LSP at PLR (LER/LSF) follows admin-group constraints

3	frr-object
	
	• **MD-CLI**
	
	`lsp propagate-admin-group true`
	
	`lsp fast-reroute propagate-admin-group false`
	
	• **classic CLI**
	
	`lsp propagate-admin-group`
	
	`lsp fast-reroute no propagate-admin-group`
		Admin color constraints sent	Admin color constraints not sent	No

4	No frr-object
	
	• **MD-CLI**
	
	`lsp propagate-admin-group true`
	
	`lsp fast-reroute propagate-admin-group true`
	
	• **classic CLI**
	
	`lsp propagate-admin-group`
	
	`lsp fast-reroute propagate-admin-group`
		Admin color constraints sent	Not present	Yes

5	No frr-object
	
	• **MD-CLI**
	
	`lsp propagate-admin-group false`
	
	`lsp fast-reroute propagate-admin group true`
		Admin color constraints not sent	Not present	No

	Ingress LER configuration	Session attribute	FRR object	Bypass LSP at PLR (LER/LSF) follows admin-group constraints

	• **classic CLI**
	
	`lsp no propagate-admin-group`
	
	`lsp fast-reroute propagate-admin group`			

6	No frr-object
	• **MD-CLI**
	
	`lsp propagate-admin-group true`
	
	`lsp fast-reroute propagate-admin group false`
	
	• **classic CLI**
	
	`lsp propagate-admin-group`
	
	`lsp fast-reroute 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 or the other constraints, the PLR node requests 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, there is no effect on existing bypass associations. The change only applies to new attempts to find a valid bypass.

2.3.6 Manual and timer resignal of RSVP-TE bypass LSP

The following 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.

```
configure router mpls bypass-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 following 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.

```
configure router mpls srlg-frr
```

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 following command performs a manual re-optimization of a specific dynamic or manual bypass LSP, or of all dynamic bypass LSPs.

```
tools perform router mpls resignal-bypass
```

The name of a manual bypass LSP is configured by the user. The name of a dynamic bypass LSP is displayed in the output for the following command.

```
show router mpls bypass-tunnel dynamic detail
```

The **delay** command 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** command option. 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** command 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.

2.3.6.1 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 using the following command 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.

    ```
    configure router mpls srlg-frr strict
    ```

    b. Otherwise, MPLS performs additional checks using the PLR of the initial PSB to determine if the new path is more optimal. Table 5: Determination of bypass LSP path optimality summarizes the possible cases of bypass path optimality determination.

Table 5: Determination of bypass LSP path optimality

PLR SRLG constraint check[2]		SRLG FRR configuration (strict/loose)	Path cumulative cost comparison[2]	Path cumulative SRLG weight comparison[2]	More optimal path
Current path	New path				
Disjoint	Disjoint	—	New Cost < Current Cost	—	New

---

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

PLR SRLG constraint check[2]		SRLG FRR configuration (strict/loose)	Path cumulative cost comparison[2]	Path cumulative SRLG weight comparison[2]	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

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:

   a. MPLS signals and programs the new path.

   b. MPLS moves to the new bypass path the PSB associations of all PLRs which evaluation against Table 5: Determination of bypass LSP path optimality results in the new bypass path being more optimal.

   c. Among the remaining PLRs, MPLS does one of the following:

   • If the updated SRLG information of the current bypass path changed and SRLG FRR loose disjointness is configured using the following command option, MPLS keeps this PLR PSB association with the current bypass path.

   ```
   configure router mpls srlg-frr loose
   ```

   • If the updated SRLG information of the current bypass path changed and SRLG strict disjointness is configured using the following command, MPLS evaluates the SRLG constraint of each PLR and performs the following actions.

   ```
   configure router mpls srlg-frr strict
   ```

---

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

– 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.

– 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:

- 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.

- If the updated SRLG information of the current bypass path changed and SRLG FRR loose disjointness is configured using the following command option, MPLS keeps the current bypass path and the PSB associations of all PLRs.

```
configure router mpls srlg-frr loose
```

- If the updated SRLG information of the current bypass path changed and SRLG strict disjointness is configured, MPLS evaluates the SRLG constraint of each PLR and performs the following actions.

```
configure router mpls srlg-frr strict
```

– 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.

– 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.

2.3.6.2 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, MPLS signals and programs the new path. Because 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:

   a. MPLS moves to the new bypass path the PSB associations in which the administrative group constraints are not violated by the new bypass path.

   b. Among the remaining PSBs, MPLS does the following:

      • 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.

      • 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:

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

   • 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:

      – 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.

      – 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.

2.3.7 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** command 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.

   - If CSPF returns a new path, MPLS performs the following actions:

      – If CSPF finds the current path is invalid, MPLS signals and programs the new path.

      – If the updated administrative group membership of the current path violates the path administrative group constraint, MPLS signals and programs the new path.

      – 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.

   - 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.

2.3.8 DiffServ traffic engineering

DiffServ 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. DiffServ 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.

2.3.8.1 Mapping of traffic to a DiffServ LSP

An LER allows the user to map traffic to a DiffServ LSP using one of the following methods:

- explicit RSVP SDP configuration of a VLL, VPLS, or VPRN service

- 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 DiffServ Class Type (CT) configuration for this LSP.

- the **auto-bind-tunnel** RSVP-TE option in a VPRN service
- static routes with indirect next-hop being an RSVP LSP name

2.3.8.2 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 only considers 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 DiffServ TE is enabled. In other words, CSPF in the head-end node must 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.

2.3.8.2.1 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 does 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 DiffServ 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 DiffServ was enabled.

2.3.8.2.2 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 ≤ 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.

2.3.8.2.2.1 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 7: RDM with two class types.

Figure 7: RDM with two class types

al_0206
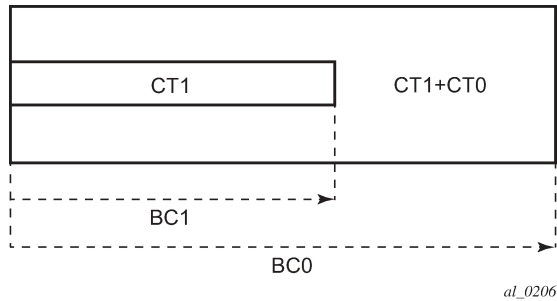
Suppose CT1 bandwidth, or the CT1 percentage of Maximum Reservable Bandwidth to be more accurate is 100 Mb/s and CT2 bandwidth is 100 Mb/s and link bandwidth is 200 Mb/s. BC constraints can be calculated as follows.

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 8: First LSP reservation

al_0207
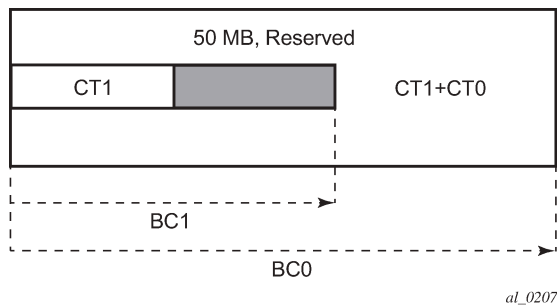
According to the RDM admission control policy:

Reserved (CT1, 0) = 50 ≤ 100 Mb/s

Reserved (CT0, 0) + Reserved (CT1, 0) = 50 ≤ 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 9: Second LSP reservation
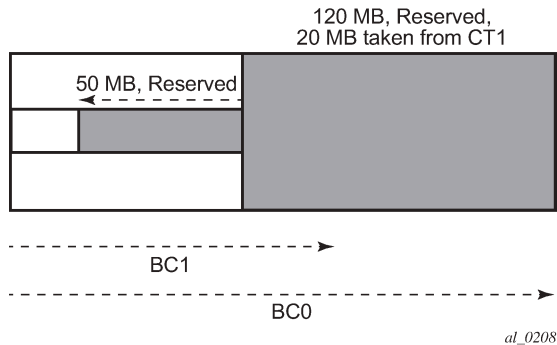
Reserved (CT0, 0) = 120 ≤ 150 Mb/s

Reserved (CT0, 0) + Reserved (CT1, 0) = 120 + 50 = 170 ≤ 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.

2.3.8.3 RSVP control plane extensions

RSVP uses 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 that is supported today on the router.

One or more forwarding classes map to a DiffServ class type trough a system level configuration.

2.3.8.4 IGP extensions

IGP extensions are defined in RFC 4124. DiffServ 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. 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 wanted class type on the LSP such that RSVP-TE can signal it in the class-type object in the path message.

IGP continues to advertise the existing Maximum Reservable Link Bandwidth TE parameter to mean the maximum bandwidth that can be booked on a specified interface by all classes. The value advertised is adjusted with the link subscription factor.

2.3.8.5 DiffServ TE configuration and operation

2.3.8.5.1 RSVP protocol level

The following are the configuration steps at the RSVP protocol level:

1. The user enables DiffServ TE. Use the following command to enable DiffServ TE.

```
configure router rsvp diffserv-te
```

When the **diffserv-te** command is enabled, IS-IS and OSPF advertise available bandwidth for each TE class configured under the **diffserv-te** context.

If required, use the following command to disable DiffServ TE globally:

- **MD-CLI**

```
configure router rsvp delete diffserv-te
```

- **classic CLI**

```
configure router rsvp no diffserv-te
```

2. In the classic CLI, enabling or disabling DiffServ TE on the system requires that the RSVP and MPLS protocols be administratively disabled.

   The user must ensure each context is administratively enabled after all configurations under both protocols are defined. When saved in the configuration file, both protocols are administratively enabled, to make sure they come up after a node reboot.

3. IGP advertises the available bandwidth in each TE class in the Unreserved Bandwidth TE command option for that class for each RSVP interface in the system.

4. In addition, IGP continues to advertise the existing Maximum Reservable Link Bandwidth TE command option, so the maximum bandwidth that can be booked on a specific interface by all classes. Use the following command to adjust the value advertised.

```
configure router rsvp interface subscription
```

5. The user can overbook or underbook the maximum reservable bandwidth of a CT by overbooking or underbooking the interface maximum reservable bandwidth by configuring the appropriate value for the **subscription** command.

6. The **diffserv-te** command only takes effect if the user has already enabled TE at the IS-IS or OSPF routing protocol levels. Use the following command to enable TE at the IS-IS routing protocol level:

- **MD-CLI**

```
configure router isis traffic-engineering true
```

- **classic CLI**

```
configure router isis traffic-engineering
```

Use the following command to enable TE at the OSPF routing protocol level:

- **MD-CLI**

```
configure router ospf traffic-engineering true
```

- **classic CLI**

```
configure router ospf traffic-engineering
```

7. The following DiffServ TE command options are configured globally under the **diffserv-te** context:

> **Note:** The command options apply to all RSVP interfaces on the system. In the classic CLI, the command options can only be changed after administratively disabling the MPLS and RSVP protocols.

- **definition of TE classes**

    A TE class is defined as the following:

    TE Class = {Class Type (CT), LSP priority}

    Use the command options in the following context to configure TE classes.

    ```
    configure router rsvp diffserv-te te-class
    ```

    Eight TE classes are supported. There is no default TE class when DiffServ is enabled. The user must explicitly define each TE class. However, when DiffServ is disabled, there is an internal use of the default CT (CT0) and eight preemption priorities, as described in the following table.

Table 6: Internal TE class definition when DiffServ TE is disabled

Class type (CT internal)	LSP priority
0	7
0	6
0	5
0	4
0	3
0	2
0	1
0	0

- **mapping of the system forwarding class to CT**

  Use the command options in the following context to map one or more system forwarding classes to a DiffServ CT.

  ```
  configure router rsvp diffserv-te fc
  ```

  The default settings are described in the following table.

Table 7: 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

- **percentage of RSVP interface bandwidth each CT shares**

  Use the command options in the following context to configure the percentage of RSVP interface bandwidth each CT shares, for example, the Bandwidth Constraint (BC).

  ```
  configure router rsvp diffserv-te class-type-bw
  ```

  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 command option, for example, the link bandwidth multiplied by the RSVP interface **subscription**.

  > **Note:** This configuration also exists at the RSVP interface level, and the interface-specific configured value overrides the globally configured value. The BC value can be changed at any time. The user can specify the BC for a CT, which is not used in any of the TE class definition and does not get used by any LSP originating or transiting this node.

- **configuration of the admission control policy**

  Use the following command to configure the admission control policy:

  - **MD-CLI**

    ```
    configure router rsvp diffserv-te admission-control-model
    ```

    – **classic CLI**

```
configure router rsvp diffserv-te
```

However, 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.

2.3.8.5.2 RSVP interface level

The following are the configuration steps at the RSVP interface level:

1. The user configures the percentage of RSVP interface bandwidth each CT shares, for example, the BC. Use the command options in the following context to configure the percentage.

```
configure router rsvp interface class-type-bw
```

The value entered at the interface level overrides the global value configured under the **diffserv-te** context.

2. The user can overbook or underbook the maximum reservable bandwidth of a specific CT by overbooking or underbooking the interface maximum reservable bandwidth. Use the following command to configure the appropriate value for either overbooking or underbooking the maximum reservable bandwidth.

```
configure router rsvp interface subscription
```

> **Note:** Both the BC and subscription values can be changed at any time.

2.3.8.5.3 LSP and LSP path levels

The following are the configuration steps at the LSP and LSP path levels:

1. The user configures the CT in which the LSP belongs. Use the following command to configure the CT for the LSP at the LSP level.

```
configure router mpls lsp class-type
```

Use the following command to configure the CT for the LSP at the path level.

```
configure router mpls lsp primary class-type
```

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 multiclass 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 user configured two 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 is not signaled and goes into the down state.

4. The user 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 user can also configure the setup and holding priorities for each LSP path.

7. An LSP which does not have the CT explicitly configured behaves like a CT0 LSP when DiffServ is enabled.

If the user configured a combination of a CT and a setup priority or a combination of a CT and a holding priority, or both, 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.

2.3.9 DiffServ TE LSP class type change under failure

The user can optionally configure a main CT and a backup CT for the primary path of a DiffServ TE LSP. 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 because of timer-based or manual resignal. The backup CT is used when the LSP retries under failure.

Use the following command to enable a backup CT for the LSP primary path.

```
configure router mpls lsp primary backup-class-type
```

When a backup CT is enabled, the LSP uses the CT configured using the following commands (whichever is inherited at the primary path level) as the main CT.

```
configure router mpls lsp class-type
configure router mpls lsp primary class-type
```

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: 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 DiffServ 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.

2.3.9.1 LSP primary path retry procedures

This feature behaves according to the following rules:

- When an LSP primary path retries because of a failure, for example, it fails after being in the up state, or undergoes any type of MBB, MPLS retries 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 as follows:

- – 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 or receipt, or both, 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 maximum value. Use the following command to configure the maximum number of retries.

```
configure router mpls lsp main-ct-retry-limit
```

  If the path did not come up, it starts using the backup CT. By default, this command is set to infinite value. The **main-ct-retry-limit** command has no effect on an LSP primary path, which retries because of a failure event. If the user configures the **main-ct-retry-limit** command to use a value that is greater than the LSP retry limit, the number of retries stops when the LSP primary path reaches the value of the LSP retry limit. In other words, the meaning of the LSP retry limit 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 administratively enabling or disabling 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 following command is executed, the retry behavior for this LSP is the same as it would be for an unmapped LSP.

```
clear router mpls lsp
```

- • If the value of the **main-ct-retry-limit** command is changed, the new value is only used at the next time the LSP path is administratively enabled.

- • The following is the behavior for 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, shut, 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 tries 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 because 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 does not check if there is a more optimal path, with the backup CT, according to the least-fill criterion and, so, does 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 tries 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.

2.3.9.2 Bandwidth sharing across class types

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 supports the Russian-Doll Model (RDM) DiffServ TE admission control policy described in RFC 4127, *Russian Dolls Bandwidth Constraints Model for Diffserv-aware MPLS Traffic Engineering*. Use the following command option to enable the feature:

- • **MD-CLI**

```
configure router rsvp diffserv-te admission-control-model rdm
```

- • **classic CLI**

```
configure 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. The following figure shows an example of RDM admission control policy.

Figure 10: RDM admission control policy example
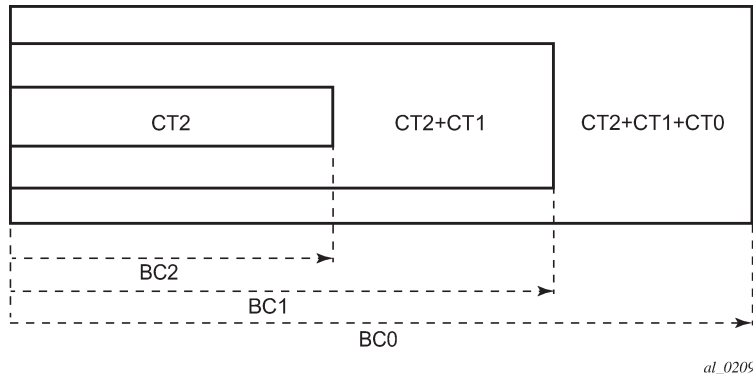
CT2 has a bandwidth constraint BC2 which represents a percentage of the maximum reservable link bandwidth. Both CT2 and CT1 can share BC1 which is the sum of the percentage of the maximum reservable bandwidth values configured for CT2 and CT1 respectively. Finally, CT2, CT1, and CT0 together can share 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 because 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.

2.3.9.3 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 11: Sharing bandwidth when an LSP primary path is downgraded to backup CT
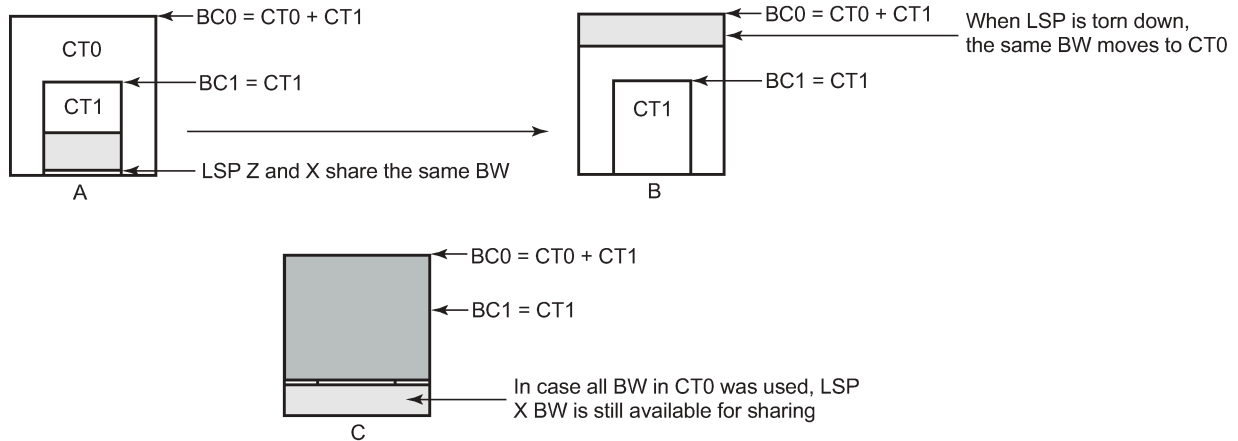
Consider an LSP path Z occupying bandwidth B at CT1. BC0 being the sum of all CTs below it, the bandwidth occupied in CT1 is guaranteed to be available in CT0. When new path X of the same LSP for CT0 is setup, it uses the same bandwidth B as used by path Z as shown in Figure 11: Sharing bandwidth when an LSP primary path is downgraded to backup CT (a). When path Z is torn down the same bandwidth now occupies CT0 as shown in Figure 11: Sharing bandwidth when an LSP primary path is downgraded to backup CT (b). Even if there were no new BW available in CT0 as can be seen in Figure 11: Sharing bandwidth when an LSP primary path is downgraded to backup CT (c), path X can always share the bandwidth with path Z.

CSPF at the head-end node and CAC at the transit LSR node shares bandwidth of an existing path when its CT is downgraded in the new path of the same LSP.

2.3.9.4 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 12: Sharing bandwidth when an LSP primary path is upgraded to main CT
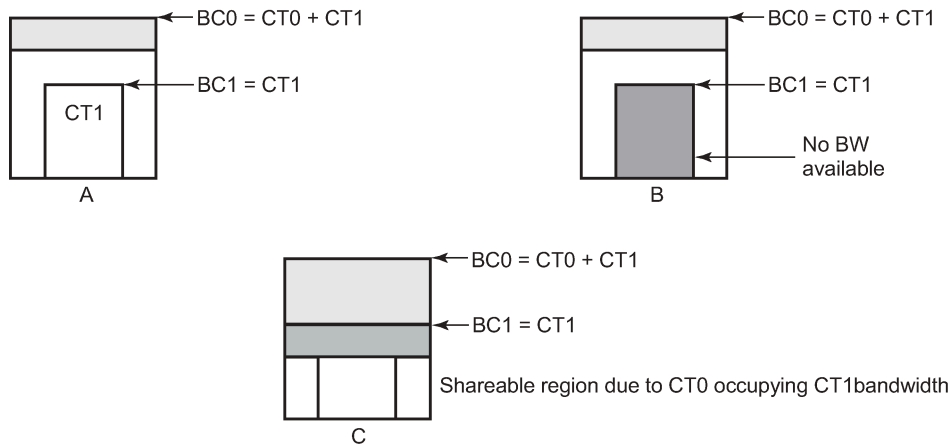
In Figure 12: Sharing bandwidth when an LSP primary path is upgraded to main CT (a), if the path X occupies the bandwidth as shown it cannot share the bandwidth with the new path Z being setup. If a condition exists, as shown in Figure 12: Sharing bandwidth when an LSP primary path is upgraded to main CT, (b) the path Z can never be setup on this particular link.

Consider Figure 12: Sharing bandwidth when an LSP primary path is upgraded to main CT (c). The CT0 has a region that overlaps with CT1 as CT0 has incursion into CT1. This overlap can be shared. However, 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 only attempts to signal the LSP path with an upgraded CT if the advertised bandwidth for that CT can accommodate the bandwidth. In other words, it assumes that in the worst case this path does not share bandwidth with another path of the same LSP using a lower CT.

2.4 Advanced MPLS/RSVP features

This section describes the advanced MPLS and RSVP features.

2.4.1 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 fails.

2.4.2 LSP path change

Use the following command to configure MPLS to replace the path of the primary or secondary LSP.

```
tools perform router mpls update-path
```

The primary or secondary LSP path is indirectly identified by the name of the current path. In the existing implementation, the same path name cannot be used more than once in a specific LSP name.

The **update-path** command is also supported on an SNMP interface.

The **update-path** command applies to both a CSPF LSP and to a non-CSPF LSP. However, it is only honored when the specified current path has the adaptive option enabled. The adaptive option can be enabled the LSP level or at the path level.

> **Note:** The new path must be first configured in CLI or provided through SNMP. Use the following command to configure the new path.
>
> ```
> configure router mpls path
> ```

The **update-path** command fails if any of the following conditions are true:

- The specified name of the current path of this LSP does not have the adaptive option enabled.
- The specified name of the new path does not correspond to a previously defined path.
- The specified name of the new path exists, but is being used by any path of the same LSP, including this one.

When the **update-path** command is used, 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, as follows:
  - MPLS writes the corresponding NHLFE in the datapath if this path is the current backup path for the primary.
  - If the current path is the active LSP path, it updates the path and writes the new NHLFE in the datapath, which causes 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.

2.4.3 Manual LSP path switch

The manual LSP path switch feature introduces a command, with both a base version and sticky version, to move the path of an LSP from a standby secondary to another standby secondary.

Use the following command (base version) to move the path of the LSP from a standby (or an active secondary) to another standby of the same priority.

```
tools perform router mpls switch-path
```

If a new standby path with a higher priority or a primary path comes up after the **switch-path** 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.

Use the following command (sticky version) to move the path of the LSP from a standby path to any other standby path, regardless of priority.

```
tools perform router mpls force-switch-path
```

The LSP remains in the specified path until either the path goes down, or the user configures the router to use the priority path.

2.4.4 MBB procedures for changing LSP and path configuration

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 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. FRR global revertive procedures

3. soft preemption of an LSP path

4. TE graceful shutdown procedures

5. update of secondary path because of an update to primary path SRLG

6. LSP primary or secondary path name change

7. LSP or path configuration command option change

In an earlier implementation, item 7 pertains to the following:

1. changing the amount of bandwidth reserved for the primary or secondary path

2. enabling the FRR option for an LSP

This feature extends the coverage of the MBB procedure to most other LSP level and path level command options as follows:

- including or excluding admin groups at LSP and path levels

- enabling or disabling **local-cspf** as the path computation method at the LSP level. Use the following command to configure the path computation method.

```
configure router mpls lsp path-computation-method
```

- configuring the **metric-type** at the LSP level. Use the following command to configure the metric type for the LSP.

```
configure router mpls lsp metric-type
```

- enabling or disabling **propagate-admin-group** at the LSP level. Use the following command to configure whether this behavior is enabled or disabled.

```
configure router mpls lsp propagate-admin-group
```

- configuring the **hop-limit** at the LSP level in the fast reroute context. Use the following command to configure the hop limit for fast reroute.

```
configure router mpls lsp fast-reroute hop-limit
```

- enabling **least-fill** at the LSP level. Use the following command to configure whether to use the least-fill path selection method.

```
configure router mpls lsp least-fill
```

- enabling or disabling **adspec** at the LSP level. Use the following command to configure whether the ADSPEC object is included in RSVP messages for the LSP.

```
configure router mpls lsp adspec
```

- enabling or disabling **node-protect** at the LSP level in the fast reroute context. Use the following command to configure whether node and link protection is enabled or disabled for the LSP.

```
configure router mpls lsp fast-reroute node-protect
```

- changing LSP primary or secondary path priority values (setup priority and hold priority). Use the command options in the following context to change the primary path priority values.

```
configure router mpls lsp primary priority
```

Use the command options in the following context to change the secondary path priority values.

```
configure router mpls lsp secondary priority
```

- changing LSP primary or secondary path **class-type** and primary path **backup-class-type**. Use the following command to change the class type for the LSP primary path.

```
configure router mpls lsp primary class-type
```

Use the following command to change the class type for the LSP secondary path.

```
configure router mpls lsp secondary class-type
```

Use the following command to change the backup class type for the LSP primary path.

```
configure router mpls lsp primary backup-class-type
```

- changing the **hop-limit** at the LSP level and path level. Use the following command to configure the hop limit at the LSP level.

```
configure router mpls lsp hop-limit
```

Use the following command to configure the hop limit for the LSP primary path.

```
configure router mpls lsp primary hop-limit
```

Use the following command to configure the hop limit for the LSP secondary path.

```
configure router mpls lsp secondary hop-limit
```

• enabling or disabling LSP primary or secondary path **record** and **record-label**. Use the following command to configure whether to record all hops that the LSP primary path traverses.

```
configure router mpls lsp primary record
```

Use the following command to configure whether to record all hops that the LSP secondary path traverses.

```
configure router mpls lsp secondary record
```

Use the following command to configure whether to record all labels at each node traversed by the LSP primary paths.

```
configure router mpls lsp primary record-label
```

Use the following command to configure whether to record all labels at each node traversed by the LSP secondary path.

```
configure router mpls lsp secondary record-label
```

This feature is not supported on a manual bypass LSP.

P2MP tree-level MBB operation is supported, if the following changes are made on the LSP template:

• changing bandwidth on the P2MP LSP template
• enabling FRR on the P2MP LSP template

2.4.5 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.

Use the command options in the following context to configure the feature.

```
configure router mpls auto-lsp
```

The user can associate multiple templates with the same or different peer prefix policies. Each application of an LSP template with a specific prefix in the prefix list results in the instantiation of a single CSPF computed LSP primary path using the LSP template command options, as long as the prefix corresponds to a router ID for a node in the TE database. Each instantiated LSP has a unique LSP ID and a unique tunnel ID.

Up to five peer prefix policies can be associated with a specific LSP template at all times. Each time the user executes the **auto-lsp** 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 tells 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 (or removed), or if a prefix range is expanded (or 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 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** command option is specified instead of a prefix policy, the **auto-lsp** 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 **auto-lsp** command is executed, the TE database keeps 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 used in conjunction with the **one-hop** command option results 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. Use the commands in the following context to configure the collection of egress statistics.

```
configure router mpls lsp-template egress-statistics
```

The user can also have ingress statistics collected at the egress LER. Use the commands in this following context to configure the collection of ingress statistics for an LSP.

```
configure router mpls ingress-statistics 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.

2.4.5.1 Automatic creation of RSVP mesh LSP: configuration and behavior

2.4.5.1.1 Feature configuration

First, use the following command to create an LSP template of type mesh P2P:
- **MD-CLI**

```
configure router mpls lsp-template type p2p-rsvp-mesh
```

- **classic CLI**

```
configure router mpls lsp-template mesh-p2p
```

Inside the template, the user configures the common LSP and path-level command options shared by all LSPs using the template.

After creating the LSP template, use the following command to associate the template with a peer prefix list:

- **MD-CLI**

```
configure router mpls auto-lsp policy
```

- **classic CLI**

```
configure router mpls auto-lsp lsp-template policy
```

> **Note:** The peer prefix list must be configured inside of a policy statement in the global policy manager.

The user can associate multiple templates with same or different peer prefix policies. Each application of an LSP template with a specific prefix in the prefix list results in the instantiation of a single CSPF computed LSP primary path using the LSP template command options, 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, the user must apply a separate LSP template to the same or different prefix list which contains the same destination node. Each instantiated LSP has 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 a non-CSPF option in the LSP template is therefore blocked.

Up to five peer prefix policies can be associated with an LSP template. Each time the user executes the **auto-lsp** 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 tells 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 (or removed) or if a prefix range is expanded (or shrunk) in a prefix list associated with an LSP template, the same prefix policy re-evaluation described above is performed.

The user must administratively enable the template before the configuration takes effect. In the classic CLI, when a template is in use, the user must administratively disable the template before making any changes to command options, except for the LSP command options for which changes can be made using MBB procedures. The user can configure the bandwidth and enable fast reroute for the LSP template without administratively disabling the template. Use the following command to configure the bandwidth.

```
configure router mpls lsp-template bandwidth
```

Use the following command to enable Fast Reroute (FRR).

```
configure router mpls lsp-template fast-reroute
```

> **Note:** FRR can be enabled with or without **hop-limit** or **node-protect**.

For all other command options, the user administratively disables the template and after it is added, removed, or modified, the existing instances of the LSP using this template are torn down and resignaled.

Finally the auto-created mesh LSP can be signaled over both numbered and unnumbered RSVP interfaces.

2.4.5.1.2 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 the auto-binding of a PW template, using an already provisioned SDP in a 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** configuration in the LSP template, an MBB is performed for all LSPs using the template. However, if automatic adjustments of the LSP bandwidth was enabled in the template, the **bandwidth** change is saved, but only takes effect at the next time the LSP bounces or is resignaled.

Use the following command to enable automatic adjustments of the LSP bandwidth for an LSP template.

```
configure router mpls lsp-template auto-bandwidth
```

Except for the MBB limitations to the configuration parameter change in the LSP template, MBB procedures for manual and timer based resignaling of the LSP, for TE Graceful Shutdown and for soft pre-emption are supported.

Note: The following command is not supported with a mesh LSP.

```
tools perform router mpls update-path
```

The one-to-one method for 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 performs an update to MPLS module which states router ID is no longer in TE database. This causes 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 is only 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 for this router ID gets deleted and re-established. All other failure conditions cause the LSP to activate the bypass backup LSP or to go down without being deleted.

2.4.5.1.3 Multi-area and multi-instance support

A router which does not have TE links within a specific IGP area or level does not have its router ID discovered in the TE database by other routers in the area or level. In other words, an auto-LSP of type P2P mesh cannot be signaled to a router which does not participate in the area or 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 IS-IS 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 IS-IS 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, it is not operational as long as the other IGP instances remain up. The LSP remains up and forwards traffic using the same TE links. The same behavior exists with a provisioned LSP.

2.4.5.1.4 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.

Use the commands in the following context to configure the collection of egress statistics for the automatically created LSP.

```
configure router mpls lsp-template egress-statistics
```

If there are no stat indexes available when an LSP is instantiated, the assignment is failed and the egress-statistics field in the show command for the LSP path displays as in the operationally down state, but in the administratively 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. Use the following command to enable ingress statistics on the egress LER:

- **MD-CLI**

```
configure router mpls ingress-statistics lsp sender ip-address lsp-name lsp-name
```

- **classic CLI**

```
configure router mpls ingress-statistics lsp lsp-name sender ip-address
```

2.4.5.2 Automatic creation of RSVP one-hop LSP: configuration and behavior

2.4.5.2.1 Feature configuration

The user first creates a one-hop LSP template. Use the following command to create a one-hop LSP template:

- **MD-CLI**

```
configure router mpls lsp-template template-name type p2p-rsvp-one-hop
```

- **classic CLI**

```
configure router mpls lsp-template template-name one-hop-p2p
```

After creating a one-hop LSP template, use the following command to enable the automatic signaling of one-hop LSP to all direct neighbors:

- **MD-CLI**

```
configure router mpls auto-lsp template-name one-hop
```

- **classic CLI**

```
configure router mpls auto-lsp lsp-template template-name one-hop
```

The LSP and path command options supported for a one-hop LSP template are the same as those supported for a mesh P2P LSP template, except for the **from** command, which is not allowed in a one-hop LSP template.

The **show** command for the auto-LSP displays the outgoing interface address in the from field.

The autocreated one-hop LSP can be signaled over both numbered and unnumbered RSVP interfaces.

2.4.5.2.2 Feature behavior

Although the provisioning model and CLI commands differ from that of a mesh LSP only by the absence of a prefix list, the actual behavior is quite different. When the **auto-lsp** command is executed, the TE database keeps track of each TE link which comes up to a directly connected IGP neighbor for which the 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, if the **auto-lsp** command is used in conjunction with the **one-hop** command option, one or more LSPs are signaled to the IGP neighbor.

Only the router ID of the first IGP instance of the neighbor which advertises a TE link causes 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 now associated with the other router ID.

The state of a one-hop LSP when 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 shut down 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 performs two separate updates to MPLS module. The first one updates the loss of the TE link association, which causes action (B) above for the one-hop LSP. The other update states router ID is no longer in TE database which causes 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 causes 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 a mesh P2P auto-LSP.

2.4.6 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 always uses 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 triggers 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 shortcuts are 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 corresponds to a router ID of a remote node.

Use the commands in the following contexts respectively to configure IGP shortcuts for an IS-IS, OSPFv2, or OSPFv3 instance.

```
configure router isis igp-shortcut
configure router ospf igp-shortcut
configure router ospf3 igp-shortcut
```

Use the following command to configure the destination address of the LSP.

```
configure router mpls lsp to
```

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 described in IGP shortcut feature configuration.

Nokia recommends disabling IGP shortcuts on RSVP LSPs which have 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, therefore, 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 8: RSVP LSP role as outcome of LSP level and IGP level configuration options summarizes the RSVP LSP role as an outcome of mixing these configuration options.

Table 8: RSVP LSP role as outcome of LSP level and IGP level configuration options

LSP level configuration	IGP instance level configurations					
	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

Use the following commands to disable the resolution of IGP routes using IGP shortcuts for an IS-IS, OSPFv2, or OSPFv3 instance respectively:

- **MD-CLI**

```
configure router isis igp-shortcut admin-state disable
```

```
configure router ospf igp-shortcut admin-state disable
configure router ospf3 igp-shortcut admin-state disable
```

- **classic CLI**

```
configure router isis igp-shortcut shutdown
configure router ospf igp-shortcut shutdown
configure router ospf3 igp-shortcut shutdown
```

2.4.6.1 IGP shortcut feature configuration

Use the commands in the following contexts respectively to enable the resolution over IGP IPv4 shortcuts of IPv4 and IPv6 prefixes within an IS-IS instance, of IPv6 prefixes within an OSPFv3 instance, and of IPv4 prefixes within an OSPFv2 instance.

```
configure router isis igp-shortcut
configure router ospf igp-shortcut
configure router ospf3 igp-shortcut
```

The resolution node **igp-shortcut** provides flexibility in the selection of the IP next hops or the tunnel types for each of the IPv4 and IPv6 prefix families.

For IS-IS and OSPF instances, when the **family** command is configured to **ipv4**, 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.

For IS-IS and OSPFv3 instances, when the **family** command is configured to **ipv6**, 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 IS-IS 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 datapath, a packet for an IPv6 prefix contains 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.

Use the following commands to control the use of an RSVP-TE LSP in IGP shortcuts:

- **MD-CLI**

```
configure router mpls lsp igp-shortcut lfa-type lfa-protect
configure router mpls lsp igp-shortcut lfa-type lfa-only
configure router mpls lsp igp-shortcut relative-metric
```

- **classic CLI**

```
configure router mpls lsp igp-shortcut [lfa-protect | lfa-only]
configure 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.

For tunneling LDP IPv4 FECs over IGP shortcuts, the user must configure the **family** command to **ipv4** and enable the **tunneling** option on the T-LDP sessions to the destinations of the RSVP-TE LSPs used as IGP shortcuts.

Tunneling of LDP IPv6 FECs over IGP shortcuts is not supported.

2.4.6.1.1 IGP shortcut binding construct

Use the commands in following contexts respectively to configure the IGP shortcut binding construct for an IS-IS, OSPFv2, or OSPFv3 instance.

```
configure router isis igp-shortcut tunnel-next-hop
configure router ospf igp-shortcut tunnel-next-hop
configure router ospf3 igp-shortcut tunnel-next-hop
```

Commands in the **tunnel-next-hop** context bind IP prefixes to IPv4 IGP shortcuts on a per-prefix family basis.

The following describes 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** command (forwarding adjacency) takes priority over **igp-shortcut**, if both commands are enabled. This applies 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

  See Table 8: RSVP LSP role as outcome of LSP level and IGP level configuration options for more information.

- IPv4 prefixes do not automatically resolve to RSVP LSPs used as IGP shortcuts when the **igp-shortcut** context is enabled. The IPv4 family must be enabled, and the **rsvp** command in the **resolution-filter** context must be configured, which selects the RSVP-TE tunnel type.

- If the **igp-shortcut** context is administratively disabled, it cannot be enabled unless at least one family is configured under the **tunnel-next-hop** context. The resolution for IGP shortcut tunnels must also not be disabled, which is the default behavior for all families. If the **resolution** command is configured to **filter**, a tunnel type must be configured using commands in the **resolution-filter** context.

- To disable IGP shortcuts globally, administratively disable 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 configured using the commands in the **tunnel-next-hop** context.

The resolution outcome for each of the IPv4 and IPv6 prefix families is summarized in Table 9: IGP shortcut binding resolution outcome (MD-CLI) and Table 10: IGP shortcut binding resolution outcome (classic CLI). The description and behavior of the SRv4 and SRv6 families are described in SR shortest path tunnel over RSVP-TE IGP shortcut feature configuration. See IPv4 IGP shortcuts using SR-TE LSP feature configuration for information about the description and behavior of the sr-te resolution option using SR-TE IGP shortcuts are described in the following tables.

Table 9: IGP shortcut binding resolution outcome (MD-CLI)

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
admin-state disable	—	—	IP (unicast RTM)	IP (mcast RTM)	IP (unicast RTM)	IP (mcast RTM)
admin-state enable	**resolution none**	**resolution none**	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)
admin-state enable	**rsvp true**	**resolution none**	RSVP+IP	IP (mcast RTM)	RSVP+IP	IP (mcast RTM)
		resolution match-family-ip	RSVP+IP	RSVP+IP	RSVP+IP	RSVP+IP
admin-state enable	**sr-te true**	**resolution none**	SR-TE+IP	IP (mcast RTM)	SR-TE+IP	IP (mcast RTM)
		resolution match-family-ip	SR-TE+IP	IP (mcast RTM)	SR-TE+IP	IP (mcast RTM)
admin-state enable	**resolution any** and **rsvp true** or **sr-te true**	**resolution none**	RSVP+IP	IP (mcast RTM)	RSVP+IP	IP (mcast RTM)
			SR-TE+IP	IP (mcast RTM)	SR-TE+IP	IP (mcast RTM)
		resolution match-family-ip	RSVP+IP	RSVP+IP	RSVP+IP	RSVP+IP
			SR-TE+IP	IP (mcast RTM)	SR-TE+IP	IP (mcast RTM)

Table 10: IGP shortcut binding resolution outcome (classic CLI)

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
shutdown	—	—	IP (unicast RTM)	IP (mcast RTM)	IP (unicast RTM)	IP (mcast RTM)

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
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)
no shutdown	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
no shutdown	sr-te	resolution disabled	SR-TE+IP	IP (mcast RTM)	SR-TE+IP	IP (mcast RTM)
		resolution match-family-ip	SR-TE+IP	IP (mcast RTM)	SR-TE+IP	IP (mcast RTM)
no shutdown	resolution any and rsvp or sr-te	resolution disabled	RSVP+IP	IP (mcast RTM)	RSVP+IP	IP (mcast RTM)
			SR-TE+IP	IP (mcast RTM)	SR-TE+IP	IP (mcast RTM)
		resolution match-family-ip	RSVP+IP	RSVP+IP	RSVP+IP	RSVP+IP
			SR-TE+IP	IP (mcast RTM)	SR-TE+IP	IP (mcast RTM)

2.4.6.2 IPv4 IGP shortcuts using SR-TE LSP feature configuration

Use the following commands respectively to configure the SR-TE LSP feature for an IS-IS, OSPFv2, or OSPFv3 instance.

```
configure router isis igp-shortcut tunnel-next-hop family resolution-filter sr-te
configure router ospf igp-shortcut tunnel-next-hop family resolution-filter sr-te
configure router ospf3 igp-shortcut tunnel-next-hop family resolution-filter sr-te
```

When enabled, this feature allows IGP to resolve IPv4 prefixes, IPv6 prefixes, and LDP IPv4 prefix FECs over SR-TE LSPs used as IGP shortcuts.

If applicable, configure the **resolution** command to use the **any** command option, to resolve IP prefixes and LDP FECs to either RSVP-TE or SR-TE LSPs used as IGP shortcuts.

For tunneling LDP FECs over IGP shortcuts, the user must enable the tunneling option on the T-LDP sessions to the destinations of the SR-TE LSPs used as IGP shortcuts. 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.

2.4.6.2.1 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 and multicast routing tables. 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:

1. OSPF supports IPv4 prefixes by configuring the **family** command to **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.

2. OSPFv3 supports IPv6 prefixes by configuring the **family** command to **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.

3. IS-IS supports IPv4 prefixes in MT=0 by configuring **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.

4. Configuring the **family** command to **ipv4** enables the resolution in the unicast routing table of LDP IPv4 prefix FEC in OSPF or IS-IS. When the following command is enabled 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.

   ```
   configure router ldp prefer-tunnel-in-tunnel
   ```

   A maximum of 32 ECMP tunnels next hops can be programmed for an LDP FEC.

   LDP IPv6 prefix FECs are not supported over IPv4 IGP shortcuts when configuring the **family** command to **ipv6**. Consequently, if the corresponding IPv6 prefix resolves to tunnel next hops only, the LDP IPv6 prefix FEC remains unresolved. After the **ipv6** command option is configured, use the following command to enable LDP IPv6 prefix FECs resolution.

   ```
   configure router ldp targeted-session resolve-v6-prefix-over-shortcut
   ```

5. In all cases, the IP reach calculation in the unicast routing table first follows 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 prefix family. 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. That is, the RSVP-TE LSP type is preferred to a SR-TE LSP type on a per-prefix basis.

6. One or more SR-TE LSPs can be selected in the unicast routing table, if the **resolution** command is configured to **filter**, and the **sr-te** command in the **resolution-filter** context is enabled.

7. One or more SR-TE LSPs can also be selected in the unicast routing table, if the **resolution** command is configured to **any**, and one or more SR-TE LSPs are available, but no RSVP-TE LSPs are available for resolving the prefix by IGP.

8. An intra-area IP prefix of IPv4 or IPv6, or an LDP IPv4 prefix FEC always resolves to a single type of tunnel, which can be either RSVP-TE or SR-TE. The RSVP-TE tunnel 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.

9. An inter-area IP prefix of IPv4 or IPv6, or an LDP IPv4 prefix FEC 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 are not used in the set of ECMP tunnel next hops for the inter-area prefix.

10. The feature does not support configuring a different tunnel type for each prefix family, using the **resolution-filter** command. Administratively disabling the **igp-shortcut** context fails if the user configured IPv4 prefixes to resolve to SR-TE and IPv6 prefixes to resolve to RSVP-TE (or the other way around). This is applies to both inter-area and intra-area prefixes.

    The feature, however, supports selecting the best tunnel type per prefix within each family, as described in (5). For example, the **family** command can be configured to **ipv4** and **ipv6**, in conjunction with configuring the **resolution** command to **any**. On a per prefix basis, the best tunnel type is selected, therefore allowing both tunnel types to be deployed in the network.

11. The user can disable the resolution for each family independently, which disables IGP shortcuts for the corresponding prefix family in this IGP instance. IP prefixes and LDP FECs of the family resolve over IP links in the multicast routing table.

2.4.6.2.2 Application support

The use of SR-TE IGP shortcuts is supported in the following applications:

1. Configuring the **family** command to **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

2. Configuring the **family** command to **ipv6** resolves IPv6 prefixes in RTM for the following:
    - IGP routes
    - indirect next hop of static routes
    - BGP next hop of BGP routes

3. When an LDP IPv4 FEC prefix is resolved to one or more SR-TE LSPs, the following applications can resolve to LDP in TTM:
    - L2 service FECs
    - BGP next hop of VPN IPv4 and 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

4. When an LDP IPv4 FEC prefix is resolved to one or more SR-TE LSPs, next hops of BGP LU routes cannot resolve to LDP in TTM.

> **Note:** SR OS supports a 3-level hierarchy in the datapath. Because SR-TE LSP is a hierarchical LSP already, this makes the BGP-over-LDP-over-SR-TE a 4-level hierarchy. Consequently, BGP keeps these BGP-LU routes unresolved.

2.4.6.2.3 LFA protection support

The following are the details of the Loop-free Alternate (LFA) protection support:

- 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.

- Use the following command option to include an SR-TE LSP in the LFA SPF, without the introduction of IGP shortcuts impacting the main SPF decision:

  - **MD-CLI**

    ```
    configure router mpls lsp igp-shortcut lfa-type lfa-only
    ```

  - **classic CLI**

    ```
    configure router mpls lsp igp-shortcut lfa-only
    ```

  If the user enables the **lfa-only** command option for a specific SR-TE LSP, and the application prefix has a single IP primary next hop (no ECMP next hops), it is protected by an LFA backup that uses an SR-TE LSP.

  > **Note:** The LFA SPF calculation cannot check whether the outgoing interface of the protecting SR-TE LSP is different from the primary next hop of the prefix. The prefix is still protected by either the ECMP next hops or the LFA backup next hop of the first segment of the protecting SR-TE LSP. However, in the case where an RSVP-TE LSP is used with the **lfa-only** command option, such an LSP is excluded from being used as an LFA backup next hop.

- 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 (1) or the LDP LFA backup using an SR-TE LSP (2).

- Assume the resolution of the IP prefix is 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 and 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 (1), the feature supports computing an 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.

- Assume the resolution of the IP prefix is 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 and 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 (2), 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 enables the remote LFA or the TI-LFA feature and also enables the use of SR IPv6 or SR IPv6 tunnels as an LFA backup next hop by the LDP IPv6 or IPv4 FEC prefix. Use the following command to protect the LDP FEC, if a backup SR tunnel is found:

– **MD-CLI**

```
configure router ldp fast-reroute backup-sr-tunnel true
```

– **classic CLI**

```
configure router ldp fast-reroute backup-sr-tunnel
```

2.4.6.3 SR shortest path tunnel over RSVP-TE IGP shortcut feature configuration

Use the following command option to allow IGP to resolve SR-ISIS IPv4 tunnels in MT=0 RSVP-TE LSPs used as IGP shortcuts.

```
configure router isis igp-shortcut tunnel-next-hop family srv4
```

Use the following command option to allow IGP to resolve SR-OSPF IPv4 tunnels over RSVP-TE LSPs used as IGP shortcuts.

```
configure router ospf igp-shortcut tunnel-next-hop family srv4
```

Use the following command option to allow IGP to resolve SR-ISIS IPv6 tunnels in MT=0 over RSVP-TE LSPs used as IGP shortcuts.

```
configure router isis igp-shortcut tunnel-next-hop family srv6
```

See Family prefix resolution and tunnel selection rules for applicable rules for the resolution of SR-ISIS IPv4 tunnels, SR-ISIS IPv6 tunnels, and SR-OSPF IPV4 tunnels, and the selection of tunnel types on a per-family basis.

2.4.6.3.1 Family prefix resolution and tunnel selection rules

The following are the details of the resolution of prefix families in the unicast and multicast routing tables:

- Configuring the **family** command to **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.

- Configuring the **family** command to **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.

- One or more RSVP-TE LSPs can be selected if the **resolution** command is configured to **match-family-ip**, and the corresponding IPv4 or IPv6 prefix is resolved to RSVP-TE LSPs.

- An SR tunnel cannot resolve to SR-TE IGP shortcuts. If the **resolution** command is configured to **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.

- For an SR tunnel corresponding to an inter-area prefix with best routes via multiple ABRs, configuring the **resolution** command to **match-family-ip** means that 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 does 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 is also tunneled over RSVP-TE LSP.

- The resolution is disabled by default for the IPv4 and IPv6 tunnel families, which means that SR-ISIS and SR-OSPF tunnels are resolved to IP links in the multicast routing table.

2.4.6.3.2 Application support

The following describes how SR-ISIS IPv4 or IPv6 or a SR-OSPF IPv4 tunnels are resolved.

1. When an SR-ISIS IPv4 or an SR-OSPF IPv4 tunnel is resolved to one or more RSVP-TE LSPs, 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

2. When an SR-ISIS IPv6 tunnel is resolved to one or more RSVP-TE LSPs, 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

3. When an SR-ISIS IPv4 or an SR-OSPF IPv4 tunnel is resolved to one or more RSVP-TE LSPs, next hops of BGP LU routes cannot resolve in TTM to a SR-TE LSP that is using an SR-ISIS or SR-OSPF segment.

   > **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 a 3-level hierarchy in the

datapath and, because SR-TE LSP is a hierarchical LSP already, this makes the BGP-over-SRTE-over-RSVPTE a 4-level hierarchy. BGP keeps these BGP-LU routes unresolved.

2.4.6.3.3 LFA protection support

The following are the details of the LFA protection support:

1. 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.

2. If the user enables the **lfa-only** command option for a specified RSVP-TE LSP, and the SR-ISIS or SR-OSPF tunnel has a single IP primary next hop (no ECMP next hops), it can be protected by a FRR backup that uses a RSVP-TE LSP.

3. Applications that resolve in TTM to an SR-ISIS or SR-OSPF tunnel, which itself is resolved to one or more RSVP-TE LSPs, are equally be protected either by the RSVP-TE LSP FRR (1) or the SR LFA using a RSVP-TE LSP (2).

4. Assume **family** configured to **ipv4** resolves to RSVP-TE in the unicast routing table, while **family** configured to **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 (1), this feature supports computing an LFA next hop for the SR IPv4 tunnel of the same prefix using IP next hops.

5. Assume **family** configured to **ipv4** or **ipv6** resolves to RSVP-TE in the unicast routing table, while **family** configured to **srv4** or **srv6** resolves to IP links in the multicast routing table. If the IP prefix of an SR IPv4 or SR IPv6 tunnel is resolved to a single IP primary next hop and is protected by an SR LFA backup using an RSVP-TE LSP FRR (2), the feature does not support computing a LFA next hop for the SR IPv4 or SR IPv6 tunnel and remains unprotected.

   If, however, the user enabled the remote LFA or the TI-LFA feature, an SR backup next hop may be found for the SR IPv4 or SR IPv6 tunnel, which then becomes protected.

2.4.6.4 Using LSP relative metric with IGP shortcut

By default, the absolute metric of the LSP is used to compute the contribution of an 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 TTM. This corresponds to the cumulative IGP metric of the LSP path returned by CSPF or the static administrative metric value of the LSP, if the user configured the metric. Use the following command to configure the static administrative metric.

```
configure router mpls lsp metric
```

Note: 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.

Use the following command option to configure the relative metric offset for an IGP shortcut.

```
configure router mpls lsp igp-shortcut relative-metric
```

IGP applies 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 that is resolved to the LSP.

The offset value is optional and is zero by default. An offset value of zero is used when the **relative-metric** command option is enabled without specifying the offset 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:** The TTM continues to show the LSP operational metric as provided by MPLS, which allows applications such as LDP-over-RSVP (when IGP shortcut is disabled) and BGP and static route shortcuts to continue to use the LSP operational metric.

The **relative-metric**, **lfa-protect**, and **lfa-only** command options are mutually exclusive. If the **relative-metric** command option is enabled for an LSP, it cannot be included in the LFA SPF, and the other way around when **igp-shortcut** is enabled in the IGP.

The **relative-metric** configuration is ignored when forwarding adjacency is enabled in IS-IS or OSPF, using the **advertise-tunnel-link** command. 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.

2.4.6.5 ECMP considerations

When ECMP is enabled on the system and multiple equal-cost paths exist for a prefix, the following selection criteria are used to select the set of next hops to program in the datapath:

- 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:
    - select lowest endpoint to destination cost
    - if same endpoint to destination cost, select lowest endpoint node router-id
    - 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 sprays 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.

2.4.6.6 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 because RSVP LSPs with a destination address different than the router-id are not included by IGP in its SPF calculation.

2.4.6.7 Forwarding adjacency

The forwarding adjacency feature can be enabled independently from the IGP shortcut feature. Use the following command to enable forwarding adjacency in IS-IS.

```
configure router isis advertise-tunnel-link
```

Use the following command to enable forwarding adjacency in OSPF.

```
configure router ospf advertise-tunnel-link
```

If both the **igp-shortcut** and **advertise-tunnel-link** commands are enabled for a specific IGP instance, the **advertise-tunnel-link** command wins. With this feature, IS-IS 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 described in 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 or LSP advertising the corresponding P2P unnumbered link is installed in the local routing database, the node performs an SPF and uses it like any other link LSA or LSP. The link bidirectional check requires that a link, regular 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 the IPv4 family in OSPFv2, the IPv6 family in OSPFv3, the IPv4 and IPv6 families in IS-IS MT=0, and the IPv6 family in IS-IS MT=2. The **igp-shortcut** command under the LSP governs the use of the LSP with both the **igp-shortcut** and **advertise-tunnel-link** commands in IGP. Table 11: Impact of LSP level configuration on IGP shortcut and forwarding adjacency features describes the interactions of the forwarding adjacency feature.

Table 11: 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 not 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
In the MD-CLI, **lfa-type** configured to **lfa-protect**	Tunnel is used in main SPF and in LFA SPF	Tunnel is advertised as a P2P link if it has the best LSP metric, is used in the main SPF if

LSP level configuration	Actions with IGP shortcut feature	Actions with forwarding adjacency feature
In the classic CLI, **igp-shortcut** configured to **lfa-protect**		advertised, and is used in LFA SPF regardless of whether it is advertised or not
In the MD-CLI, **lfa-type** configured to **lfa-only** In the classic CLI, **igp-shortcut** configured to **lfa-only**	Tunnel is not used in main SPF, but used in LFA SPF	Tunnel is not advertised as a P2P link, if not used in main SPF, but is used in LFA SPF

2.4.6.8 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 a 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.

2.4.6.9 LDP forwarding over IGP shortcut

The user can enable the resolution of LDP FECs over IGP shortcuts by enabling IGP shortcut for family IPv4 (see IGP shortcut feature configuration) and by configuring T-LDP sessions to the destination of the RSVP LSP. In this case, LDP IPv4 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** commands are mutually exclusive under OSPF or IS-IS.

Similarly, LDP IPv4 FECs can be tunneled over SR-TE LSPs as detailed in IPv4 IGP shortcuts using SR-TE LSP feature configuration.

Tunneling of LDP IPv6 FECs over IGP shortcuts is not supported.

2.4.6.10 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, see the *7705 SAR Gen 2 Router Configuration Guide*.

2.4.6.11 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 because PIM requires a bidirectional path to the source but IGP shortcuts are unidirectional.

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** and **multicast-import** commands 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 does 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) only work with ECMP next-hops when IGP shortcuts are enabled.

Concurrently enabling the **advertise-tunnel-link** and **multicast-import** commands results a multicast RTM that is a copy of the unicast RTM and is populated with mix of IP and tunnel NHs. RPF succeeds for a prefix resolved to a IP NH, but fails for a prefix resolved to a tunnel NH. Table 12: Impact of IGP shortcut and forwarding adjacency on unicast and multicast RTM summarizes the interaction of the **igp-shortcut** and **advertise-tunnel-link** commands with unicast and multicast RTMs.

Table 12: 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**	✓	✓[3]	✓	X[4]
	advertise-tunnel-link	✓	✓[5]	✓	✓[6]

---

[3] 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.

[4] 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, however, is computed and installed in the unicast RTM for a primary IP next-hop of a prefix.

[5] Multicast RTM is a copy of the unicast RTM and, so, is populated with mix of IP and tunnel NHs. RPF succeeds for a prefix resolved to a IP NH but fails for a prefix resolved to a tunnel NH.

		Unicast RTM (primary SPF)	Multicast RTM (primary SPF)	Unicast RTM (LFA SPF)	Multicast RTM (LFA SPF)
IS-IS	igp-shortcut	✓	✓[3]	✓	X[4]
	advertise-tunnel-link	✓	✓[5]	✓	✓[6]

2.4.7 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:

Use the following global commands to configure TTL handling for all RSVP LSP shortcuts originating on an ingress LER.

```
configure router mpls shortcut-transit-ttl-propagate
configure router mpls shortcut-local-ttl-propagate
```

The configuration of the preceding commands applies to all RSVP LSPs used to resolve static routes, BGP routes, and IGP routes.

If the **shortcut-local-ttl-propagate** command is not enabled, TTL propagation is disabled on all locally generated IP packets, including ICMP ping, traceroute, and OAM packets that are destined for 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, if the **shortcut-transit-ttl-propagate** command is not enabled, TTL propagation is disabled on all IP packets received on any IES interface and destined for a route that is resolved to the LSP shortcut. In this case, a TTL of 255 is programmed onto the pushed label stack.

2.4.8 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:

---

[6] Multicast RTM is a copy of the unicast RTM and, so, is populated with mix of IP and tunnel LFA NHs. RPF succeeds for a prefix resolved to a primary or LFA IP NH but fails for a prefix resolved to a primary or LFA tunnel NH.

- adaptive

- admin-group

- bandwidth

- CSPF calculation

- fast-reroute

- hop-limit

- record-label

- retry-timer

2.4.9 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 SRLGs are applied to MPLS interfaces, Constraint-based Shortest Path First (CSPF) at an LER excludes the SRLGs of interfaces used by the LSP primary path when computing the path of the secondary path. CSPF at an LER or LSR also excludes the SRLGs of the outgoing interface of the primary LSP path in the computation of the path of the Fast Reroute (FRR) backup LSP. This provides path disjointedness between the primary path and the secondary path or FRR backup path of an LSP.

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.

When SRLGs are applied to IES, VPRN, or network IP interfaces, they are evaluated in the route next-hop selection by adding the SRLG option in a route next-hop policy template applied to an interface or a set of prefixes. For instance, the user can enable the SRLG constraint to select an LFA next-hop for a prefix that avoids all interfaces that share the outcome of the primary next hop.

During provisioning, the system rejects the creation of an SRLG if it reuses the same name with a different group value from an existing group, or if it reuses the same group value with a different name.

Only the SRLGs bound to an MPLS interface are advertised area-wide in TE link TLVs and sub-TLVs when the traffic-engineering option is enabled in IS-IS or OSPF. IES and VPRN interfaces do not have their attributes advertised in TE TLVs.

A user can specify a penalty weight associated with an SRLG. This controls the likelihood of bypass or detour LSP using paths with links that share SRLG values with a primary path. The higher the penalty weight, the less preferred it is to use the link with the SRLG.

Use the following command to specify a penalty weight associated with an SRLG:

- **MD-CLI**

```
configure routing-options if-attribute srlg-group penalty-weight
```

- **classic CLI**

```
configure router if-attribute srlg-group value penalty-weight
```

2.4.9.1 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 or detour LSPs that minimizes the probability of fate sharing with the path of the primary LSP (Figure 13: Shared Risk Link Groups).

The following describes the steps necessary to create SRLGs:

- For primary or standby SRLG disjoint configuration, perform the following steps:

  1. Create an SRLG group, similar to admin groups.

  2. Link the SRLG group to MPLS interfaces.

  3. Configure primary and secondary LSP paths and enable SRLG on the secondary LSP path.

     > **Note:** The SRLG secondary LSP paths always perform a strict CSPF query. The **srlg-frr** command is irrelevant in this case.

- For FRR detours or bypass SRLG disjoint configuration, perform the following steps:

  1. Create an SRLG group, similar to admin groups.

  2. Link the SRLG group to MPLS interfaces.

  3. Use the following system-wide command to force every LSP path CSPF calculation to take the configured SRLG memberships (including those propagated through the IGP opaque TE database) into account.

     ```
     configure router mpls srlg-frr
     ```

  4. Configure primary FRR (one-to-one or facility) LSP paths. Each PLR creates a detour or bypass that only avoids 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 does 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, priority is given in the following order:

     a. manual bypass disjoint

     b. manual bypass non-disjoint (only eligible if the **srlg-frr** command is not configured to **strict**)

c. dynamic disjoint

d. dynamic non-disjoint (only eligible if the **srlg-frr** command is not configured to **strict**)

> **Note:** Non-CSPF manual bypass is not considered.

Figure 13: 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
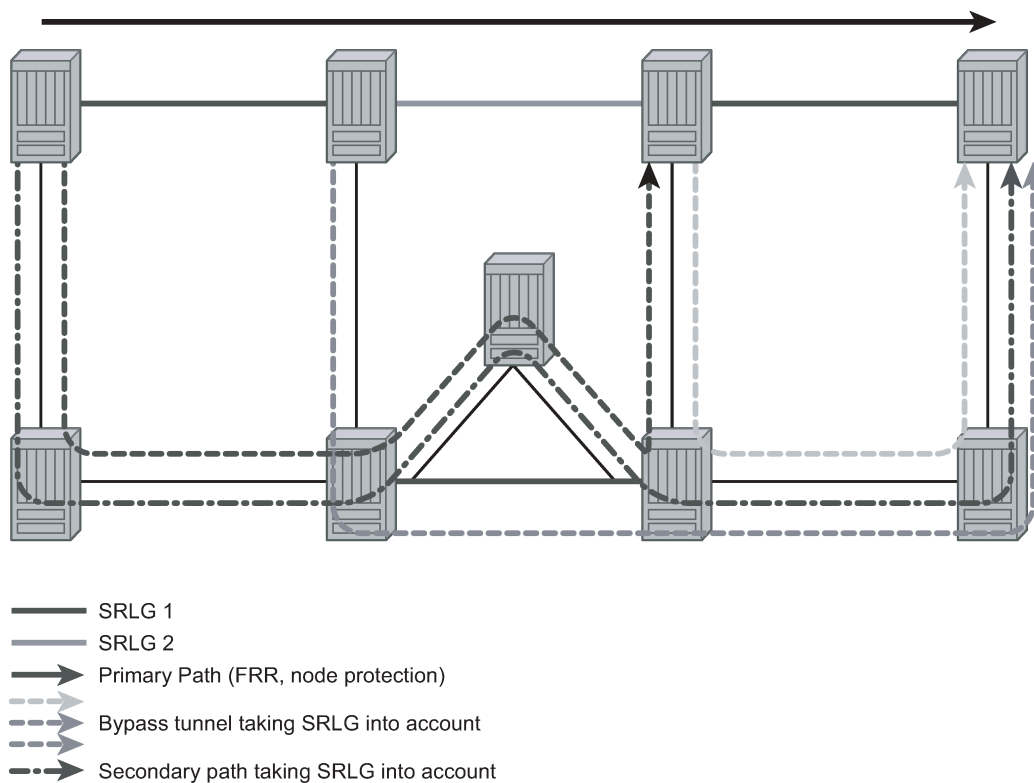
Fig_33

This feature is supported on OSPF and IS-IS interfaces on which RSVP is enabled.

2.4.9.2 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 specific SRLG.

Figure 14: SRLG penalty weight operation illustrates the operation of SRLG penalty weights.

Figure 14: SRLG penalty weight operation
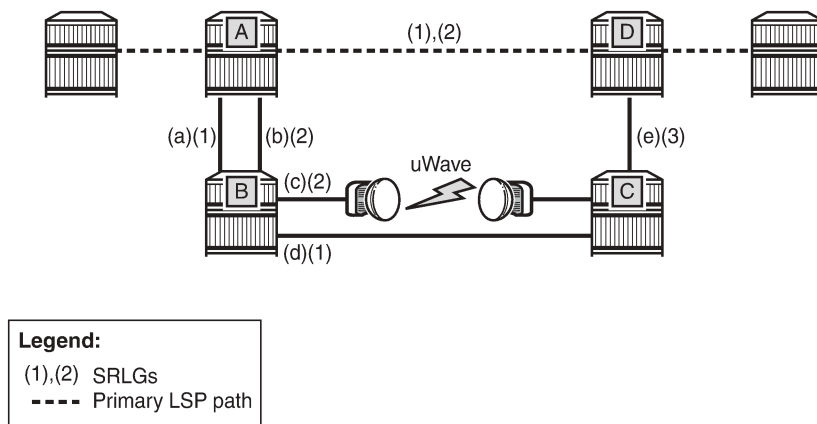
The primary LSP path includes a link between A and D with SRLG (1) and (2). The bypass around this link through nodes B and C includes links (a) and (d), which are members of SRLG (1), and links (b) and (c), which are members of SRLG 2. If the link metrics are equal, then this gives four ECMP paths from 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 because 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.

Use the following command option to configure a penalty weight value to associate with an SRLG group.

- **MD-CLI**

```
configure routing-options if-attribute srlg-group penalty-weight
```

- **classic CLI**

```
configure router if-attribute srlg-group value penalty-weight
```

If an SRLG penalty weight is configured, CSPF includes 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: 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** command is not configured to **strict**.

2.4.9.3 Static configurations of SRLG memberships

This feature allows users to manually enter the link members of SRLG groups for the entire network on any SR OS router that needs to signal LSP paths (for example, a head-end node).

The user may explicitly enable the use by CSPF of the SRLG database. In that case, CSPF does not query the TE database for IGP advertised interface SRLG information.

Note, however, that the SRLG secondary path computation and FRR bypass or detour path computation remains unchanged.

There are deployments where SR OS interoperates 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 manually enter the link members of SRLG groups for the entire network on any SR OS router that needs to signal LSP paths, for example, a head-end node.

Use the following command to enter the SRLG membership information for any link in the network.

```
configure router mpls srlg-database router-id interface
```

An interface can be associated with up to five 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.

Use the following command to delete a specific interface entry in the database:

- **MD-CLI**

```
configure router mpls srlg-database router-id delete interface
```

- **classic CLI**

```
configure router mpls srlg-database router-id no interface srlg-group
```

The SLRG group must already be configured using the **srlg-group** command. The maximum number of distinct SRLG groups the user can configure on the system is 1024.

The value for the **router-id** command must correspond to the router ID configured under the base router instance, the base OSPF instance or the base IS-IS instance of a specific node. Note however that a single user SRLG database is maintained per node, regardless of whether the listed interfaces participate in static

routing, OSPF, IS-IS, or both routing protocols. Use the following command to temporarily disable CSPF from using interface membership information for a specific router ID:

- **MD-CLI**

```
configure router mpls srlg-database router-id admin-state disable
```

- **classic CLI**

```
configure router mpls srlg-database router-id shutdown
```

When the preceding command is used, CSPF assumes these interfaces have no SRLG membership association.

Use the following command to delete all interface entries for a specific router ID in the database:

- **MD-CLI**

```
configure router mpls srlg-database delete router-id
```

- **classic CLI**

```
configure router mpls srlg-database no router-id
```

CSPF does 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 does not query the TE database for IGP advertised interface SRLG information.

Use the following command option to allow CSPF to query the SRLG database:

- **MD-CLI**

```
configure router mpls user-srlg-db true
```

- **classic CLI**

```
configure router mpls user-srlg-db enable
```

When the MPLS module makes a request to CSPF for the computation of an SRLG secondary path, CSPF queries 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.

Use the following command to disable CSPF from querying the SRLG database:

- **MD-CLI**

```
configure router mpls user-srlg-db false
```

- **classic CLI**

```
configure router mpls user-srlg-db disable
```

When the preceding command is used, CSPF resumes querying the TE database for SRLG membership information. However, the user SRLG database is maintained.

Use the following command to delete the entire SRLG database:

- **MD-CLI**

```
configure router mpls delete srlg-database
```

- **classic CLI**

```
configure router mpls no srlg-database
```

When the SRLG database is deleted, CSPF assumes all interfaces have no SRLG membership association if the user has not disabled the use of this database.

2.4.10 TE graceful shutdown

Graceful shutdown provides a method to bulk re-route transit LSPs away from the node during a software upgrade of a node. A solution is described in RFC 5817, *Graceful Shutdown in MPLS and Generalized MPLS Traffic Engineering Networks*. The solution is achieved by using a PathErr message with a Local Maintenance error code on a flag required by a TE link. When a LER receives this message, it performs a make-before-break on the LSP path to move the LSP away from the links/nodes on 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 can 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 traffic engineering metric set to 0xffffffff and 0 as unreserved bandwidth.

2.4.11 Soft preemption of DiffServ RSVP LSP

A DiffServ 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.

2.4.12 Least-fill bandwidth rule in CSPF ECMP selection

When multiples equal-cost paths satisfy the constraints of a specific RSVP LSP path, CSPF in the router head-end node selects a path so that LSP bandwidth is balanced across the network links. In releases before 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 applies this method to all path calculations of the LSP, also at the time of the initial configuration.

2.4.13 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.

2.4.13.1 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 does 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 multisegment intra-area or inter-area LSP path in one operation.

Figure 15: Automatic ABR node selection for inter-area LSP illustrates the role of each node in the signaling of an inter-area LSP with automatic ABR node selection.

Figure 15: Automatic ABR node selection for inter-area LSP
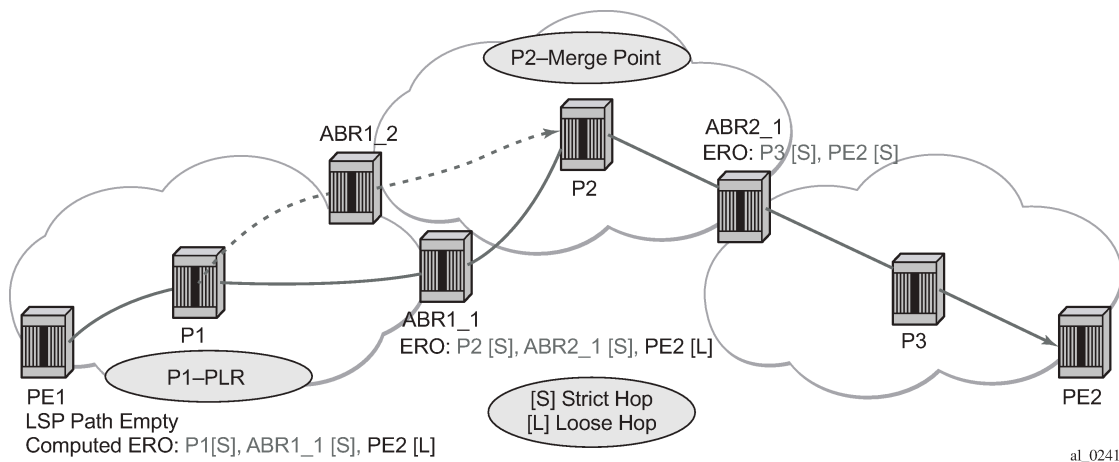
CSPF for an inter-area LSP operates as follows:

- 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.

- When the path of the LSP is empty, CSPF computes a single-segment intra-area path to an ABR node that advertised a prefix matching with the destination address of the LSP.

- When the path of the LSP contains one or more hops, CSPF computes a multisegment intra-area path including the hops that are in the area of the Ingress LER node.

- 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.

- 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.

- Note the following special case of a multisegment 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 16: CSPF for an inter-area LSP. 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 further expands the ERO by including link D-E as part of the path in area 0.

Figure 16: CSPF for an inter-area LSP
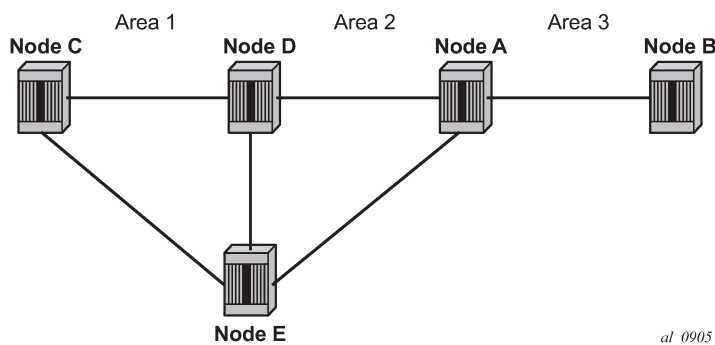
- If there is more than 1 ABR that advertised a prefix, CSPF calculates a path for all ABRs. Only the shortest path is withheld. If more than one path has the shortest path, CSPF picks a path randomly or based on the least-fill criterion if enabled. If more than one ABR satisfies the least-fill criterion, CSPF also picks one path randomly.

- The path for an intra-area LSP path is not 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.

2.4.13.1.1 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 remains 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 retries 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 receives 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 then performs 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 performs 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.

2.4.13.1.2 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 command options. There is, however, a number of enhancements introduced by the automatic ABR selection feature, as described in the following:

- Features such as path bandwidth reservation and admin groups continue to operate within the scope of all areas because they rely on propagating the command option information in the Path message across the area boundary.

- The TE graceful shutdown and soft preemption features continues 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 is not able to avoid the link or node when it performs the MBB because 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 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 indeed avoids 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 **te** command option for the **metric-type** command in CSPF cannot be propagated across the area boundary and operates within the scope of the local area of the ingress LER node. This behavior is as a result of the automatic ABR selection feature.

  Use the following command to configure the metric type used for the LSP path computation.

  ```
  configure router mpls lsp metric-type
  ```

- The **srlg** command for bypass LSP continues to operate locally at each PLR within each area. The PLR node protecting the ABR checks the SRLG constraint for the path of the bypass within the local area. Use the following command to enable the SRLG constraint:

  - **MD-CLI**

    ```
    configure router mpls lsp secondary srlg true
    ```

– **classic CLI**

```
configure router mpls lsp secondary srlg
```

- The **srlg** command 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.

- Support for the **least-fill** command with an inter-area LSP is introduced with the automatic ABR selection feature. Use the following command to enable the least-fill path selection method:

    – **MD-CLI**

    ```
    configure router mpls lsp least-fill true
    ```

    – **classic CLI**

    ```
    configure router mpls lsp least-fill
    ```

    When this command is configured, CSPF applies the least-fill criterion to select the path segment to the exit ABR node in the local area.

- The 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** command under the LSP still needs to be enabled on the inter-area LSP, if the user needs to have admin groups propagated across the areas. Use the following command to enable the propagation of admin groups:

    – **MD-CLI**

    ```
    configure router mpls lsp propagate-admin-group true
    ```

    – **classic CLI**

    ```
    configure router mpls lsp propagate-admin-group
    ```

- With the automatic ABR selection feature, timer based re-signal of the inter-area LSP path is supported and re-signals 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.

2.4.13.2 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 because 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 because 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 automatically computes and signals 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 computes and signals an intra-area LSP using the transit area TE links, even when the destination router-id is only part of area 0.

2.4.13.3 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 correct 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 17: ABR node protection using dynamic bypass LSP illustrates the role of each node in the ABR node protection using a dynamic bypass LSP.

Figure 17: ABR node protection using dynamic bypass LSP
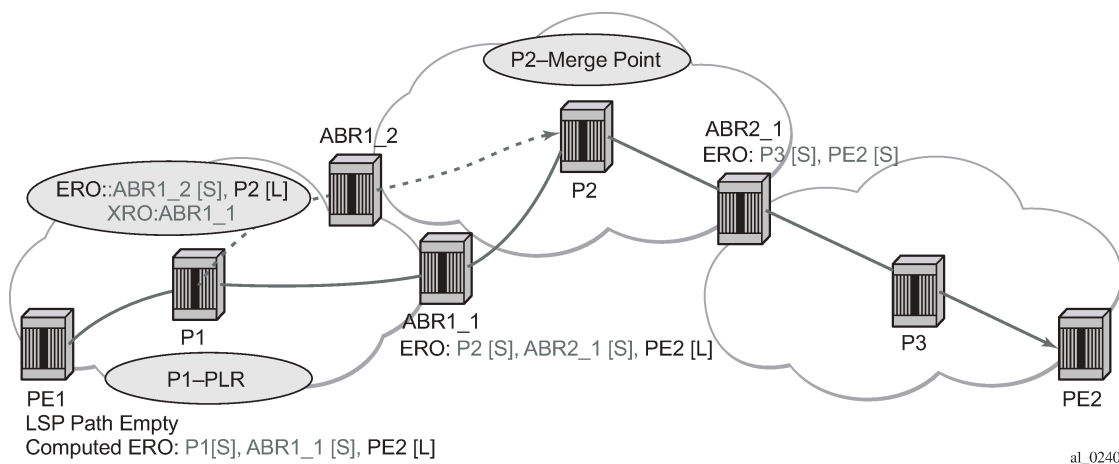
In order for a PLR node within the local area of the ingress LER to provide ABR node protection, it must dynamically signal a bypass LSP and associate it with the primary path of the inter-area LSP using the following new procedures:

- 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 automatically includes 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 only works if the node-id RRO also contains the interface address. Otherwise the neighbor is not created when the bypass is activated by the PLR node. The Path state then times 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 does not signal a one-to-one detour LSP for ABR protection. In addition, an ABR node rejects 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** command option is enabled or not on the node. In other words, the router as a transit ABR for the detour path rejects the signaling of an inter-area detour backup LSP.

2.4.14 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, instead of 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 fallback timer-based mechanism is required, in case a user fails to execute the switch back to the primary path. This function is also useful to stagger reversion for large numbers of LSPs.

Use the following command to configure a reversion timer for an LSP.

```
configure router mpls lsp revert-timer
```

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). After a timer has started, it can be modified using the same **revert-timer** command. If a new value is entered, 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 datapath.

Use the following command to cancel any currently outstanding revert timer and allow the LSP to revert to the primary path, if it is up:

- **MD-CLI**

```
configure router mpls lsp delete revert-timer
```

- **classic CLI**

```
configure router mpls lsp no revert-timer
```

If the LSP secondary path fails while the revert timer is still running, the system cancels the revert-timer and the LSP reverts to the primary path immediately. Use the following command to manually force an LSP to revert to the primary path while the revert timer is still running.

```
tools perform router mpls revert lsp
```

The preceding command forces the early expiry of the revert timer for the LSP. The primary path must be up, for this command to work.

2.4.15 LSP tagging and autobind using tag information

RSVP and SR-TE LSPs can be configured with an administrative tag.

LSP tagging enables 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:

- Associating a color with an exported BGP route. This configuration is signaled using the BGP Color Extended Community described in Section 4.3 of *draft-ietf-idr-tunnel-encaps-03*. This configuration 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.

- Defining 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.

- Configuring 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.

- Applying one or more administrative tags to include or exclude as an action to a matching route in a BGP route policy. Different administrative 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 administrative tags associated with their VPN routes via matching on RT extended community values.

- Matching 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 autobind-tunnel

  - EVPN with autobind-tunnel

The following is an overview of how the feature is intended to operate:

1. Configure a nodal database of administrative tags.

   Use the following command to configure administrative tags:

© 2025 Nokia.
Use subject to Terms available at: www.nokia.com/terms.

- **MD-CLI**

```
configure routing-options admin-tags admin-tag
```

- **classic CLI**

```
configure router admin-tags admin-tag
```

> **Note:** Each tag is automatically assigned an internal color.

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 route administrative tag policy containing a list of administrative tags to include or exclude.

   Use the following command to configure a route administrative tag policy:

   - **MD-CLI**

   ```
   configure routing-options admin-tags route-admin-tag-policy
   ```

   - **classic CLI**

   ```
   configure router admin-tags route-admin-tag-policy
   ```

   > **Note:** Up to eight include and exclude statements are supported per policy.

4. Configure a **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 are given the associated internal color.

5. Configure administrative tags on RSVP or SR-TE LSPs, so that different groups of LSPs can be treated differently by applications that intend to use them.

   > **Note:** More than one administrative tag can be configured against a specified LSP.

6. Apply a route policy to a service or other object as an import policy.

   The system 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, use the following respective commands to restrict the resolution filter to only LSPs matching the pattern of administrative tags in the **route-admin-tag-policy**:

   - **MD-CLI**

   ```
   configure router bgp next-hop-resolution labeled-routes transport-tunnel family enforce-
   strict-tunnel-tagging true
   ```

```
configure router bgp next-hop-resolution shortcut-tunnel family enforce-strict-tunnel-
tagging true
```

- **classic CLI**

```
configure router bgp next-hop-resolution labeled-routes transport-tunnel family enforce-
strict-tunnel-tagging
configure router bgp next-hop-resolution shortcut-tunnel family enforce-strict-tunnel-
tagging
```

If the **enforce-strict-tunnel-tagging** command is not enabled, the router falls back to untagged LSPs. The tunnels that VPRN and EVPN services can autobind to can also be restricted using the **enforce-strict-tunnel-tagging** command in the **auto-bind-tunnel** configuration for the service. The following subsections provide more details about how the matching algorithm works.

2.4.15.1 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. See LSP administrative tag use in tunnel selection for VPRN and E-VPN autobind and LSP administrative tag use for BGP next hop or BGP prefix for labeled and unlabeled unicast routes for more information about how the algorithm applies to specific use cases.

Internal color matching occurs before any resolution filter is applied.

The following selection process steps assume that 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 resulting set of colored LSPs match, the default behavior is to fall back to using uncolored LSPs, if they exist. If they do not exist, the route does not resolve. Depending on the context, the fall-back method can be disabled on a per-service basis, as described in LSP administrative tag use in tunnel selection for VPRN and E-VPN autobind and LSP administrative tag use for BGP next hop or BGP prefix for labeled and unlabeled unicast routes .

3. If an administrative tag policy is not configured for a route, it is assumed that the user does not want to express a preference for the LSP to use. Therefore, routes with no administrative 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.

2.4.15.2 LSP administrative tag use in tunnel selection for VPRN and E-VPN autobind

For VPRN, EVPN-VPLS, and EVPN-VPWS, routes may be imported using the following methods used for the autobind-tunnel:

- peer route import policies that contain route administrative tag policies

- VRF import policies for VPRN

- VSI import policies for E-VPN VPLS

- VSI import policies for E-VPN VPLS or Epipe

VRF import and VSI import policies take precedence over the peer route import policy.

For policies that contain route administrative 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, this is applied across the resulting set.

Routes with no administrative tag, or a tag that is not explicitly excluded by the route administrative 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 LSPs exist. By default, the system falls back to regular autobind behavior using LDP, SR-ISIS, SR-OSPF, or any other lower priority configured tunnel type; otherwise, the resolution fails. That is, matching administrative 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 the way that only specific tagged or untagged tunnels are considered, using the **enforce-strict-tunnel-tagging** and **enforce-untagged-route** commands in the **auto-bind-tunnel** context, as follows:

- If **enforce-strict-tunnel-tagging** is configured and **enforce-untagged-route** is set to **none** or is not configured, only tagged LSPs are considered for routes that have an administrative tag policy applied. If there is no match, the resolution fails. Routes with no administrative tag policy applied can still resolve using either tagged or untagged LSPs.

- If **enforce-untagged-route untagged-tunnel** is configured, routes with no administrative tag policy applied can only resolve using untagged LSPs, and if no such LSPs exist, resolution of those routes fails.

2.4.15.3 LSP administrative 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 administrative tag policies. For route import policies that contain route administrative 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, this is applied across the resulting set.

Routes with no administrative 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 administrative-tagged RSVP or SR-TE LSPs are preferred to other LSP types, whether tagged or untagged. On a per-address family basis, it is possible to enforce the way that tagged or untagged tunnels are considered for BGP labeled routes or shortcut tunnels using the **enforce-strict-tunnel-tagging** and **enforce-untagged-route** commands that follow.

If **enforce-strict-tunnel-tagging** is configured and **enforce-untagged-route** is set to **none** or is not configured, only tagged LSPs are considered for routes that have an administrative tag policy applied. If there is no match, the resolution fails. Routes with no administrative tag policy applied can still resolve using either tagged or untagged LSPs.

If **enforce-untagged-route untagged-tunnel** is configured, routes with no administrative tag policy applied can only resolve using untagged LSPs, and if no such LSPs exist, resolution of those routes fails.

Use the following commands to configure how tagged or untagged tunnels are considered for BGP labeled routes or shortcut tunnels:

- **MD-CLI**

```
configure router bgp next-hop-resolution labeled-routes transport-tunnel family enforce-
strict-tunnel-tagging true
configure router bgp next-hop-resolution shortcut-tunnel family enforce-strict-tunnel-
tagging true
```

- **classic CLI**

```
configure router bgp next-hop-resolution labeled-routes transport-tunnel family enforce-
strict-tunnel-tagging
configure router bgp next-hop-resolution shortcut-tunnel family enforce-strict-tunnel-
tagging
```

2.4.16 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 datapath 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.

An SR OS router can use LSP Self-ping to test that the datapath 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 datapath of the LSP must have been programmed. LSP Self-ping is used instead of the LSP hold timer, which is configured using the following command.

```
configure router mpls hold-timer
```

LSP Self-ping is particularly useful in multivendor networks, where specific nodes may take unexpectedly long times to program their datapath.

LSP BFD is not supported if LSP Self-ping is enabled. The router ignores the LSP Self-ping configuration, if the following command option is configured for an LSP.

```
configure router mpls lsp bfd failure-action failover-or-down
```

Use the commands in the following context to configure LSP Self-ping.

```
configure router mpls lsp-self-ping
```

Use the following command option to enable LSP Self-ping for all RSVP-TE LSPs:

- **MD-CLI**

```
configure router mpls lsp-self-ping rsvp-te true
```

- **classic CLI**

```
configure router mpls lsp-self-ping rsvp-te enable
```

> **Note:** 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.

Use the following command to set the interval, in seconds, that periodic LSP Self-ping packets are sent.

```
configure router mpls lsp-self-ping interval
```

Use the following command to configure a timer that is started when the first LSP Self-ping packet for a specific event is sent on an LSP path.

```
configure router mpls lsp-self-ping timeout
```

Use the following command to specify which action is taken, if no LSP Self-ping reply is received before the timer expires.

```
configure router mpls lsp-self-ping timeout-action
```

If the **timeout-action** command is configured to **retry**, the router tries to signal a new path and the process repeats (see Detailed behavior of LSP Self-ping for more information). If the **timeout-action** command is configured to **switch**, the router uses the new path regardless and stops the LSP Self-ping cycle.

By default, LSPs and LSP templates inherit the configuration at the MPLS level. However, LSP Self-ping can be explicitly enabled or disabled for a specific LSP or LSP template.

Use the following command option to enable LSP Self-ping for an LSP:

- **MD-CLI**

```
configure router mpls lsp lsp-self-ping true
```

- **classic CLI**

```
configure router mpls lsp lsp-self-ping enable
```

Use the following command option to enable LSP Self-ping for an LSP template:

- **MD-CLI**

```
configure router mpls lsp-template lsp-self-ping true
```

- **classic CLI**

```
configure router mpls lsp-template lsp-self-ping enable
```

2.4.16.1 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

> **Note:** This happens if the path preference changes or if the following command is used.
>
> ```
> tools perform router mpls switch-path
> ```

- degraded standby or secondary path → degraded primary path (degraded primary is preferred to degraded standby or secondary path)
- MBB on active path

A path can go to a degraded state either because of 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 then periodically sends 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** command is configured to **retry**, the system tears down the candidate path (in the case of switching between paths) and goes back to CSPF for a new path. The system then starts 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 does 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 completes, and the new value is not used until the next LSP Self-ping session.

If no timeout is configured, the default value is used.

2.4.16.2 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, the system continues with the 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 are 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. Use the following commands to display LSP Self-ping and OAM resource exhaustion timeout values.

```
show router mpls lsp detail
show router mpls lsp-self-ping
```

2.4.17 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 the 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.

Use the following steps to configure dark bandwidth accounting:

1. Use the following command option to enable the collection of statistics for dark bandwidth:

   • **MD-CLI**

   ```
   configure router mpls aux-stats sr true
   ```

   • **classic CLI**

   ```
   configure router mpls aux-stats sr
   ```

   > **Note:** The **aux-stats** command can only be configured to consider MPLS-SR traffic as contributing to dark bandwidth.

2. Use the following command to enable dark bandwidth accounting on each SE.

   ```
   configure router rsvp dbw-accounting
   ```

   > **Note:** After dark bandwidth has been enabled, auxiliary statistics collection cannot be disabled. Dark bandwidth accounting must be disabled before auxiliary statistics collection can be disabled. Use the following command to disable dark bandwidth accounting:
   >
   > • **MD-CLI**
   >
   > ```
   > configure router rsvp delete dbw-accounting
   > ```
   >
   > • **classic CLI**
   >
   > ```
   > configure router rsvp no dbw-accounting
   > ```

3. Configure the dark bandwidth accounting command options as needed 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 commands are generally applied globally, but it is possible to override the global value of some commands on a per-interface basis.

Use the following command to display (on a global or per-interface basis) key values associated with the dark bandwidth accounting process.

```
show router rsvp status
```

2.5 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 in which the 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.

2.5.1 Application in video broadcast

Figure 18: Application of P2MP LSP in video broadcast shows a triple play application (TPSDA).

Figure 18: Application of P2MP LSP in video broadcast
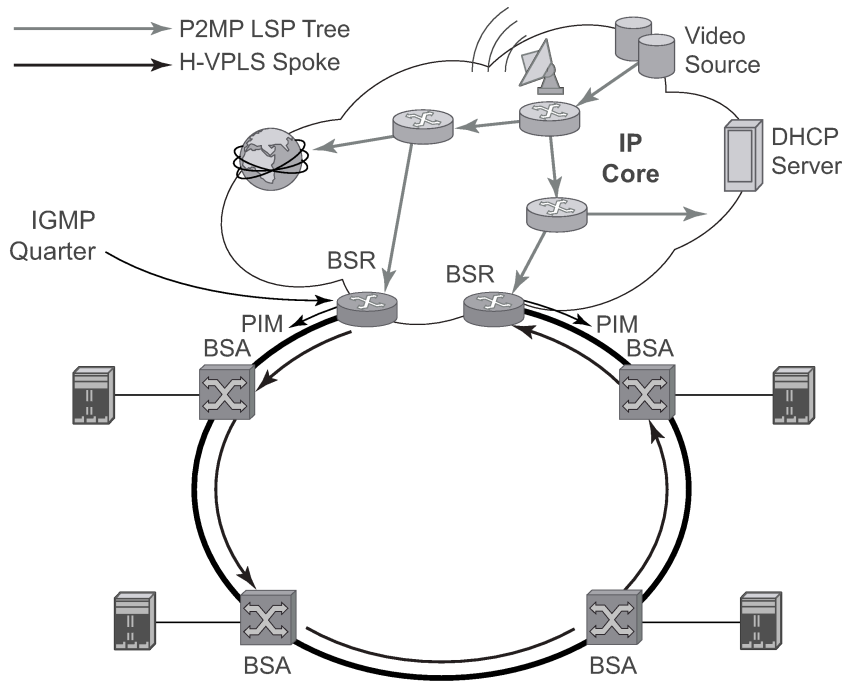
A PIM-free core network can be achieved by deploying P2MP LSPs using other core routers. The router can act as the ingress LER receiving the multicast packets from the multicast source and forwarding them over the P2MP LSP.

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.

2.5.2 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 or at a transit LSR, or both, 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.

2.5.2.1 Ingress LER node

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 has 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 replicated packet.

2.5.2.2 LSR node

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.

For control packets received on an ILM in an LSR, packets arriving with an expired TTL in the outer label are sent to the CPM for further processing and are not forwarded to the egress NHLFE.

2.5.2.3 Branch LSR node

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.

2.5.2.4 Egress LER node

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.

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.

2.5.2.5 BUD LSR node

At an LSR node that is both a branch node and an egress leaf node (bud node), a pop operation is performed 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. 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 that 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 any of the IP header exception handling conditions are set after the label is popped: 127/8 destination address, router alert option set, or any other options set.

2.5.3 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, IOM/IMM (on the 7705 SAR Gen 2) ingress datapaths 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 determines 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, 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 removes 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.

2.5.3.1 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 have a set of per forwarding class multicast or broadcast queues, or both, 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 or when a P2MP LSP ILM exists on the ingress IOM/IMM, the remaining multicast paths are also enabled. 16 multicast paths are supported by default. One path remains as a secondary path and the rest are primary paths.

Use the following command to configure the fabric speed for the chassis.

```
tools perform system set-fabric-speed
```

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. Use the commands in the following context to enable IMPM on the ingress XMA datapath or the FP datapath.

```
configure card fp ingress mcast-path-management
```

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 causes routed and VPLS (IGMP and PIM) snooped IP multicast groups to be managed by IMPM.

The above ingress datapath 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.

2.5.4 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 the P2MP ID, tunnel ID, and extended tunnel ID part of the P2MP session object, and the tunnel sender address and 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 forwards the path message of each S2L sub-LSP to the downstream LSR without replicating it. It also forwards 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 drops aggregated RSVP messages on the receive side if originated by another vendor's implementation.

- Use the following command option to configure a P2MP LSP:

  - **MD-CLI**

    ```
    configure router mpls lsp type p2mp-rsvp
    ```

  - **classic CLI**

    ```
    configure router mpls lsp p2mp-lsp
    ```

  Next, use the following command to create a primary P2MP instance.

  ```
  configure router mpls lsp primary-p2mp-instance
  ```

  After creating the primary P2MP instance, use the command options in the following context to add a S2L sub-LSP path to the P2MP instance.

  ```
  configure router mpls lsp primary-p2mp-instance s2l-path
  ```

  > **Note:** The paths can be empty paths or can specify a list of explicit hops. The path name must exist and must have been defined, using the following command.
  >
  > ```
  > configure router mpls path
  > ```

- The same path name can be reused by more than one S2L of the primary P2MP instance. However the **to** command 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 specific time.

- The following command options can be used with a P2MP LSP.

```
configure router mpls lsp adaptive
configure router mpls lsp path-computation-method local-cspf
configure router mpls lsp fast-reroute
configure router mpls lsp from
configure router mpls lsp hop-limit
configure router mpls lsp metric
configure router mpls lsp retry-limit
configure router mpls lsp retry-timer
configure router mpls resignal-timer
```

In the MD-CLI, the following additional commands can be used with a P2MP LSP.

```
configure router mpls lsp exclude-admin-group
configure router mpls lsp include-admin-group
```

In the classic CLI, the following additional commands can be used with a P2MP LSP.

```
configure router mpls lsp exclude
configure router mpls lsp include
```

- The following command options cannot be used with a P2MP LSP.

```
configure router mpls lsp adspec
configure router mpls lsp primary
configure router mpls lsp secondary
configure router mpls lsp to
```

- The node ingress LER does 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** command is not available at the LSP level. It is, however, available at the path level of each S2L sub-LSP of the primary or secondary instance of this P2MP LSP.

- Use the following command to configure a hold timer that applies when signaling or resignaling an individual S2L sub-LSP path.

```
configure router mpls hold-timer
```

It does not apply when the entire tree is signaled or resignaled.

- The head-end node can add or remove, or both, 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 resignaling 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 of the following other configuration changes result in tearing down and retrying all affected S2L paths:

  – enabling or disabling the **adaptive** command

  – configuring the **metric-type** command to **te**

  – disabling **fast-reroute**

  – disabling or changing the configuration of the **path-computation-method** command

- MPLS requests CSPF to re-compute the whole set of S2L paths of a specific active P2MP instance each time the P2MP resignal timer expires. The P2MP resignal 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 requests CSPF to recompute the whole set of S2L paths of a specific active P2MP instance each time the user performs a manual resignal 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. Use the following command to perform a manual resignal of the P2MP LSP instance.

  ```
  tools perform router mpls resignal p2mp-lsp p2mp-instance
  ```

- 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 retries the S2L using the retry timer and retry count values inherited from P2MP LSP configuration. However, there is a global MBB timer set to 600 seconds, which is not configurable. If the global MBB succeeds, and for example, all S2L MBBs have succeeded, before the global timer expires, MPLS moves 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 continues until all S2Ls have retried once and then aborts the global MBB. After global MBB is aborted, MPLS moves 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 continues forwarding packets on S2L sub-LSP paths which are not being reoptimized and on the older S2L sub-LSP paths for which make-before-break operation was not successful. MBB therefore results in duplication of packets until the old path is torn down.

- The MPLS datapath 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 only makes 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

  - DiffServ TE

  - Soft preemption of RSVP P2MP LSP

2.5.5 P2MP RSVP-TE preemption behavior

P2MP S2Ls can be preempted by or preempt other P2P or P2MP LSPs. SR OS supports P2MP S2L soft preemption and hard preemption, as described in Soft preemption and Hard preemption.

The P2P LSPs and P2MP S2L LSPs reserve the bandwidth they require throughout the network. The P2P and P2MP LSPs only compete for bandwidth if a link failure occurs and a single link is overloaded by additional P2P and S2L LSPs.

The user can configure the setup and hold priority on all LSP types to ensure the higher priority LSPs (P2P or S2L) can preempt lower priority LSPs.

The following figure shows a network example with three LSP types:

- gold P2P LSP, with setup 1 and hold 1

- silver P2MP LSP, with setup 3 and hold 3

- bronze P2P LSP, with setup 7 and hold 7

Figure 19: LSP priority example
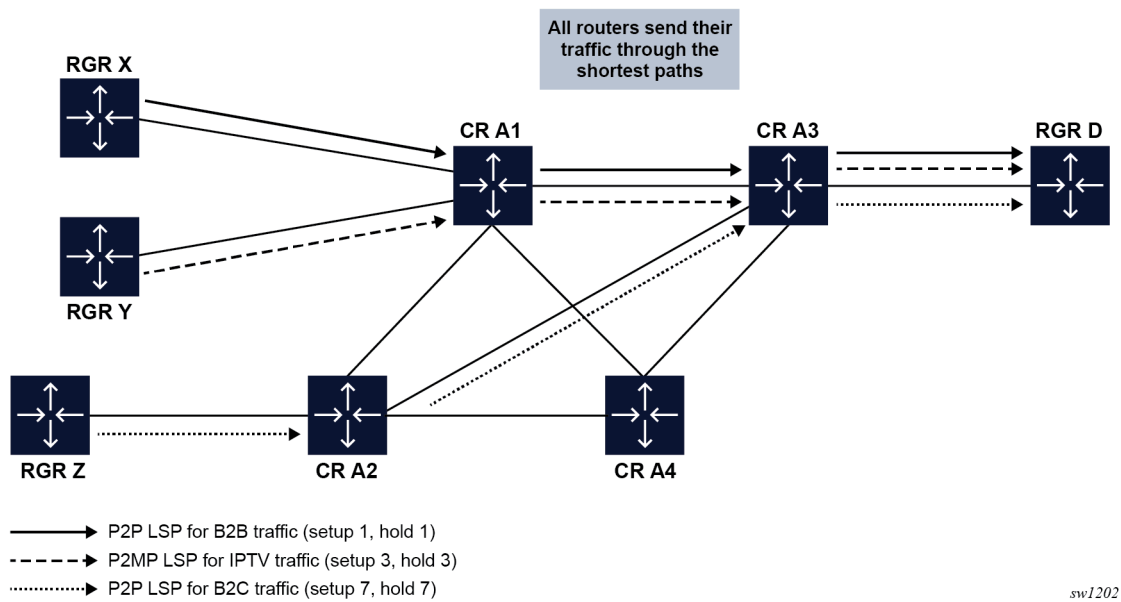
If a link failure occurs between CR A1 and CR A3, as shown in the following figure, the following occurs:

- The gold P2P LSPs take the shortest path and preempts all other P2P and P2MP LSPs that have lower hold values.

- The silver P2MP LSPs (S2L) take the second shortest path because there is insufficient bandwidth on the shortest path after the gold P2P LSPs reserve all bandwidth. The bronze LSPs on this path are preempted.

- The bronze P2P LSPs take the longest path because there is insufficient bandwidth on the other two paths. They cannot preempt any other LSPs and can always be preempted by other LSPs that have higher priority values.

Figure 20: Link failure example
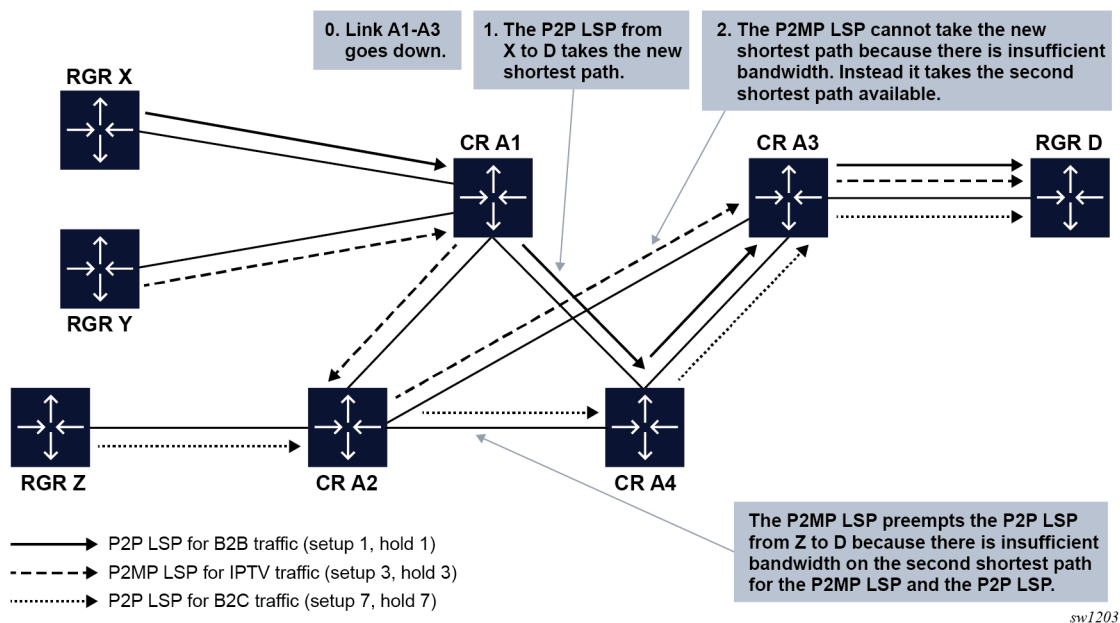
2.5.5.1 Soft preemption

When soft preemption is enabled for P2MP S2L LSPs, the S2L preemption is governed by the timer value configured using the following command.

```
configure router rsvp preemption-timer
```

When the S2L is preempted at an LSR node, the preempting node sends to the head-end node an 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). The preemption timer (configured as described in the preceding paragraph) starts. When the timer expires, the node performs MBB on the affected adaptive CSPF LSP. Both IGP metric and TE metric-based CSPF LSPs are included. If an alternative path that excludes the flagged interface is not found, the LSP is placed on a retry list in a similar way to the global revertive procedure at a head-end node.

When the preemption timer expires, the preempting node tears down the S2L and sends a path error to the head-end node, and the head-end node places the S2L on the retry list.

2.5.5.2 Hard preemption

When soft preemption is not enabled for the P2MP LSP, the value of the preemption timer is hard coded as 0 for S2L LSPs. That is, S2Ls are hard preempted by higher priority LSPs. The MBB operation is not performed at the root of the preempted S2L. That is, the S2L is immediately torn down to the root PE when it is preempted and must signal again.

When an S2L is preempted, it sends an RESVTEAR message to the root PE (head-end) indicating the lack of resources, and then the S2L is torn down throughout the network. After the P2MP S2L retry or fast retry

timer expires, the S2L signals again by sending a new PATH message from the headend, based on the newly calculated Constraint-based Shortest Path First (CSPF) path where the bandwidth resources are available.

S2Ls can preempt other P2P LSPs or other S2Ls based on the hold and setup priority.

2.5.6 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.

2.5.6.1 Procedures at ingress LER node

Perform the following steps to forward multicast packets over a P2MP LSP:

1. Use the following command option to create a tunnel interface associated with the P2MP LSP:
   - **MD-CLI**

     ```
     configure router tunnel-interface rsvp-p2mp-root
     ```

   - **classic CLI**

     ```
     configure router tunnel-interface rsvp-p2mp
     ```

2. Use the following command to add static multicast group joins to the PIM interface as a specific <S,G>:
   - **MD-CLI**

     ```
     configure router igmp tunnel-interface rsvp-p2mp-root static group source
     ```

   - **classic CLI**

     ```
     configure router igmp tunnel-interface static group source
     ```

   Use the following command to add static multicast group joins to the PIM interface as a <*,G>:
   - **MD-CLI**

     ```
     configure router igmp tunnel-interface ldp-p2mp-root static group starg
     ```

   - **classic CLI**

     ```
     configure router igmp tunnel-interface 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 actually passes to PIM a more structured tunnel interface identifier. The structure follows 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 specific <*,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 are also 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.

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 needs to be combined into a superset MCID.

2.5.6.2 Procedures at egress LER node

2.5.6.2.1 Procedures with a primary tunnel interface

Use the following command to configure a tunnel interface and associate it with a terminating P2MP LSP leaf:

- **MD-CLI**

```
configure router tunnel-interface rsvp-p2mp-leaf sender-address
```

- **classic CLI**

```
configure router tunnel-interface rsvp-p2mp sender
```

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 actually passes to PIM a more structured tunnel interface identifier. The structure follows 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:

- the regular RPF check on unlabeled IP multicast packets, which is based on routing table lookup

- 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 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 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 does 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 toward 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.

2.6 Pipe mode support for RSVP-TE MPLS trees

RSVP-TE P2MP LSPs can operate in uniform mode or pipe mode.

In uniform mode (default behavior), the multicast packet TTL value is copied to the P2MP LSP EXP field on the ingress label edge router (iLER). The MPLS TTL value is copied to the multicast PDU TTL on the egress label edge router (eLER), .

In pipe mode for P2MP LSPs, the iLER and LSR set the EXP value of the P2MP LSP header to 255 and the multicast PDU TTL value is not propagated to the MPLS header TTL. On the eLER, the behavior is the same as unicast, that is, the multicast PDU TTL = MIN{transport label stack TTL-1, service packet TTL-1}.

Use the following command to configure the pipe mode for iLER and eLER.

```
configure router mpls p2mp-ttl-propagate
```

The iLER and LSR default behavior is uniform mode.

2.6.1 Switching between uniform and pipe modes

When the following command is modified, the new TTL mode applies to future P2MP LSPs only.

```
configure router mpls p2mp-ttl-propagate
```

The existing operational LSPs are not affected. If a new S2L is added to an existing P2MP tree at any node, the new S2L uses the same TTL mode as the rest of the tree. For the new configuration to take effect, the user must manually resignal the P2MP LSPs from the iLER. S2Ls cannot be resignaled from the eLER.

Use the following command at the iLER to resignal the specified P2MP LSP using Make-Before-Break (MBB).

```
tools perform router mpls resignal p2mp-lsp p2mp-instance
```

Use the following command to make the P2MP resignal timer expire faster.

```
tools perform router mpls resignal p2mp-delay
```

Use the following command at the iLER to bounce the specified P2MP LSP.

```
clear router mpls lsp
```

When the **p2mp-ttl-propagate** configuration changes, an information message is displayed in the classic CLI indicating that the P2MP LSPs must be bounced for the change to take effect. In the MD-CLI, this information message is not supported currently.

2.7 MPLS service usage

Nokia routers enable service providers to deliver VPNs and Internet access using Generic Routing Encapsulation (GRE) or MPLS tunnels, or both, with Ethernet interfaces or SONET/SDH interfaces, or both.

2.7.1 Service distribution paths

A service distribution path (SDP) acts as a logical way of directing traffic from one router to another through a unidirectional (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, see "Service Distribution Paths (SDPs)" in the *7705 SAR Gen 2 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.

2.8 MPLS/RSVP configuration process overview

The following figure shows the process to configure MPLS and RSVP command options.

Figure 21: 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    │
└──────────┘
```
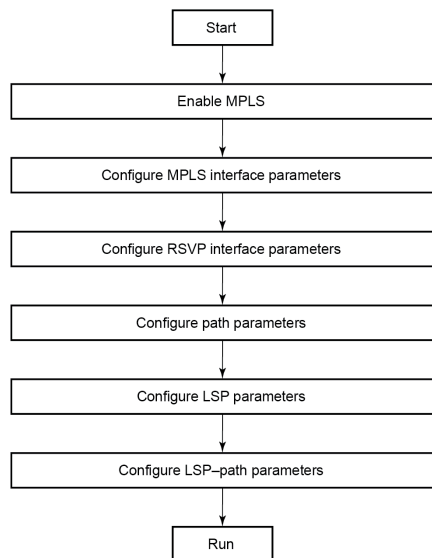
al_0212

2.9 Configuration notes

The following are MPLS and RSVP restrictions:

- Interfaces must already be configured, before they can be specified in MPLS and RSVP. Use the commands in the following context to configure an interface.

```
configure router interface
```

- A router interface must be specified in the **mpls** context to apply it or modify command options in the following context.

```
configure router rsvp
```

- A system interface must be configured and specified. Use the commands in the following context to configure MPLS protocol support for a system interface.

```
configure router mpls interface
```

- Paths must be created before they can be applied to an LSP.

2.10 Configuring MPLS and RSVP with CLI

This section provides information to configure MPLS and RSVP using the command line interface.

2.10.1 MPLS configuration overview

MPLS enables routers to forward traffic based on a simple label embedded into the packet header. The 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.

To implement MPLS, the user must first configure at least one LSP, path, and router interface.

2.10.1.1 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 LSPs to allow the software to make the forwarding decisions, or manually configure some or all routers in the path. The LSP is set up by Resource Reservation Protocol (RSVP), through RSVP signaling messages. SR OS 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 next-hop IP address and the outgoing label are explicitly specified.

2.10.1.2 Paths

To configure signaled LSPs, the user must first create one or more named paths on the ingress router. For each path, the transit routers (hops) in the path are specified.

2.10.1.3 Router interface

At least one router interface and one system interface must be defined to configure MPLS on an interface. Use the commands in the following context to configure interfaces.

```
configure router interface
```

2.10.1.4 Choosing the signaling protocol

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 advertises the Traffic Engineering (TE) information for the link to other routers in the network 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.

2.10.2 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 command options are optional. The following is an MPLS configuration example.

Example: MD-CLI

```
[ex:/configure routing-options]
A:admin@node-2# info
    if-attribute {
        admin-group "green" {
            value 15
        }
        admin-group "red" {
            value 25
        }
        admin-group "yellow" {
            value 20
        }
    }
[ex:/configure router "Base" mpls]
A:admin@node-2# info
    admin-state enable
    interface "StaticLabelPop" {
        admin-group ["green"]
        label-map 35 {
            admin-state enable
            swap {
                out-label 36
                next-hop 10.10.10.91
            }
        }
        label-map 50 {
            admin-state enable
            pop
        }
    }
    path "secondary-path" {
        admin-state enable
    }
    path "to-NYC" {
        admin-state enable
        hop 1 {
            ip-address 10.10.10.104
            type strict
        }
    }
    lsp "lsp-to-eastcoast" {
        admin-state enable
        type p2p-rsvp
        from 10.10.10.103
        to 10.10.10.104
        fast-reroute {
        }
        primary "to-NYC" {
        }
        secondary "secondary-path" {
```

```
        }
    }
    static-lsp "StaticLabelPush" {
        admin-state enable
        to 10.10.11.105
        push {
            out-label 60
            next-hop 10.10.11.105
        }
    }
```

Example: classic CLI

```
A:node-2>config>router>if-attr# info
----------------------------------------------
        admin-group "green" value 15
        admin-group "red" value 25
        admin-group "yellow" value 20
----------------------------------------------
A:node-2>config>router>mpls# info
----------------------------------------------
        interface "system"
            no shutdown
        exit
        interface "StaticLabelPop"
            admin-group "green"
            label-map 35
                swap 36 nexthop 10.10.10.91
                no shutdown
            exit
            label-map 50
                pop
                no shutdown
            exit
            no shutdown
        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
----------------------------------------------
```

2.10.3 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

For MPLS to run, you must configure at least one MPLS interface in the **configure router mpls** context.

- An interface must be created in the **configure router interface** context before it can be applied to MPLS.
- In the **configure router mpls** context, configure path command options 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 following command and at least one primary or secondary path must be specified.

```
configure router mpls lsp to
configure router mpls static-lsp to
```

All other statements under the LSP hierarchy are optional.

2.10.4 Configuring MPLS components

Use the MPLS and RSVP CLI syntax shown in the following sections to configure MPLS components.

2.10.4.1 Configuring global MPLS command options

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.

Use the command options in the following context to configure MPLS admin groups:

- **MD-CLI**

```
configure routing-options if-attribute admin-group
```

- **classic CLI**

```
configure router if-attribute admin-group
```

Use the following additional commands to finish configuring MPLS admin groups.

```
configure router mpls frr-object
configure router mpls resignal-timer
```

> 📝 **Note:** The **frr-object** command is enabled by default.

The following example shows an admin group configuration.

Example: MD-CLI

```
[ex:/configure routing-options]
A:admin@node-2# info
    if-attribute {
        admin-group "green" {
            value 15
        }
        admin-group "red" {
            value 25
        }
        admin-group "yellow" {
            value 20
        }
    }
[ex:/configure router "Base" mpls]
A:admin@node-2# info
    resignal-timer 500
```

Example: classic CLI

```
A:node-2>config>router>if-attr# info
----------------------------------------------
            admin-group "green" value 15
            admin-group "yellow" value 20
            admin-group "red" value 25
----------------------------------------------
A:node-2>config>router>mpls# info
----------------------------------------------
...
            resignal-timer 500
...
----------------------------------------------
```

2.10.4.2 Configuring an MPLS interface

Use the commands in the following context to configure an MPLS interface on the router.

```
configure router mpls interface
```

If the interface is used in a static LSP, use the commands in the following context to configure the label map.

```
configure router mpls interface label-map
```

The following example shows an MPLS interface configuration.

Example: MD-CLI

```
[ex:/configure router "Base" mpls]
A:admin@node-2# info
    admin-state enable
    interface "to-104" {
        admin-group ["green" "red" "yellow"]
        label-map 35 {
            admin-state enable
```

```
            swap {
                out-label 36
                next-hop 10.10.10.91
            }
        }
    }
```

Example: classic CLI

```
A:node-2>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
                no shutdown
            exit
            no shutdown
----------------------------------------------
```

2.10.4.3 Configuring MPLS paths

Configure an LSP path to use in MPLS. When configuring an LSP, the IP address of the hops that the LSP traverses 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 commands in the following context to configure a path.

```
configure router mpls path
```

The following example shows a path configuration.

Example: MD-CLI

```
[ex:/configure router "Base" mpls]
A:admin@node-2# info
    admin-state enable
    path "secondary-path" {
        admin-state enable
        hop 1 {
            ip-address 10.10.0.121
            type strict
        }
        hop 2 {
            ip-address 10.10.0.145
            type strict
        }
        hop 3 {
            ip-address 10.10.0.1
            type strict
        }
    }
    path "to-NYC" {
```

```
            hop 1 {
                ip-address 10.10.10.103
                type strict
            }
            hop 2 {
                ip-address 10.10.0.210
                type strict
            }
            hop 3 {
                ip-address 10.10.0.215
                type loose
            }
        }
```

Example: classic CLI

```
A:node-2>config>router>mpls$ info
----------------------------------------------
            interface "system"
                no shutdown
            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"
                shutdown
                hop 1 10.10.10.103 strict
                hop 2 10.10.0.210 strict
                hop 3 10.10.0.215 loose
            exit
            no shutdown
----------------------------------------------
```

2.10.4.4 Configuring an MPLS LSP

Configure an LSP path for MPLS. When configuring an LSP, the user must specify the IP address of the egress router in the **to** command, and 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 example shows an MPLS LSP configuration.

Example: MD-CLI

```
[ex:/configure router "Base" mpls]
A:admin@node-2# info
    admin-state enable
...
    lsp "lsp-to-eastcoast" {
        admin-state enable
        type p2p-rsvp
        to 192.168.200.41
        rsvp-resv-style ff
        adspec true
        path-computation-method local-cspf
        include-admin-group ["green" "red"]
        fast-reroute {
        }
```

```
        primary "to-NYC" {
            hop-limit 10
        }
        secondary "secondary-path" {
            bandwidth 50000
        }
    }
```

Example: classic CLI

```
A:node-2>config>router>mpls$ info
----------------------------------------------
...
            lsp "lsp-to-eastcoast"
                to 192.168.200.41
                rsvp-resv-style ff
                path-computation-method local-cspf
                include "green"
                include "red"
                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
----------------------------------------------
```

2.10.4.5 Configuring a static LSP

An LSP can be explicitly (statically) configured. Static LSPs are configured on every node along the path. The forwarding information of the label includes the address of the next-hop router.

Use the commands in the following context to configure a static LSP.

```
configure router mpls static-lsp
```

The following example shows a static LSP configuration.

Example: MD-CLI

```
[ex:/configure router "Base" mpls]
A:admin@node-2# info
...
    static-lsp "static-LSP" {
        admin-state enable
        to 10.10.10.124
        push {
            out-label 60
            next-hop 10.10.42.3
        }
    }
```

Example: classic CLI

```
A:node-2>config>router>mpls# info
----------------------------------------------
...
            static-lsp "static-LSP"
                to 10.10.10.124
                push 60 nexthop 10.10.42.3
                no shutdown
            exit
...
----------------------------------------------
```

2.10.4.6 Configuring manual bypass tunnels

Consider the following network setup:

A----B----C----D

| |

E----F

The user first configures the command option to disable the dynamic bypass tunnels on node B, if required. Use the following command option to disable dynamic bypass tunnels:

- **MD-CLI**

```
configure router mpls dynamic-bypass false
```

- **classic CLI**

```
configure router mpls dynamic-bypass disable
```

> **Note:** Dynamic bypass tunnels are enabled by default.

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.

Defining an LSP as a manual bypass LSP, exclusively, disables the following commands under the LSP configuration:

- bandwidth
- fast-reroute
- secondary

Use the following command option to define an LSP as a manual bypass LSP:

- **MD-CLI**

```
configure router mpls lsp type p2p-rsvp-bypass
```

- **classic CLI**

```
configure router mpls lsp bypass-only
```

The following LSP configuration command options are allowed.

```
configure router mpls lsp adaptive
configure router mpls lsp adspec
configure router mpls lsp hop-limit
configure router mpls lsp metric-type
configure router mpls lsp path-computation-method local-cspf
```

In the MD-CLI, the following additional commands are allowed.

```
configure router mpls lsp include-admin-group
configure router mpls lsp exclude-admin-group
```

In the classic CLI, the following additional commands are allowed.

```
configure router mpls lsp include
configure router mpls lsp exclude
```

The following example shows a bypass tunnel configuration.

Example: MD-CLI

```
[ex:/configure router "Base" mpls]
A:admin@node-2# info
    path "BEFC" {
        admin-state enable
        hop 10 {
            ip-address 10.10.10.11
            type strict
        }
        hop 20 {
            ip-address 10.10.10.12
            type strict
        }
        hop 30 {
            ip-address 10.10.10.13
            type strict
        }
    }
    lsp "bypass-BC" {
        admin-state enable
        type p2p-rsvp
        to 10.10.10.15
        primary "BEFC" {
        }
    }
```

Example: classic CLI

```
A:node-2>config>router>mpls$ 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
                exit
    ----------------------------------------------
```

Next, the user configures an LSP from A to D. Use the following command option to indicate fast-reroute bypass protection for the LSP:

- **MD-CLI**

```
configure router mpls lsp fast-reroute frr-method facility
```

- **classic CLI**

```
configure 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, node B uses that.

2.10.4.7 Configuring RSVP command options

RSVP is used to set up LSPs. RSVP must be enabled on the router interfaces that are participating in signaled LSPs. The default values for the following commands can be modified in the RSVP context.

```
configure router rsvp keep-multiplier
configure router rsvp refresh-time
```

Use the commands in the following context to configure MPLS interfaces.

```
configure router mpls interface
```

Only existing MPLS interfaces can be modified for RSVP in the following context.

```
configure router rsvp
```

Interfaces cannot be added directly in the RSVP context.

The following example shows an RSVP configuration.

Example: MD-CLI

```
[ex:/configure router "Base" rsvp]
A:admin@node-2# info
    admin-state enable
    interface "to-104" {
        hello-interval 4
    }
```

Example: classic CLI

```
A:node-2>config>router>rsvp# info
----------------------------------------------
            interface "system"
                no shutdown
            exit
```

```
                interface "to-104"
                    hello-interval 4000
                    no shutdown
                exit
                no shutdown
---------------------------------------------
```

2.10.4.8 Configure RSVP message pacing

RSVP message pacing maintains a count of the messages that were dropped because the output queue for the egress interface was full.

Use the commands in the following context to configure RSVP message pacing.

```
configure router rsvp msp-pacing
```

The following example shows an RSVP message pacing configuration.

Example: MD-CLI

```
[ex:/configure router "Base" rsvp]
A:admin@node-2# info
    admin-state enable
    keep-multiplier 5
    refresh-time 60
    msg-pacing {
        max-burst 400
        period 400
    }
    interface "system" {
        admin-state enable
    }
    interface "to-104" {
        hello-interval 60
    }
```

Example: classic CLI

```
A:node-2>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 60
                no shutdown
            exit
            no shutdown
---------------------------------------------
```

2.10.4.9 Configuring graceful shutdown

Use the following command to enable TE graceful shutdown on a specific interface:

- **MD-CLI**

```
configure router rsvp interface graceful-shutdown true
```

- **classic CLI**

```
configure router rsvp interface graceful-shutdown
```

An interface on which TE graceful shutdown is enabled is considered the maintenance interface.

Use the following command to disable TE graceful shutdown at the RSVP level:

- **MD-CLI**

```
configure router rsvp graceful-shutdown false
```

- **classic CLI**

```
configure router rsvp no graceful-shutdown
```

Use the following command to disable TE graceful shutdown at the RSVP interface level:

- **MD-CLI**

```
configure router rsvp interface graceful-shutdown false
```

- **classic CLI**

```
configure router rsvp interface no graceful-shutdown
```

In this case, the user-configured TE command options of the maintenance links are restored and the maintenance node floods them.

2.11 MPLS configuration management tasks

This section provides information about MPLS configuration management tasks.

2.11.1 Deleting MPLS

Before removing the MPLS instance, MPLS must be administratively disabled and all SDP bindings to LSPs must be removed. Use the following command to administratively disable MPLS:

- **MD-CLI**

```
configure router mpls admin-state disable
```

- **classic CLI**

```
configure router mpls shutdown
```

Use the following command to remove MPLS on the router:

> ✎ **Note:** If MPLS is not administratively disabled first, a warning message on the console displays, indicating that MPLS is still administratively up.

- **MD-CLI**

```
configure router delete mpls
```

- **classic CLI**

```
configure router no mpls
```

When MPLS is administratively disabled, the preceding command deletes the protocol instance and removes all configuration command options for the MPLS instance.

2.11.2 Modifying MPLS command options

In the classic CLI, the user must administratively disable MPLS entities to modify command options. After the command options have been modified, administratively enable the entities again for the modifications to take effect.

2.11.3 Modifying an MPLS LSP

In the classic CLI, some MPLS LSP commands such as **primary** and **secondary**, must be administratively disabled before they can be edited or deleted from the configuration.

The following example shows an MPLS LSP configuration.

Example: MD-CLI

```
[ex:/configure router "Base" mpls lsp "example-lsp"]
A:admin@node-2# info
    type p2p-rsvp
    from 10.10.10.103
    to 10.10.10.104
    rsvp-resv-style ff
    include-admin-group ["red"]
    exclude-admin-group ["green"]
    fast-reroute {
    }
    primary "to-NYC" {
        hop-limit 50
    }
    secondary "secondary-path" {
    }
```

Example: classic CLI

```
A:node-2>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
---------------------------------------------
```

See for more information.

2.11.4 Modifying MPLS path command options

Use the following command to administratively disable the path.

- **MD-CLI**

```
configure router mpls path admin-state enable
```

- **classic CLI**
  In the classic CLI, before modifying path command options, the path must be administratively disabled.

```
configure router mpls path shutdown
```

In the MD-CLI, you can modify the configuration without first administratively disabling the path.

The following example shows an MPLS path configuration.

Example: MD-CLI

```
[ex:/configure router "Base" mpls]
A:admin@node-2# info
    path "secondary-path" {
        admin-state enable
        hop 1 {
            ip-address 10.10.0.111
            type strict
        }
        hop 2 {
            ip-address 10.10.0.222
            type strict
        }
        hop 3 {
            ip-address 10.10.0.123
            type strict
        }
    }
    path "to-NYC" {
        admin-state enable
        hop 1 {
            ip-address 10.10.10.104
            type strict
        }
```

```
            hop 2 {
                ip-address 10.10.0.210
                type strict
            }
        }
...
```

Example: classic CLI

```
A:node-2>config>router>mpls$ info
----------------------------------------------
...
            path "to-NYC"
                hop 1 10.10.10.104 strict
                hop 2 10.10.0.210 strict
                no shutdown
            exit
            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
...
----------------------------------------------
```

See Configuring MPLS paths for another MPLS path configuration example.

2.11.5 Modifying MPLS static LSP command options

Use the following command to administratively disable the path:

- **MD-CLI**

```
configure router mpls path admin-state disable
```

- **classic CLI**

    In the classic CLI before modifying static LSP command options, the path must be administratively disabled.

```
configure router mpls path shutdown
```

The following example shows a static LSP configuration.

Example: MD-CLI

```
[ex:/configure router "Base" mpls]
A:admin@node-2# info
    admin-state enable
...
    static-lsp "static-LSP" {
        admin-state enable
        to 10.10.10.234
        push {
            out-label 102704
            next-hop 10.10.8.114
        }
```

Example: classic CLI

```
A:node-2>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
----------------------------------------------
```

See Configuring a static LSP for another static LSP configuration example.

2.11.6 Deleting an MPLS interface

Use the following command to administratively disable an MPLS interface:

- **MD-CLI**

```
configure router mpls interface admin-state disable
```

- **classic CLI**

  In the classic CLI before deleting an interface from the MPLS configuration, the interface must be administratively disabled.

```
configure router mpls interface shutdown
```

Use the following command to delete an interface from the MPLS configuration:

- **MD-CLI**

```
configure router mpls delete interface
```

- **classic CLI**

```
configure router mpls no interface
```

Admin groups are preserved when an interface is deleted from the MPLS configuration, as shown in the following configuration example.

Example: MD-CLI

```
[ex:/configure routing-options if-attribute]
A:admin@node-2# info
    admin-group "green" {
        value 15
    }
    admin-group "red" {
        value 25
    }
    admin-group "yellow" {
        value 20
    }
```

```
[ex:/configure router "Base" mpls]
A:admin@node-2# info detail
...
    interface "system" {
     ## apply-groups
     ## apply-groups-exclude
        admin-state enable
     ## te-metric
     ## admin-group
     ## srlg-group
     ## label-map
    }
...
```

Example: classic CLI

```
A:node-2>config>router>if-attr# info
----------------------------------------------
            admin-group "green" value 15
            admin-group "red" value 25
            admin-group "yellow" value 20
----------------------------------------------
A:node-2>config>router>mpls$ info
----------------------------------------------
            interface "system"
                no shutdown
            exit
            no shutdown
----------------------------------------------
```

2.12 RSVP configuration management tasks

This section provides information about RSVP configuration management tasks.

2.12.1 Modifying RSVP command options

Only interfaces configured in the MPLS context can be modified in the RSVP context.

Use the following command to delete an RSVP protocol instance and remove all command-option configurations under it:

- **MD-CLI**

```
configure router delete rsvp
```

- **classic CLI**

```
configure router no rsvp
```

Administratively disabling the RSVP protocol instance suspends the execution and maintains the existing configuration. Use the following command to administratively disable the RSVP protocol instance:

- **MD-CLI**

```
configure router rsvp admin-state disable
```

- **classic CLI**

```
configure router rsvp shutdown
```

The following example shows a modified RSVP configuration.

Example: MD-CLI

```
[ex:/configure router "Base" rsvp]
A:admin@node-2# info
    admin-state enable
    keep-multiplier 5
    refresh-time 60
    msg-pacing {
        max-burst 400
        period 400
    }
    interface "test1" {
        hello-interval 5
    }
```

Example: classic CLI

```
A:node-2>config>router>rsvp# info
----------------------------------------------
            keep-multiplier 5
            refresh-time 60
            msg-pacing
                period 400
                max-burst 400
            exit
            interface "system"
                no shutdown
            exit
            interface "test1"
                hello-interval 5000
                no shutdown
            exit
            no shutdown
----------------------------------------------
```

2.12.2 Modifying RSVP message pacing command options

RSVP message pacing maintains a count of the messages dropped because the output queue for the egress interface was full.

The following example shows a modified RSVP message pacing configuration.

Example: MD-CLI

```
[ex:/configure router "Base" rsvp]
A:admin@node-2# info
    admin-state enable
    keep-multiplier 5
    refresh-time 60
    msg-pacing {
        max-burst 200
        period 200
    }
```

```
        interface "to-104" {
        }
```

Example: classic CLI

```
A:node-2>config>router>rsvp# info
----------------------------------------------
            keep-multiplier 5
            refresh-time 60
            msg-pacing
                period 200
                max-burst 200
            exit
            interface "system"
                no shutdown
            exit
            interface "to-104"
                no shutdown
            exit
            no shutdown
----------------------------------------------
```

See Configure RSVP message pacing for another RSVP message pacing configuration example.

2.12.3 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 about deleting an MPLS interface.

3 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 allow other LSRs in the path to 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 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).

3.1 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 bidirectional) 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. 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, and liberal label retention mode support.

3.2 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 LDP session, while buffering alleviates TCP-related congestion issues.

The LDP subsystems and their relationships to other subsystems are illustrated in Figure 22: Subsystem interrelationships. 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. LDP maintains the following primary data structures:

- **FEC/label database**

  This database contains all the FEC to label mappings that include both sent and received. It also contains both address FECs (prefixes and host addresses) and service FECs (Layer 2 VLLs and VPLS).

- **timer database**

  This database contains all the timers for maintaining sessions and adjacencies.

- **session database**

  This database contains all the session and adjacency records, and serves as a repository for the LDP MIB objects.

3.3 Subsystem interrelationships

The following sections describe how LDP and the other subsystems work to provide services.

The following figure shows the subsystem interrelationships.

Figure 22: Subsystem interrelationships

OSSRG017
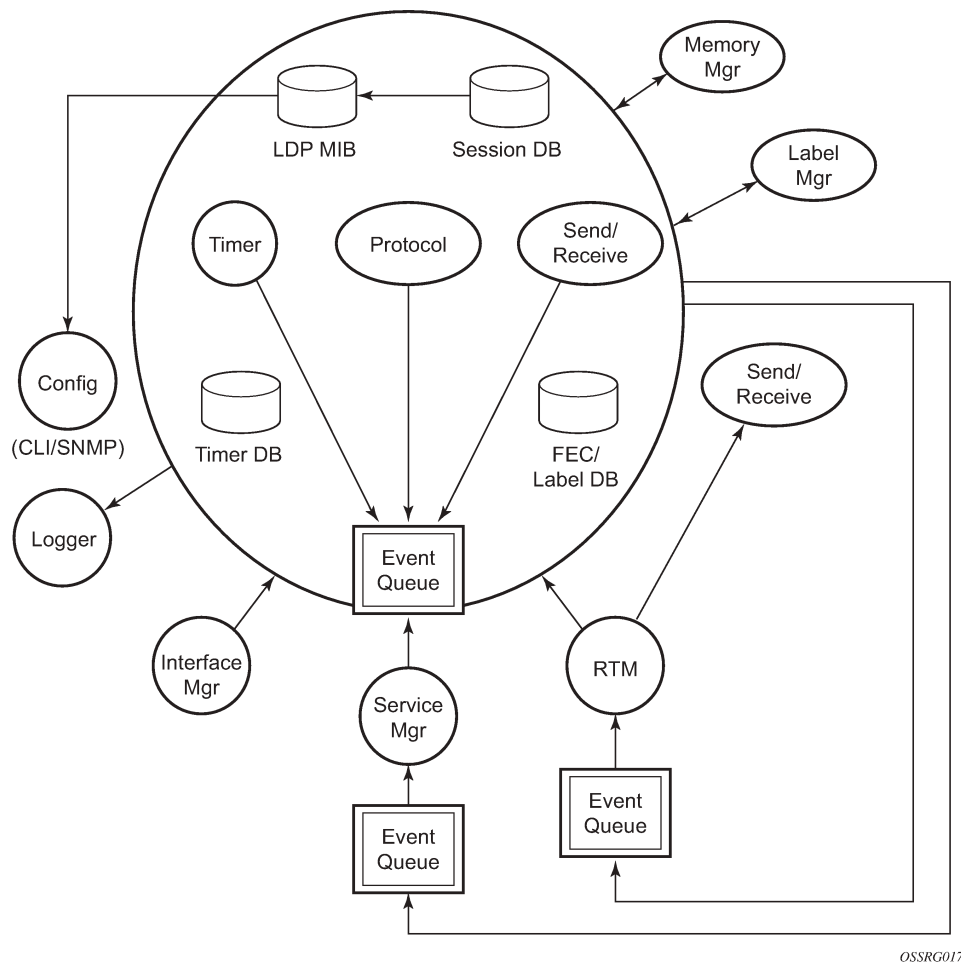
3.3.1 Memory manager and LDP

LDP does not use any memory until it is instantiated. It preallocates 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.

3.3.2 Label manager

LDP assumes that the label manager is up and running. LDP aborts initialization if the label manager is not running. The label manager is initialized at system boot up; therefore, anything that causes it to fail likely implies 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.

3.3.3 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.

3.3.4 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.

3.3.5 Service manager

All interaction occurs between LDP and the service manager because LDP is used primarily to exchange labels for Layer 2 services. 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 the service manager of relevant LDP events, such as connection setups and failures, timeouts, and labels signaled/withdrawn.

3.4 Execution flow

LDP activity is limited to service-related signaling. Therefore, the configurable parameters are restricted to system-wide parameters, such as hello and keepalive timeouts.

3.4.1 Initialization

LDP ensures 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 a pool of memory and initializes its databases.

3.4.2 Session lifetime

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.

3.4.2.1 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 the following main entities:

- customers

- services

- Service Access Paths (SAPs) on the local routers

- service destination 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 following:

- SDP ID

- transport encapsulation type (either MPLS or GRE)

- far-end system IP address

If the SDP is identified as using LDP signaling, then an LDP extended Hello adjacency is attempted.

If the **tldp** command option is selected as the mechanism for exchanging service labels over an MPLS or GRE SDP and the T-LDP session is automatically established, the explicit T-LDP session that is subsequently configured takes precedence over the automatic T-LDP session.

> **Note:** Use the following command option to enable ingress and egress pseudowire signaling using T-LDP.

```
configure service sdp signaling tldp
```

However, if the explicit, manually configured session is removed, the system does not revert to the automatic session and the automatic session is also deleted. To address this, recreate the T-LDP session by disabling and re-enabling the SDP. Use the following command to administratively disable the SDP:

- **MD-CLI**

```
configure service sdp admin-state disable
```

- **classic CLI**

```
configure service sdp shutdown
```

Use the following command to administratively re-enable the SDP:

- **MD-CLI**

```
configure service sdp admin-state enable
```

- **classic CLI**

```
configure service sdp no shutdown
```

If another SDP is created to the same remote destination, and if LDP signaling is enabled, no further action is taken, because only one adjacency and one LDP session exists between the pair of nodes.

An SDP is a unidirectional object, so a pair of SDPs pointing at each other must be configured in order for an LDP adjacency to be established. When an adjacency is established, it is maintained through periodic Hello messages.

3.4.2.2 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.

Because 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 reattempted as the back-pressure eases.

3.5 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.

3.5.1 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.

3.5.2 Cleanup

LDP closes all sockets, frees all memory, and shuts down all its tasks when it is deleted. LDP does this so that it does not use any memory while it is not running.

3.5.3 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.

Use the following command to enable the implicit null option for all LDP FECs for which the node is the egress LER:

- **MD-CLI**

```
configure router ldp implicit-null-label true
```

- **classic CLI**

```
configure router ldp implicit-null-label
```

When the user changes the implicit null configuration, LDP withdraws all the FECs and re-advertises them using the new label value.

3.6 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 the following:

- prefix list (match on bindings with the specified prefix or 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. Use the following command to configure the router 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:

- **MD-CLI**

```
configure router ldp consider-system-ip-in-gep true
```

- **classic CLI**

```
configure router ldp consider-system-ip-in-gep
```

Export policy enables configuration of a policy to advertise label bindings based on the following:

- 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 of a global import policy is not considered by LDP.

3.6.1 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.

Use the following command to configure the FEC prefix export policy.

```
configure router ldp session-parameters peer export-prefixes
```

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.

Use the following command to configure the FEC prefix import policy.

```
configure router ldp session-parameters peer import-prefixes
```

By default, all FEC prefixes are imported from this peer.

3.7 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 does 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 auto-establishes one but uses 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 specific 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 does 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 also uses the address of that interface as the transport address.

3.7.1 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. Use the following command to configure the advertisement of the local LSR ID in the session parameters for a specified peer.

```
configure router ldp session-parameters peer adv-local-lsr-id
```

Use the following command to configure the advertisement of the local LSR ID for the targeted-session peer template.

```
configure router ldp targeted-session peer-template adv-local-lsr-id
```

> **Note:** When the **adv-local-lsr-id** command is configured, the FEC is effectively advertised if no export prefixes policy prevents the advertisement.

3.8 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 policy in the following example was assigned as an import policy to mLDP, any FEC arriving with an IP address of 100.0.1.21 is dropped.

Example: MD-CLI

```
[ex:/configure policy-options]
A:admin@node-2# info
    prefix-list "100.0.1.21/32" {
        prefix 100.0.1.21/32 type exact {
        }
    }
    policy-statement "policy1" {
        entry 10 {
            from {
                prefix-list ["100.0.1.21/32"]
            }
            action {
                action-type reject
            }
        }
        entry 20 {
        }
        default-action {
            action-type accept
        }
    }
```

Example: classic CLI

```
A:node-2>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
----------------------------------------------
```

Use the following command to assign the policy to mLDP.

```
configure router ldp import-mcast-policy
```

If the preceding command is configured, the prefix list matches the mLDP outer FEC and the action is 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.

3.8.1 Recursive FEC behavior

In the case of recursive FEC, the prefix list matches the outer root. For example, for recursive FEC <outerROOT, opaque <ActualRoot, opaque<lspID>> the import policy works 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.

3.8.2 Import policy

For mLDP, a policy can be assigned as an import policy only. Import policies only affect FECs arriving to the node, and do not affect the self-generated FECs on the node. The import policy causes the multicast FECs received from the peer to be rejected and stored in the LDP database but not resolved. Therefore, the following command displays the FEC, but the **active** command under the same context does not.

```
show router ldp bindings
```

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 only matches the prefix against the root node address field of the FEC, and no other fields. This means that the policy works on all P2MP Opaque types.

The following figure shows the P2MP FEC element encoding.

Figure 23: P2MP FEC element encoding
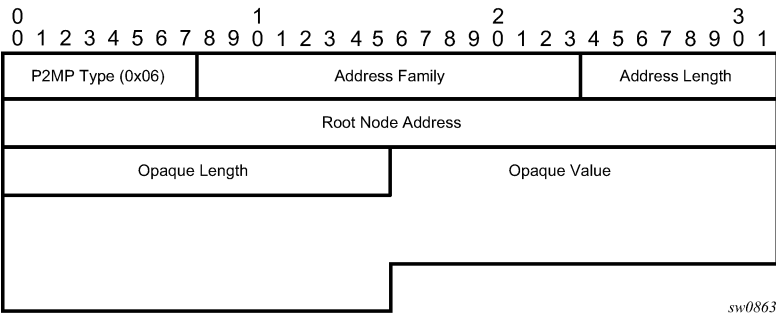
3.9 LDP FEC resolution per specified community

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 specific 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.

3.9.1 Configuration

> **Note:** The **no local-lsr-id** or **local-lsr-id system** commands only apply to the classic CLI. 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 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 for the following commands.

```
configure router ldp interface-parameters interface ipv4 local-lsr-id
configure router ldp interface-parameters interface ipv6 local-lsr-id
configure router ldp targeted-session peer local-lsr-id
configure router ldp targeted-session peer-template local-lsr-id
```

It is never applied to a system FEC or local static FEC. 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 toward a specified peer or within a specified targeted peer template. The FEC advertised by the following commands is automatically put in the community configured on the session.

```
configure router ldp session-parameters peer adv-local-lsr-id
configure router ldp targeted-session peer-template adv-local-lsr-id
```

The specified community is only associated with 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.

3.9.2 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 because of 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 specific 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 advertised local LSR ID FECs are exchanged, and the FECs are activated correctly as per routing. On each node, for each FEC there is 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 shown in the output of the following command.

```
show router ldp bindings prefixes
```

The community mismatch state is visible through 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

3.10 T-LDP Hello reduction

This feature implements a 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 specific network IP interface and therefore, the periodic transmission of Hello messages is required to discover new peers in addition to the periodic keepalive message transmission to maintain the existing LDP sessions. A targeted LDP session is established to a single peer. Consequently, after the Hello adjacency is established and the LDP session is brought up over a TCP connection, keepalive 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 starts 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 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), Hello messages are sent every 21844 seconds.

Both LDP peers must be configured with this feature to gradually bring 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 continues to be governed by the smaller of the two Hold-Time values. This feature complies with *draft-pdutta-mpls-tldp-hello-reduce*.

3.11 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.

3.12 Link LDP Hello adjacency tracking with BFD

LDP can only track an LDP peer using the Hello and keepalive 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 because the LDP peer is not directly tracked by BFD.

To properly track the link LDP peer, LDP needs to track the Hello adjacency to its peer by registering with BFD.

Use the following command to enable the Hello adjacency tracking of IPv4 LDP sessions with BFD:

- **MD-CLI**

```
configure router ldp interface-parameters interface bfd-liveness ipv4 true
```

- **classic CLI**

```
configure router ldp interface-parameters interface bfd-enable ipv4
```

Use the following command to enable the Hello adjacency tracking of IPv6 LDP sessions with BFD:

- **MD-CLI**

```
configure router ldp interface-parameters interface bfd-liveness ipv6 true
```

- **classic CLI**

```
configure router ldp interface-parameters interface bfd-enable ipv6
```

Use the command options in the following context to configure BFD sessions for IPv4:

- **MD-CLI**

```
configure router interface ipv4 bfd
```

- **classic CLI**

```
configure router interface bfd
```

Use the command options in the following context to configure BFD sessions for IPv6.

```
configure router interface ipv6 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 exists to the peer and uses 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 immediately brings 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 results 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.

When the last Hello adjacency goes down because of BFD timing out, the LDP session goes down and the LDP FRR procedures are triggered. LDP FRR procedures result in moving the traffic to an LFA backup next hop on a different LDP peer.

3.13 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 datapath 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 datapath statistics is provided whenever applicable.

3.14 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. Use the following command to allow for TTM to be populated with LDP tunnels to non-host prefixes in a controlled manner for both IPv4 and IPv6.

```
configure router ldp import-tunnel-table
```

3.15 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. Because 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 multihop sessions. When a multihop 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 catches a vast majority of observed distributed DoS (DDoS) attacks against EBGP.

TSH can be used to protect LDP peering sessions as well. For more information, 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.

3.16 ECMP support for LDP

ECMP support for LDP performs load balancing for LDP-based LSPs by using multiple outgoing next hops for an IP prefix on ingress and transit LSRs.

An LSR that has multiple equal cost paths to an IP prefix can receive an LDP label mapping for this prefix from each of the downstream next-hop peers. The LDP implementation uses the liberal label retention mode, that is, 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 select the next-hop peer 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 the outgoing label to the installed next hop is no longer valid (for example, 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 (peers that sent a label mapping for an IP prefix) are installed in the forwarding plane.

In both cases, an ingress LER and a transit LSR, an ingress label is 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 specific next hop.

The hash algorithm at LER and transit LSR are described in "Traffic Load Balancing Options" in the *7705 SAR Gen 2 Interface Configuration Guide*.

LDP supports up to 64 ECMP next hops. LDP derives its maximum limit from the whichever value is lower between the following commands.

```
configure router ecmp
configure router ldp max-ecmp-routes
```

3.16.1 Label operations

If an LSR is the ingress for an IP prefix, LDP programs a push operation for the prefix in the forwarding engine and creates an LSP ID for the next-hop Label Forwarding Entry (NHLFE) (LTN) mapping and an LDP tunnel entry in the forwarding plane. LDP also informs the TTM of this tunnel. Both the LTN entry and the tunnel entry have an NHLFE for the label mapping that the LSR received from each of its next-hop peers.

If the LSR behaves as a transit for an IP prefix, LDP programs a swap operation for the prefix in the forwarding engine. Programming a swap operation involves creating an Incoming Label Map (ILM) entry in the forwarding plane. The ILM entry must map an incoming label to multiple NHLFEs. If the LSR is an egress for an IP prefix, LDP programs a POP entry in the forwarding engine. Similarly, programming a POP entry results in the creation of an ILM entry in the forwarding plane but with no NHLFEs.

When unlabeled packets arrive at the ingress LER, the forwarding plane consults the LTN entry and uses a hashing algorithm to map the packet to one of the NHLFEs (push label), and then forwards the packet to the corresponding next-hop peer. For labeled packets arriving at a transit or egress LSR, the forwarding plane consults the ILM entry and either uses a hashing algorithm to map it to one of the NHLFEs (swap label), or routes the packet if there are no NHLFEs (pop label).

Static FEC swap is not activated unless there is a matching route in the system route table that also matches the user configured static FEC next-hop.

3.17 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 support for LSP ping, P2MP LSP ping, and LDP tree trace, allowing these features to test an LDP unicast or multicast FEC which is resolved over an unnumbered LDP interface.

3.17.1 Feature configuration

This feature does not implement a command for adding an unnumbered interface into LDP. Instead, use the following command to specify the interface name, as an unnumbered interface does not have an IP address of its own:

- **MD-CLI**

```
configure router ldp fec-originate interface
```

- **classic CLI**

```
configure router ldp fec-originate
```

The user can, however, specify the interface name for numbered interfaces.

3.17.2 Operation of LDP over an unnumbered IP interface

Consider the setup shown in Figure 24: LDP adjacency and session over unnumbered interface.

Figure 24: LDP adjacency and session over unnumbered interface
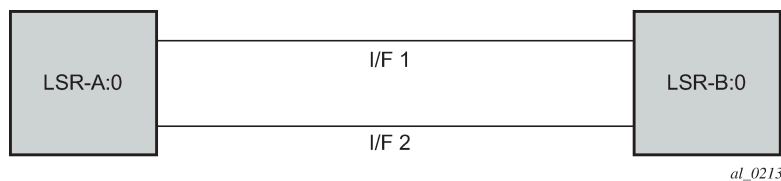
LSR A and LSR B have the following LDP identifiers respectively:

<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. Use the command options in following context to change the borrowed IP interface to any configured IP interface, regardless of whether it is a loopback interface:

- **MD-CLI**

```
configure router interface ipv4 unnumbered
```

- **classic CLI**

```
configure router interface unnumbered
```

Subsequent sections describe the behavior of an unnumbered interface when it is added into LDP.

3.17.2.1 Link LDP

When the IPv6 context of interfaces I/F1 and I/F2 are brought up, the following operations and procedures are executed.

1. 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.

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 does not come up and an error is returned (lsrInterfaceNoValidIp [17]) in the output of the following command.

   ```
   show router ldp interface detail
   ```

3. 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 does not come up and an error is returned (interfaceNoValidIp [16]) in the output of the following command.

   ```
   show router ldp interface detail
   ```

4. 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, LSR A (LSR B) resolves IPv6 FECs only because IPv6 addresses are not advertised in Address messages, in accordance with 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) does not bring up the Hello adjacency and discards the Hello message. If the LDP session was already established, LSRA(B) sends a fatal Notification message with status code of "Transport Connection Mismatch (0x00000032)" and

restarts the LDP session as defined in RFC 7552 [*ldp-ipv6-rfc*]. In both cases, a new counter for the transport connection mismatch is incremented in the output of the following command.

```
show router ldp statistics
```

5. 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.

3.17.2.2 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** command 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 is used as the LSR-ID. In all cases, the transport address for the LDP session and the source IP address of targeted Hello message is updated to the new LSR-ID value.

The LSR with the highest transport address, that is, LSR-ID in this case, bootstraps 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.

3.17.2.3 FEC resolution

LDP advertises or withdraws unnumbered interfaces using the Address or 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 are supported.

All LDP FEC types are supported.

In the classic CLI, the **fec-originate** command is supported when the next hop is over an unnumbered interface.

In the MD-CLI, the commands in the **fec-originate** context are 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 are not 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.

3.18 LDP over RSVP tunnels

LDP over RSVP-TE provides end-to-end tunnels that have fast reroute (FRR) and traffic engineering (TE) while still running shortest-path-first-based LDP at both endpoints of the network; FRR and TE are not typically available in LDP. LDP over RSVP-TE is typically used in large networks where a portion of the network is RSVP-TE-based with access segments of the network making use of simple LDP-based transport. In large topologies, while an LER may not have that many tunnels, any transit node potentially has thousands of LSPs, and if each transit node must also deal with detours or bypass tunnels, then the LSR can become strained.

LDP over RSVP-TE allows tunneling of services using an LDP LSP inside an RSVP LSP. The main application of this feature is for deployment of MPLS-based services, such as VPRN, VLL, and VPLS services, in large scale networks across multiple IGP areas without requiring full mesh of RSVP LSPs between PE routers.

The following figure shows an example LDP over RSVP application.

Figure 25: LDP over RSVP application
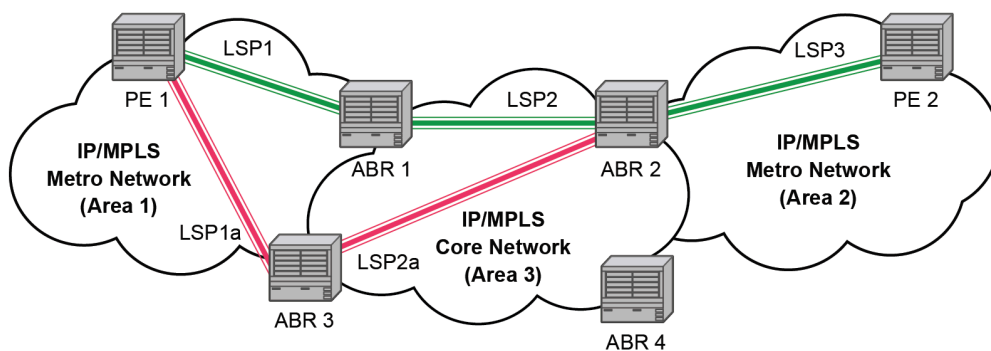
The network displayed in the preceding figure consists of two metro areas, Area 1 and 2, and a core area, Area 3. Each area makes use of TE LSPs to provide connectivity between the edge routers. To enable services between PE1 and PE2 across the three areas, LSP1, LSP2, and LSP3 are set up using RSVP-TE. There are six LSPs required for bidirectional operation, and each bidirectional LSP has 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 for the tunnel between ABR2 and PE2. The loopback address of each of these routers is advertised using T-LDP. Backup bidirectional LDP over RSVP tunnels, LSP1a and LSP2a, are configured via ABR3.

This setup effectively creates an end-to-end LDP connectivity that 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 bidirectional 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 tunnel types match the destination PE FEC lookup, LDP prefers an LDP tunnel over an LDP-over-RSVP tunnel by default.

The design in Figure 25: LDP over RSVP application 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, 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 three-label stack is required.

To provide a VLL service, PE1 and PE2 are required to set up a targeted LDP session directly between them. A three-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 25: LDP over RSVP application, in which area 1 is an LDP area. In that case, PE1 pushes a two-label stack while ABR1 swaps the LDP label and pushes the RSVP label, as shown in the following figure. LDP-over-RSVP tunnels can also be used as IGP shortcuts.

Figure 26: LDP over RSVP application variant
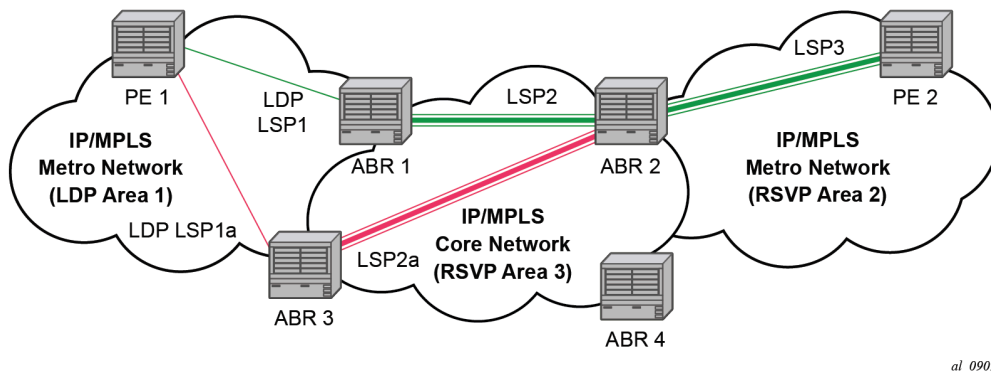
3.18.1 Signaling and operation

This section describes LDP signaling and operation procedures.

3.18.1.1 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 are 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.

3.18.1.2 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 prefers the LDP tunnel by default, unless the user explicitly changed the default preference, using the following system-wide command.

```
configure router ldp prefer-tunnel-in-tunnel
```

If the LDP tunnel becomes unavailable, LDP selects an LDP-over-RSVP tunnel if available.

When searching for an LDP-over-RSVP tunnel, LDP selects the advertising routers with best route. If the advertising router matches the T-LDP peer, LDP then performs a second lookup for the advertising router in the TTM which returns the user configured RSVP LSP with the best metric. If there is 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 returns the first available RSVP LSP which terminates on the advertising router with the lowest metric.

3.18.1.3 FEC resolution procedure when prefer-tunnel-in-tunnel is enabled

When LDP tries to resolve a prefix received over a T-LDP session, LDP performs a lookup in the 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 routers with the best route. If the advertising router matches the targeted LDP peer, LDP then performs a second lookup for the advertising router in the TTM that returns the user configured RSVP LSP with the best metric. If there is more than one configured LSP with the best metric, LDP selects the first available LSP.

If all user configured RSVP LSPs are down, an LDP tunnel is selected if available.

If the user did not configure any LSPs under the T-LDP session, a lookup in TTM returns the first available RSVP LSP that terminates on the advertising router. If none are available, then an LDP tunnel is selected.

3.18.2 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.

3.18.2.1 LDP-over-RSVP tunnel protection

An RSVP LSP can deal with a failure in the following ways.

- If the LSP is a loosely routed LSP, RSVP finds a new IGP path around the failure, and traffic follows this new path. The discovery of a new path may cause some network disruption if the LSP comes down and is then rerouted. 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 FRR enabled, RSVP switches to the detour path very quickly. A new LSP is 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.

3.18.2.2 ABR protection

If an ABR fails, then routing around the ABR requires that a new next-hop LDP-over-RSVP tunnel be assigned 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.

3.19 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 27: LDP over RSVP without ABR stitching point
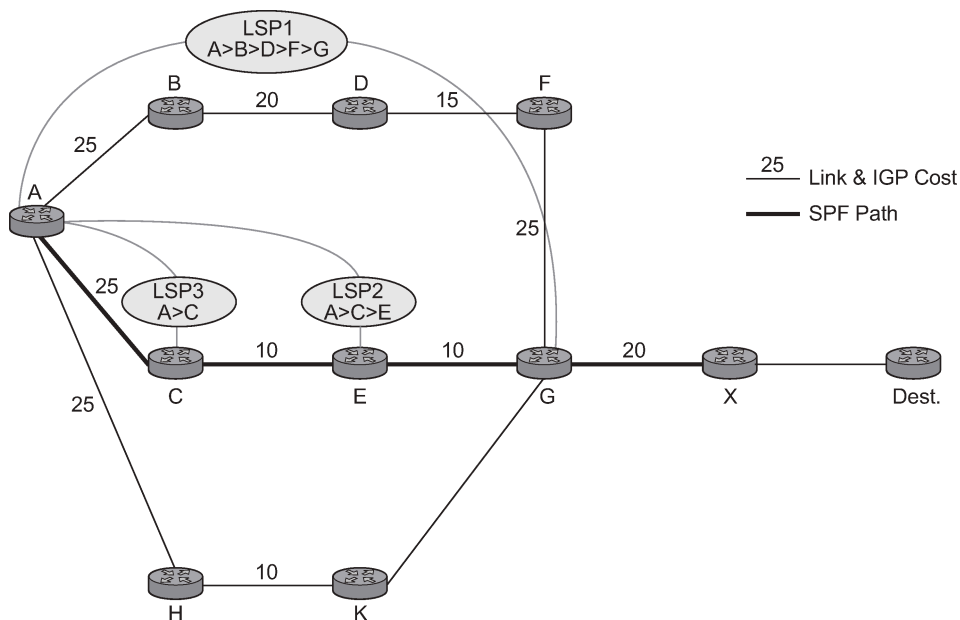
al_0214

In Figure 27: LDP over RSVP without ABR stitching point, 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 performs 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 has the following path: A>C>E>G>X.

Using this SPF tree, router A searches 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 is 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) is 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 is 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 is resolved over RSVP LSPs within an area.

3.19.1 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 is 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 chooses the endpoint that is highest cost in the route entry and does further tunnel selection over those endpoints. If there are multiple endpoints with equal highest cost, then LDP considers all of them.

3.20 Weighted load-balancing for LDP over RSVP and SR-TE

Weighted load-balancing (Weighted ECMP) is supported for LDP over RSVP (LoR) in the following scenarios:

*   The LDP next hop resolves to an IGP shortcut tunnel over RSVP.

*   The LDP resolves to a static route with next hops, which, in turn, uses RSVP tunnels.

*   The following command is configured for the LDP peer (classic LDP over RSVP).

```
configure router ldp targeted-session peer tunneling
```

It is also supported when the LDP next hop resolves to an IGP shortcut tunnel over SR-TE. Weighted load-balancing is supported for both push and swap NHLFEs.

At a high level, weighted load-balancing operates as follows:

1. All the RSVP or SR-TE LSPs in the ECMP set must have a load balancing weight configured; otherwise, non-weighted ECMP behavior is used.

2. The normalized weight of each RSVP or SR-TE LSP is calculated based on its configured load-balancing weight. LDP performs the calculation to a resolution of 64, meaning if there are values between 1 and 200, the system buckets these into 64 values. These LSP next hops are then populated in TTM.

3. RTM entries are updated accordingly for LDP shortcuts.

4. 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.

5. LDP labeled packets are then sprayed in proportion to the normalized weight of the RSVP or SR-TE LSPs that they are forwarded over.

6. No per-service differentiation exists between packets. LDP-labeled packets from all services are sprayed in proportion to the normalized weight.

7. Tunnel-in-tunnel takes precedence over the existence of a static route with a tunneled next hop. If tunneling is configured, then LDP uses these LSPs instead of 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, or LDP over SR-TE, when using IGP shortcuts, is enabled as follows:

Use the following commands to enable weighted ECMP for LDP over RSVP (when using IGP shortcuts or static routes) or LDP over SR-TE (when using IGP shortcuts).

```
configure router weighted-ecmp
configure router ldp weighted-ecmp
```

However, in the case of classic LoR, the user only needs to configure weighted ECMP needs under LDP. The maximum number of ECMP tunnels is taken from whichever of the following commands is configured to have a lower value.

```
configure router ecmp
configure router ldp max-ecmp-routes
```

The following example shows LDP resolving to a static route with one or more indirect next hops and a set of RSVP tunnels specified in the resolution filter.

Example: MD-CLI

```
[ex:/configure router "Base"]
A:admin@node-2# info
...
    static-routes {
        route 192.0.2.102/32 route-type unicast {
            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" { }
                        }
                    }
                }
            }
            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" { }
                        }
                    }
                }
            }
```

```
            }
        }
```

Example: classic CLI

```
A:node-2>config>router# info
...
#--------------------------------------------------
echo "Static Route Configuration"
#--------------------------------------------------
        static-route-entry 192.0.2.102/32
            indirect 192.0.2.2
                shutdown
                tunnel-next-hop
                    resolution disabled
                    resolution-filter
                        rsvp-te
                            lsp "LSP-ABR-1-1"
                            lsp "LSP-ABR-1-2"
                            lsp "LSP-ABR-1-3"
                        exit
                    exit
                exit
            exit
            indirect 192.0.2.3
                shutdown
                tunnel-next-hop
                    resolution disabled
                    resolution-filter
                        rsvp-te
                            lsp "LSP-ABR-2-1"
                            lsp "LSP-ABR-2-2"
                            lsp "LSP-ABR-2-3"
                        exit
                    exit
                exit
            exit
        exit
#--------------------------------------------------
...
```

If **weighted-ecmp** is enabled at both the LDP level and router level, 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.

3.20.1 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, (**weighted-ecmp** is enabled at the router level), 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 the router context.

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, if **weighted-ecmp** is not configured at the router level, or if one or more of the RSVP tunnels does not have a load balancing weight configured, 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.

3.21 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 coexist, 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

   See chapter IP Router Configuration, Section 2.3 in *7705 SAR Gen 2 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

   See chapter Virtual Private Routed Network Service, Section 3.2.27 in *7705 SAR Gen 2 Layer 3 Services Guide: IES and VPRN*, for a description of this case.

IGP IPv4 shortcuts, in all four cases, see MPLS RSVP-TE or SR-TE LSPs.

3.21.1 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.

Enable the following to activate CBF:

- 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)

Use the command options in the following context to configure FC-to-Set based configuration mode.

```
configure router mpls class-forwarding-policy
```

Additionally, use the command options in the following context to configure the mapping of a class-forwarding policy and forwarding set ID to an LSP template, so that LSPs created from the template acquire the assigned CBF configurations.

```
configure router mpls lsp-template class-forwarding forwarding-set
```

Multiple FCs can be assigned to a specific 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 or LSR roles, or both. 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 described in subsequent sections.

3.21.1.1 LSR and 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)
- 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 specific 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 specific 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 with 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 next hop from the set.

3.22 LDP ECMP uniform failover

LDP ECMP uniform failover allows the fast re-distribution by the ingress datapath 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 the other way around.

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 64, 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 therefore 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 updates 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 keepalive
- failure of LDP Hello adjacency detected via link LDP BFD or LDP Hello

In addition, PIP sends 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 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 keepalive, 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 triggers 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 causes 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 toward 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.

3.23 LDP FRR for IS-IS and OSPF prefixes

LDP FRR allows the user to provide local protection for an LDP FEC by precomputing 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 LFA next hop.

The LFA next-hop precomputation 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 specific 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 LFA next hop ahead of time and is stored in the Label Information Base because 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 RTM.

3.23.1 LDP FRR configuration

Use the follow commands to enable Loop-Free Alternate (LFA) computation by SPF under the IS-IS or OSPF routing protocol level:

- **MD-CLI**

```
configure router isis loopfree-alternate
configure router ospf loopfree-alternate
```

- **classic CLI**

```
configure router isis loopfree-alternates
configure router ospf loopfree-alternates
```

The preceding commands instruct the IGP SPF to attempt to precompute 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 command.

```
configure router ldp fast-reroute
```

When this command is enabled, LDP uses 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 results 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 or next hop independently, it is possible that LDP switches to the LFA next-hop while IGP is still using the primary next-hop. To avoid this situation, Nokia recommends enabling IGP-LDP synchronization on the LDP interface using the following command.

```
configure router interface ldp-sync-timer
```

3.23.2 LDP FRR procedures

LDP FEC resolution when LDP FRR is not enabled operates as follows. When LDP receives a *FEC, label* binding for a prefix, it resolves the binding by checking if the exact prefix, or a longest match prefix when the following command 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.

```
configure router ldp aggregate-prefix-match
```

When the next hop is no longer available, LDP deactivates the FEC and deprograms the NHLFE in the datapath. LDP immediately withdraws the labels it advertised for the FEC and deletes the ILM in the datapath, unless the user configured the following command to delay this operation.

```
configure router ldp label-withdrawal-delay
```

Traffic that is received while the ILM is still in the datapath is dropped. When routing computes and populates the routing table with a new next hop for the prefix, LDP resolves the FEC and programs the datapath 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 the **aggregate-prefix-match** command is enabled in LDP, LDP resolves the FEC, but programs the datapath with both a primary NHLFE and a backup NHLFE for each next hop of the FEC.

To 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 occur, LDP instructs in the fast path of 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 flipping a single state bit associated with the failed interface or neighbor/next hop:

- An LDP interface goes operationally down, or is administratively shut down. 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.

- An LDP session to a peer goes down because of the Hello or keepalive 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.

- The TCP connection used by a link LDP session to a peer goes 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.

- A BFD session, enabled on a T-LDP session to a peer, times out. As a result, the link LDP session to the same peer using the same TCP connection as the T-LDP session also goes down. In this case, LDP sends a neighbor/next-hop down message to the IOM line cards for this LDP peer only.

- 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 following commands, when enabled, do not cause the corresponding timer to be activated for a FEC as long as a backup NHLFE is still available.

```
configure router ldp tunnel-down-damp-time
configure router ldp label-withdrawal-delay
```

3.23.2.1 ECMP considerations

Whenever the SPF computation determines there is more than one primary next hop for a prefix, SPF does not program any LFA next hop in RTM. As such, the LDP FEC resolves to the multiple primary next hops, providing the required protection.

When the system ECMP value is configured as **configure router ecmp 1**, which is the default value, SPF can use the overflow ECMP links as LFA next hops.

3.23.2.2 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.

3.23.2.3 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 activates the LFA next hop as soon as the FEC is resolved to direct primary next hop.

LDP FEC tunneled over an RSVP LSP because of enabling the LDP-over-RSVP feature therefore does not support the LDP FRR procedures and follows the slow path procedure of prior implementation.

When the user enables the following command option for an RSVP LSP, as described in "Loop-Free Alternate calculation in the presence of IGP shortcuts" in the *7705 SAR Gen 2 Unicast Routing Protocols Guide*, the LSP is not used by LDP to tunnel an LDP FEC even when IGP shortcut is disabled but LDP-over-RSVP is enabled in IGP.

- **MD-CLI**

```
configure router mpls lsp igp-shortcut lfa-type lfa-only
```

- **classic CLI**

```
configure router mpls lsp igp-shortcut lfa-only
```

3.23.2.4 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 specific 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:

- A FEC which is resolved to a direct primary next hop can be backed up by a LFA tunneled next hop.
- A FEC which is resolved to a tunneled primary next hop does not have an LFA next hop. It relies on RSVP FRR for protection.

The LFA SPF is extended to use IGP shortcuts as LFA next hops as described in "Loop-Free Alternate calculation in the presence of IGP shortcuts" in the *7705 SAR Gen 2 Unicast Routing Protocols Guide*.

3.23.3 IS-IS and OSPF support for LFA calculation

See "OSPF and IS-IS support for LFA calculation" in the *7705 SAR Gen 2 Unicast Routing Protocols Guide*.

3.23.3.1 LFA calculation in the presence of IGP shortcuts

See "Loop-Free Alternate calculation in the presence of IGP shortcuts" in the *7705 SAR Gen 2 Unicast Routing Protocols Guide*.

3.23.3.2 LFA calculation for inter-area/inter-level prefixes

See "LFA calculation for inter-area and inter-level prefixes" in the *7705 SAR Gen 2 Unicast Routing Protocols Guide*.

3.23.3.3 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.

For more information, see "Loop-free alternate shortest path first policies" in the *7705 SAR Gen 2 Unicast Routing Protocols Guide*.

3.24 LDP FEC to BGP labeled 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. The following figure shows a network with a core area and regional areas.

Figure 28: Application of LDP to BGP FEC stitching
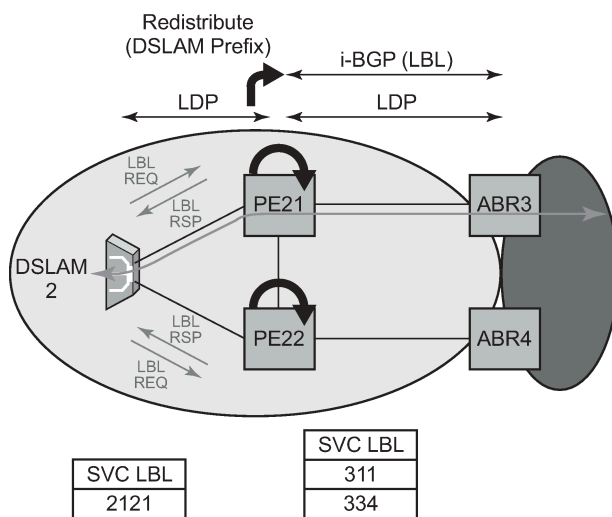
Specific /32 routes in a regional area are not redistributed into the core area. Therefore, only nodes within a regional area and the ABR nodes in the same area exchange LDP FECs. A PE router in a regional area, for example, PE21, learns the reachability of PE routers in other regional areas by way of RFC 8277 BGP labeled routes, which are 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 redistributing them into the core area. In this case, ABR3, which is the local ABR for PE2, may or may not set next-hop self when it redistributes 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 acts as PE router for VLL services and needs 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 redistribute 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 redistribute it to its LDP neighbors in the same area. In the application in Figure 28: Application of LDP to BGP FEC stitching, 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.

3.24.1 Configuration

> **Note:** The **no local-lsr-id** or **local-lsr-id system** commands only apply to the classic CLI. 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 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 for the following commands.

```
configure router ldp interface-parameters interface ipv4 local-lsr-id
configure router ldp interface-parameters interface ipv6 local-lsr-id
configure router ldp targeted-session peer local-lsr-id
configure router ldp targeted-session peer-template local-lsr-id
```

It is never applied to a system FEC or local static FEC. 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 toward a specified peer or within a specified targeted peer template. The FEC advertised by the following commands is automatically put in the community configured on the session.

```
configure router ldp session-parameters peer adv-local-lsr-id
configure router ldp targeted-session peer-template adv-local-lsr-id
```

The specified community is only associated with 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.

3.24.2 Detailed LDP FEC resolution

When an LSR receives a FEC-label binding from an LDP neighbor for a specific FEC1 element, the following is the LDP FEC resolution workflow.

1. LDP installs the FEC if:

   - the following command is enabled in LDP, and it was able to perform a successful exact match or a longest match of the FEC /32 prefix with a prefix entry in the routing table

   ```
   configure router ldp aggregate-prefix-match
   ```

   - the advertising LDP neighbor is the next hop to reach the FEC prefix

2. After such a FEC-label binding has been installed in the LDP FIB, LDP performs the following steps:

   a. Program a push and a swap NHLFE entries in the egress datapath to forward packets to FEC1.

   b. Program the CPM tunnel table with a tunnel entry for the NHLFE.

   c. Advertise a new FEC-label binding for FEC1 to all its LDP neighbors according to the global and per-peer LDP prefix export policies.

   d. 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, BGP performs the following steps:

   a. 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, such as the local ABR/ASBR nodes which have the following command enabled:

      - **MD-CLI**

      ```
      configure router bgp neighbor advertise-ldp-prefix
      ```

      - **classic CLI**

      ```
      configure router bgp group neighbor advertise-ldp-prefix
      ```

   b. Install the ILM entry pointing to the swap NHLFE programmed by LDP.

3.24.3 Detailed BGP labeled route resolution

When an LSR receives a BGP labeled route by way of IBGP for a specific /32 prefix, the following is the BGP labeled route resolution workflow.

1. BGP resolves and installs the route in BGP if an LDP LSP to the BGP neighbor exists, such as the ABR or ASBR that advertised it, and it is the next hop of the BGP labeled route.

2. When the BGP route is installed, BGP programs the following:

   a. a push NHLFE in the egress datapath to forward packets to this BGP labeled route

   b. 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, LDP performs the following:

   a. 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 has already advertised a FEC for the same /32 prefix after receiving it from an LDP neighbor, 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.

    b. 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, no update to the ILM entry is performed. The LDP route is always preferred over the BGP labeled route.

The following command (in the LDP context) has no effect on LDP-to-BGP stitching except for one specific case as described below.

```
configure router ldp prefer-protocol-stitching
```

Typically BGP does not add a TTM entry if the BGP-LU route is not the most preferred route in RTM. Because a BGP-LU route cannot be used for LDP FEC resolution, there are no two TTM entries to choose from, and the command has no effect. However, it is possible to program BGP-LU tunnels for prefixes available in the IGP by blocking those prefixes from the label IPv4 RIB using the following command.

```
configure router bgp rib-management label-ipv4 route-table-import
```

In this case, the **prefer-protocol-stitching** command impacts the stitching and prefers stitching to BGP instead of LDP.

> ✎ **Note:** The following BGP command, if set to a lower value than the IGP preference in the route table, overrides the IGP preference.
>
> ```
> configure router bgp label-preference
> ```

When resolving a FEC, LDP prefers the RTM over the TTM when both resolutions are possible. That is, swapping the LDP ILM to an LDP NHLFE is preferred over stitching the LDP ILM to an SR NHLFE. This behavior can be overridden by enabling the **prefer-protocol-stitching** command in the LDP context, in which case LDP prefers stitching to the SR tunnel, even if an LDP tunnel exists. This capability interacts with SR-to-LDP stitching. When SR stitches to LDP, no SR tunnel entry is added to the TTM and the command has no effect.

3.24.4 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, such as an ABR/ASBR that 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.

3.25 LDP-SR stitching for IPv4 prefixes

This feature enables stitching between an LDP FEC and SR node segment ID (SID) route for the same IPv4 /32prefix.

3.25.1 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 TTM into LDP (IGP).

In the LDP-to-SR datapath 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 following IS-IS or OSPF configuration information:

- **IS-IS (MD-CLI)**

```
configure policy-options policy-statement entry from protocol name isis
configure policy-options policy-statement entry from protocol instance
```

- **IS-IS (classic CLI)**

```
configure router policy-options policy-statement entry from protocol isis instance
```

- **OSPF (MD-CLI)**

```
configure policy-options policy-statement entry from protocol name ospf
configure policy-options policy-statement entry from protocol instance
```

- **OSPF (classic CLI)**

```
configure router policy-options policy-statement entry from protocol ospf instance
```

The preceding configuration information is added to the LDP tunnel table export policy using the following command.

```
configure router ldp export-tunnel-table
```

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 IS-IS, 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 using the following command:

- **MD-CLI**

```
configure policy-options policy-statement default-action action-type accept
```

- **classic CLI**

```
configure router policy-options policy-statement default-action accept
```

When the accept action is enabled, the LDP follows the TTM selection rules as described in the "Segment Routing Tunnel Management" in the *7705 SAR Gen 2 Unicast Routing Protocols Guide* to select a tunnel to which it stitches the LDP ILM to:

1. LDP selects the tunnel from the lowest TTM preference protocol.

2. If IS-IS and BGP protocols have the same preference, LDP uses the default TTM protocol preference to select the protocol.

3. 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 an SR tunnel of type SR IS-IS 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 a FEC for the prefix and redistributes it to its LDP and T-LDP peers. The latter allows an LDP FEC that is tunneled over an RSVP-TE LSP to have its ILM stitched to an SR tunnel endpoint. When an LDP FEC is stitched to an SR tunnel, forwarded packets benefit from the protection provided by the LFA or 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 an LDP NHLFE is preferred over stitching the LDP ILM to an SR NHLFE. This behavior can be overridden by enabling the **prefer-protocol-stitching** command in the LDP context, in which case LDP prefers stitching to the SR tunnel, even if an LDP tunnel exists. This capability interacts with SR-to-LDP stitching. When SR stitches to LDP, no SR tunnel entry is added to the TTM and the command has no effect.

> ✏️ **Note:** Forcing the stitching to SR affects forwarding at the LER and LSR roles. Typically, a specific prefix has a "push" and a "swap" binding for the LER and LSR roles, respectively. When **prefer-protocol-stitching** is enabled, the "swap" binding points to an SR tunnel and the "push" binding is removed. Services using the LDP tunnel should use the SR tunnel instead.

In the SR-to-LDP datapath direction, the SR mapping server provides a global policy for prefixes corresponding to the LDP FECs the SR needs to stitch to. For more information, see "Segment routing mapping server" in the *7705 SAR Gen 2 Unicast Routing Protocols Guide*. 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:

```
configure router isis segment-routing export-tunnel-table ldp
configure 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, forwarded packets benefit from the FRR protection of the LFA backup next hop of the LDP FEC.

When resolving a node SID, IGP prefers 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 an SR NHLFE is preferred over stitching it to an LDP tunnel endpoint. For more information about prefix SID resolution, see "Segment routing mapping server prefix SID resolution" in the *7705 SAR Gen 2 Unicast Routing Protocols Guide*.

Nokia recommends enabling the BFD option on the interfaces in both LDP and IGP instance contexts to speed up the failure detection and the activation of the LFA or 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.

3.25.2 Stitching in the LDP-to-SR direction

Stitching in the data plane from the LDP-to-SR direction is based on the LDP module monitoring the TTM for an SR tunnel of a prefix matching an entry in the LDP TTM export policy.

Figure 29: Stitching in the LDP-to-SR direction
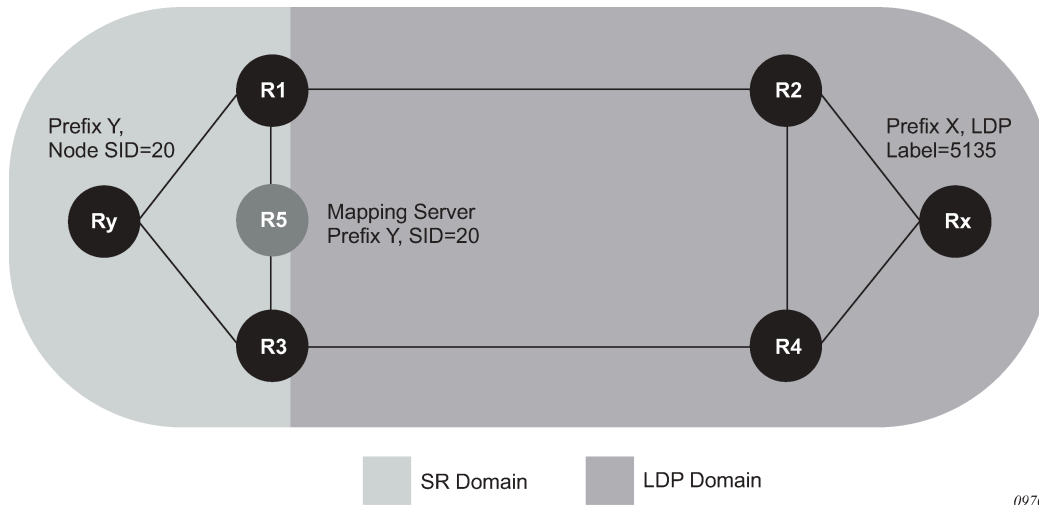
In the preceding figure, the boundary router R1 performs the following procedure to effect stitching.

1. Router R1 is at the boundary between an SR domain and LDP domain and is configured to stitch between SR and LDP.

2. Link R1-R2 is LDP-enabled, but router R2 does not support SR (or SR is disabled).

3. Router R1 receives a prefix-SID sub-TLV in an IS-IS IP reachability TLV originated by router Ry for prefix Y.

4. R1 resolves the prefix-SID and programs an NHLFE on the link toward the next hop in the SR domain. R1 programs an SR ILM and points it to this NHLFE.

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.

6. R1 advertises the LDP FEC for the prefix Y to all its LDP peers. R2 is now able to install an LDP tunnel toward Ry.

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.

   a. R1 selects the tunnel from the lowest preference IGP protocol.

   b. R1 selects the protocol using the default TTM protocol preference.

   c. Within the same IGP protocol, R1 uses the lowest instance ID to select the tunnel.

8. If the user concurrently configured BGP, IS-IS, and OSPF from protocol statements (as follows) in the same LDP tunnel table export policy, or did not include the from statement but added a default action of accept, R1 selects the tunnel to destination prefix Y to stitch the LDP ILM to using the TTM tunnel selection rules.

   - **MD-CLI**

   ```
   configure policy-options policy-statement entry from protocol name {bgp | isis | ospf}
   ```

- **classic CLI**

```
configure router policy-options policy-statement entry from protocol {bgp | isis | ospf}
```

The TTM tunnel selection rules are as follows:

a. R1 selects the tunnel from the lowest preference protocol.

b. 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.

c. Within the same IGP protocol, R1 uses the lowest instance ID to select the tunnel.

> ✏️ **Note:** If R1 has already resolved an LDP FEC for prefix Y, it has an ILM for it, but this ILM is not updated to point toward 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.

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 to forward packets over the SR tunnel to router Ry. No processing occurs in R1 because the SR ILM is already programmed.

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 toward Ry.

3.25.3 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 an LDP tunnel of a prefix matching an entry in the SR TTM export policy.

In Figure 29: Stitching in the LDP-to-SR direction, the boundary router R1 performs the following procedure to effect stitching:

1. Router R1 is at the boundary between an SR domain and an 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).

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.

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 toward R2 with egress label L2.

4. Later on, R1 receives a prefix-SID sub-TLV from the mapping server R5 for prefix X.

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 SR and, therefore, SID resolution for prefix X in routing table fails.

6. IGP in R1 attempts to resolve prefix SID of X in TTM because it is configured to stitch SR-to-LDP. R1 finds an LDP tunnel to X in TTM, instructs the SR module to program an 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 is 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 that failed resolution was part of mapping server TLV or a prefix TLV.

7. The user enables SR on R2, which causes IGP in R1 to discover 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.

8. 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.

9. This triggers the re-evaluation of the SIDs. R1 first attempts the resolution in routing table and because 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 an 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.

10. Later, router Rx, which owns prefix X, is upgraded to support SR. R1 now receives a prefix-SID sub-TLV in an IS-IS 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 prefers the prefix-SID originated by Rx and updates the SR ILM and NHLFE with appropriate labels.

11. Finally, the operator cleans up the mapping server and removes the mapping entry for prefix X, which gets withdrawn by IS-IS.

3.26 LDP FRR LFA backup using SR tunnel for IPv4 prefixes

The user enables the use of an SR tunnel as a remote LFA or as a TI-LFA backup tunnel next hop by an LDP FEC via the following command.

```
configure router ldp fast-reroute backup-sr-tunnel
```

As a prerequisite, the user must enable the stitching of LDP and SR in the LDP-to-SR direction as described 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** command option is enabled in LDP and **remote-lfa** or **ti-lfa**, or both are enabled by the IGP instance:

- **MD-CLI**

```
configure router isis loopfree-alternate remote-lfa
configure router ospf loopfree-alternate remote-lfa
configure router isis loopfree-alternate ti-lfa
configure router ospf loopfree-alternate ti-lfa
```

- **classic CLI**

```
configure router isis loopfree-alternates remote-lfa
configure router ospf loopfree-alternates remote-lfa
configure router isis loopfree-alternates ti-lfa
configure router ospf loopfree-alternates ti-lfa
```

and that LDP was able to resolve the primary next hop of the LDP FEC in RTM. IGP SPF runs 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 also runs 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 LDP-SR stitching for IPv4 prefixes, 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 effectively points to the primary next hop of the SR tunnel and traffic of the LDP ILM/LTN follows 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** command 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.

3.27 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. See "Remote LFA with Segment Routing" in the *7705 SAR Gen 2 Unicast Routing Protocols Guide* for more information about rLFA computation. In SR OS, a repair tunnel can also be an SR tunnel; this section describes an LDP-in-LDP tunnel.

As a prerequisite for LDP rLFA configuration, enable Remote LFA computation using the following commands:

- **MD-CLI**

```
configure router isis loopfree-alternate remote-lfa
configure router ospf loopfree-alternate remote-lfa
```

- **classic CLI**

```
configure router isis loopfree-alternates remote-lfa
configure router ospf loopfree-alternates remote-lfa
```

Enable attaching rLFA information to RTM entries using the following commands:

- **MD-CLI**

```
configure router isis loopfree-alternate augment-route-table
configure router ospf loopfree-alternate augment-route-table
```

- **classic CLI**

```
configure router isis loopfree-alternates augment-route-table
configure router ospf loopfree-alternates augment-route-table
```

The preceding command attaches rLFA-specific information to route entries that are necessary for LDP to program repair tunnels toward the PQ node using a specific neighbor.

In addition, enable tunneling on both the PQ node and the source node using the following command.

```
configure router ldp targeted-session peer tunneling
```

The following figure shows the general principles of LDP rLFA operation.

Figure 30: General principles of LDP rLFA operation
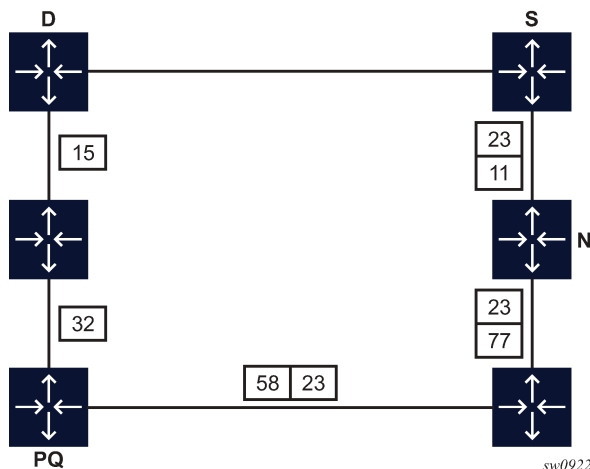
In the preceding figure, S is the source node and D is the destination node. The primary path is the direct link between S and D. The rLFA algorithm has determined the PQ node. In the event of a failure between S and D, for traffic not to loopback to S, the traffic must be sent directly to the PQ node. 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 toward the PQ nodes that the rLFA algorithm has identified). These targeted sessions must be set up with router IDs that match the IDs the rLFA algorithm uses. LDP rLFA is designed to operate in LDP-only environments; therefore, LDP does not establish rLFA backups in the presence of LDP over RSVP-TE or LDP over SR-TE tunnels. The following OAM command is not supported over the repair tunnels.

```
oam lsp-trace
```

3.28 Automatic LDP rLFA

The manual LDP rLFA configuration method requires the user to specify beforehand, on each node, the list of peers that are used to establish a targeted session. 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 30: General principles of LDP rLFA operation, considering a failure on the shortest path between S and D nodes, S needs a targeted LDP session toward 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:

- **MD-CLI**

```
configure router isis loopfree-alternate remote-lfa
```

- **classic CLI**

```
configure router isis loopfree-alternates remote-lfa
```

Enable attaching rLFA information to RTM entries using the following command:

- **MD-CLI**

```
configure router isis loopfree-alternate augment-route-table
```

- **classic CLI**

```
configure router isis loopfree-alternates augment-route-table
```

In the Figure 30: General principles of LDP rLFA operation scenarios, because the S node requires the T-LDP session, it should initiate the T-LDP session request. The PQ node receives the request for this session. Therefore, S node configuration is as follows.

Example: MD-CLI

```
[ex:/configure router "Base" ldp targeted-session auto-tx ipv4]
A:admin@node-2# info
    admin-state enable
    tunneling false
```

Example: classic CLI

```
A:node-2>config>router>ldp>targ-session>auto-tx>ipv4# info
----------------------------------------------
                        no shutdown
----------------------------------------------
```

And PQ node configuration is as follows:

Example: MD-CLI

```
[ex:/configure router "Base" ldp targeted-session auto-tx ipv4]
A:admin@node-2# info
    admin-state enable
    tunneling true
```

Example: classic CLI

```
A:node-2>config>router>ldp>targ-session>auto-tx>ipv4# info
----------------------------------------------
                        tunneling
                        no shutdown
----------------------------------------------
```

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 automatically kills the session to the previous PQ node and establish a new one (toward the new PQ node).

> **Note:** It is not possible to configure command options specifically for automatic T-LDP sessions. The system inherits command options, either those defined for the IPv4 family (under targeted-session) or the default command options of the system. This applies to the following configurations.
>
> ```
> configure router ldp targeted-session ipv4 hello
> configure router ldp targeted-session ipv4 hello-reduction
> configure router ldp targeted-session ipv4 keepalive
> ```

Also, the automatic T-LDP session can use parameters defined for the following configuration if the specified address is the router ID of the peer.

```
configure router ldp tcp-session-parameters peer-transport
```

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 as follows:

```
configure router ldp targeted-session auto-tx
configure router ldp targeted-session auto-rx
```

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 (item 1 being highest priority and item 5 being lowest):

1. peer

2. template

3. auto-tx

4. auto-rx

5. sdp

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 establishes 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 does not work because the PQ node does 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 then acts 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 only works if tunneling is enabled under **auto-tx**. The configuration from which the session command options are taken is indicated in the following command ("creator" label).

```
show router ldp targ-peer detail
```

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
```

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 verifies 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 following command.

```
tools dump router ldp timers
```

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

3.29 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.

3.29.1 Feature configuration

The user first creates a targeted LDP session peer parameter template by using the following command.

```
configure router ldp targeted-session peer-template
```

Inside the template the user configures the common T-LDP session command options shared by all peers using this template with the following commands:

- **MD-CLI**

```
configure router ldp targeted-session peer-template bfd-liveness
configure router ldp targeted-session peer-template hello
configure router ldp targeted-session peer-template hello-reduction
configure router ldp targeted-session peer-template keepalive
configure router ldp targeted-session peer-template local-lsr-id
configure router ldp targeted-session peer-template tunneling
```

- **classic CLI**

```
configure router ldp targeted-session peer-template bfd-enable
configure router ldp targeted-session peer-template hello
configure router ldp targeted-session peer-template hello-reduction
configure router ldp targeted-session peer-template keepalive
configure router ldp targeted-session peer-template local-lsr-id
configure router ldp targeted-session peer-template tunneling
```

The tunneling option does not support adding explicit RSVP LSP names. LDP selects 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 using the following command:

- **MD-CLI**

```
configure router ldp targeted-session peer-template-map template-map-name
configure router ldp targeted-session peer-template-map policy-map
```

- **classic CLI**

```
configure router ldp targeted-session peer-template-map peer-template policy
```

Each application of a targeted session template to a specific prefix in the prefix list results 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 either triggers a new LDP session or is 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 specific peer prefix. In all cases, the targeted session parameters applied to a specific 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 tells 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 enabled state and therefore it takes effect immediately. After 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.

3.29.2 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 following command option in IS-IS or OSPF.

```
configure router isis traffic-engineering
configure router ospf traffic-engineering
```

Each mapping of a targeted session peer parameter template to a policy prefix which exists in the TE database results in LDP establishing a targeted Hello adjacency to this peer address using the targeted session parameters configured in the template. This Hello adjacency then either gets associated with an LDP session to the peer if one exists or it triggers 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 command options inherited from the following top level context.

    ```
    configure router ldp targeted-session ipv4
    ```

    User configuration of the peer with the targeted session command options explicitly configured for this peer in the following context and which overrides the top level command options shared by all targeted peers.

    ```
    configure router ldp targeted-session peer
    ```

    This allows us to refer to the top level configuration context as the global context. Some command options only exist in the global context; their value is always inherited by all targeted peers regardless of which event triggered it.

*   User configuration of an SDP of any type to a peer with the following command enabled (default configuration). In this case the targeted session command option values are taken from the global context.

    ```
    configure service sdp signaling tldp
    ```

*   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 multisegment 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 command option 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 command options or configure a mapping of a targeted session peer parameter template to a prefix policy. These features are the following:

    –   BGP shortcut

        ```
        configure router bgp next-hop-resolution shortcut-tunnel
        ```

    –   IGP shortcut

        ```
        configure router isis igp-shortcut
        configure router ospf igp-shortcut
        configure router ospf3 igp-shortcut
        ```

    –   LDP shortcut for IGP routes

        *   **MD-CLI**

            ```
            configure router ldp ldp-shortcut
            ```

- • **classic CLI**

```
configure router ldp-shortcut
```

- – static route LDP shortcut (**ldp** option in a static route)
  - • **MD-CLI**

```
configure router static-routes route indirect tunnel-next-hop resolution-filter ldp
```

  - • **classic CLI**

```
configure router static-route-entry indirect tunnel-next-hop resolution-filter ldp
```

- – VPRN service

```
configure service vprn bgp-ipvpn mpls auto-bind-tunnel resolution-filter ldp
configure service vprn bgp-evpn mpls auto-bind-tunnel resolution-filter ldp
```

Because the above triggering events can occur simultaneously or in any arbitrary order, the LDP code implements a priority handling mechanism to decide which event overrides the active targeted session parameters. The overriding trigger becomes the owner of the targeted adjacency to a specific peer and is displayed using the following command.

```
show router ldp targ-peer
```

Table 13: Targeted LDP adjacency triggering events and priority summarizes the triggering events and the associated priority.

Table 13: 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	—

Triggering event	Automatic creation of targeted Hello adjacency	Active targeted adjacency parameter override priority
LDP-over-RSVP in BGP Labeled 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** value. In this case, the Hello adjacency is brought down which also causes 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 then get established to the peer using the new value of the local LSR ID.

- The triggering event caused the targeted peer to become disabled. In this case, the Hello adjacency is brought down which also causes 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 following context.

```
configure router ldp session-parameters peer
```

This includes MD5 authentication, LDP prefix per-peer policies, label distribution mode (DU or DOD), and so on.

3.30 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 toward 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 31: Video distribution using P2MP LDP 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 31: Video distribution using P2MP LDP
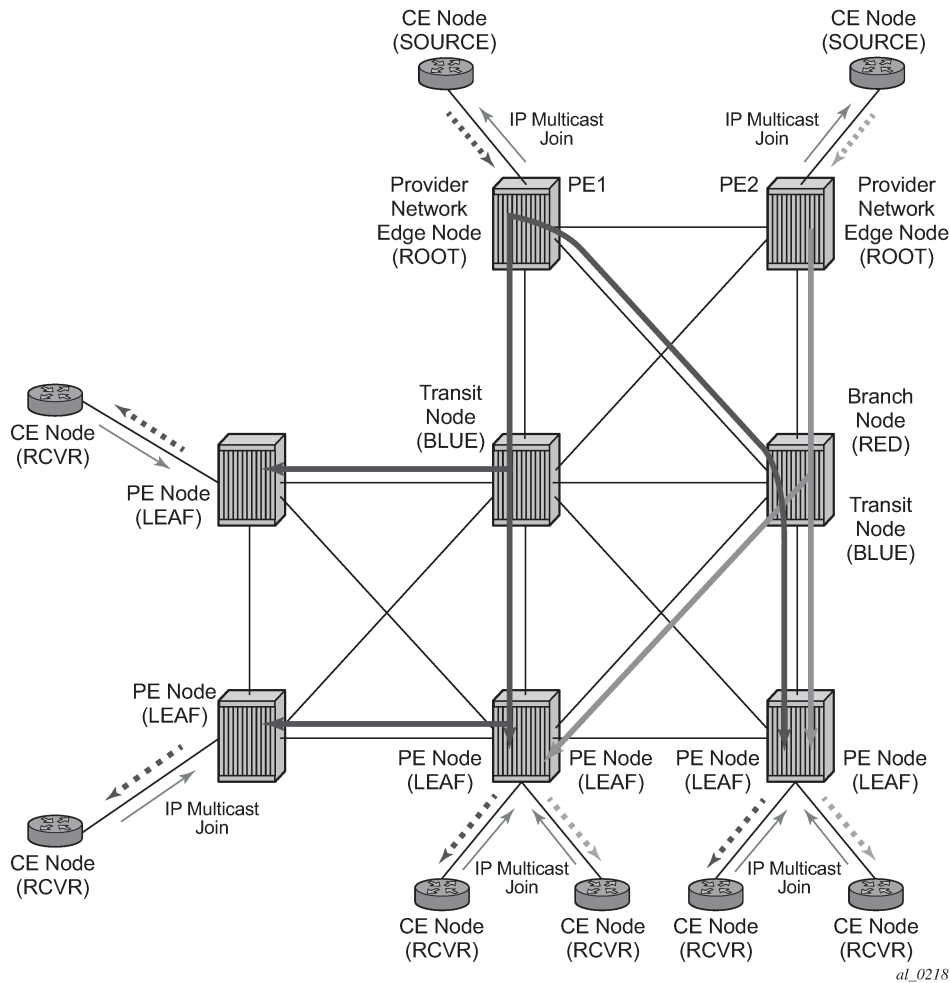
3.31 LDP P2MP support

3.31.1 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 toward 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.

3.31.2 LDP P2MP protocol

Only a single generic identifier range is defined for signaling multipoint tree for all client applications. Implementation on the 7705 SAR Gen 2 reserves the range (1..8292) of generic LSP P2MP-ID on root node for static P2MP LSP.

3.31.3 MBB

When a transit or leaf node detects that the upstream node toward 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.

3.31.4 ECMP support

In the following figure, the leaf discovers the ROOT-1 from all three ASBRs (ASBR-3, ASBR-4 and ASBR-5).

Figure 32: ECMP support

The leaf uses the following process to choose the ASBR used for the multicast stream:

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 (**configure router ecmp**). For example, if the ECMP value is set to 2, only two of the ASBRs are 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 **configure 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 32: ECMP support, 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 32: ECMP support, ASBR-5 goes 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 is 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 specific LSP.

3.31.5 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 because of 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.

3.31.5.1 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 actual root (the root with the source connected to it) and the leaf are not in the same AS or area. In these cases, on the leaf, the actual root is resolved using a local root which is usually the ASBR or the ABR. As such, the intermediate routers in the leaf AS or area do not know anything about the actual 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 33: Inter-AS Option C 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 more information needed to reach the root from ASBR-3. As a result, the SR OS 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 33: Inter-AS Option C
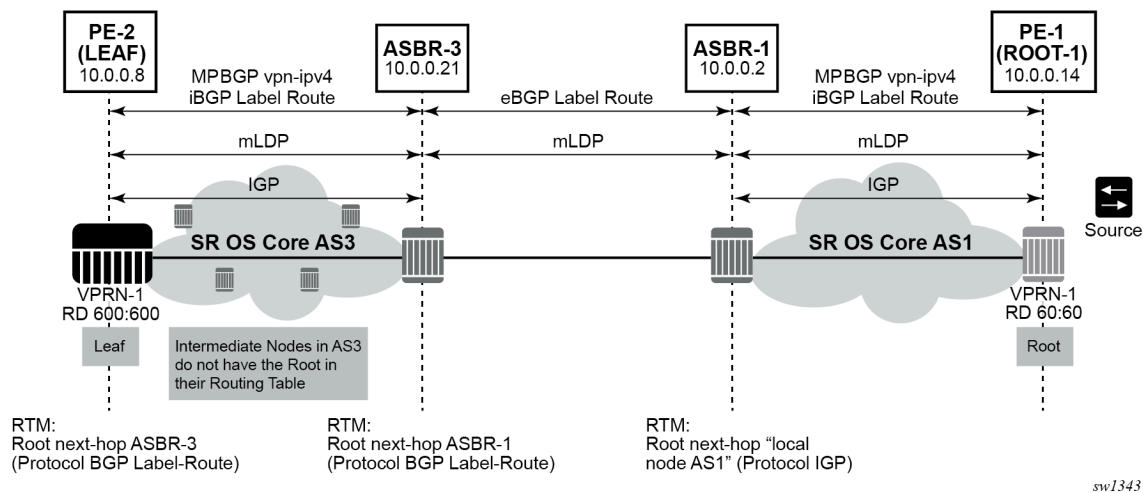
Because non-segmented d-mLDP requires end-to-end mLDP signaling, the ASBRs support both mLDP and BGP signaling between them.

3.31.5.2 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. The following figure shows this FEC representation.

Figure 34: mLDP FEC for single AS with transit IPv4 opaque
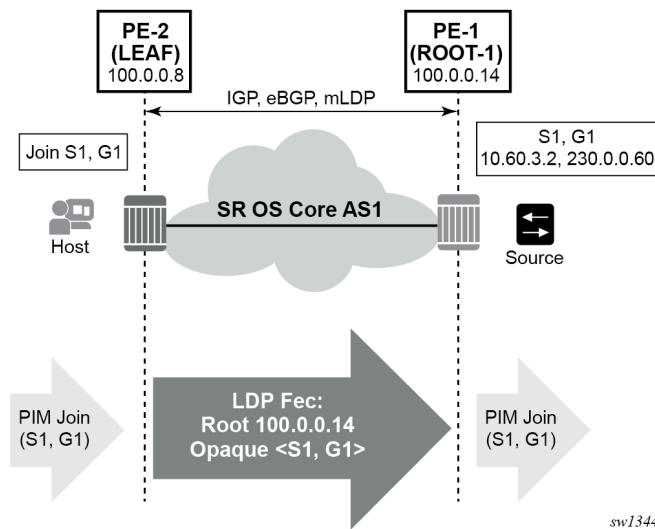
The following figure shows an inter-AS FEC with recursive opaque based on RFC 6512.

Figure 35: mLDP FEC for inter-AS with recursive opaque value
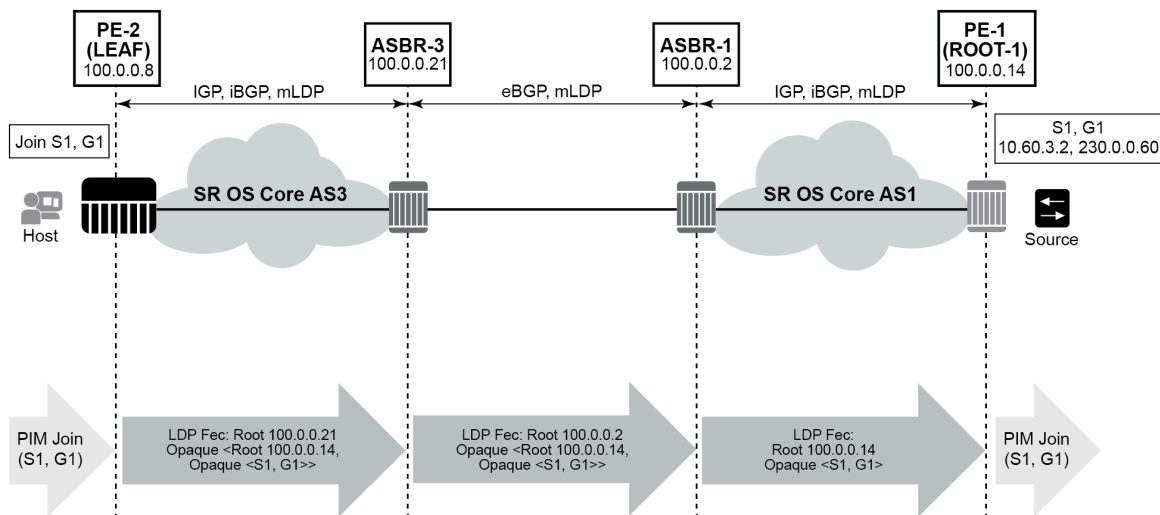
As shown in the preceding figure, the root "10.0.0.21" is an ASBR and the opaque value contains the original mLDP FEC. As such, in the AS of the leaf where the actual root "10.0.0.14" is not known, the LDP FEC can be routed using the local root of ASBR. When the FEC arrives at an ASBR that co-locates 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 the following figure.

Figure 36: Non-VPN mLDP with recursive opaque for inter-AS
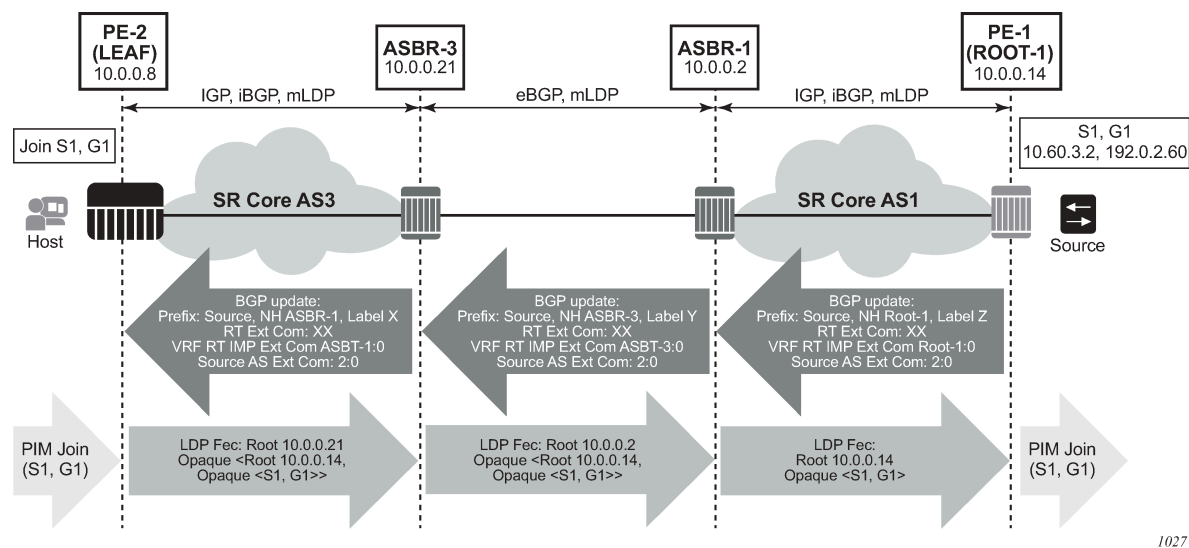
As shown in the preceding figure, the leaf is in AS3s where the AS3 intermediate nodes do not have the ROOT-1 in their RTM. The leaf has the S1 installed in the RTM via BGP. All ASBRs are acting as next-hop-self in the BGP domain. The leaf resolving the S1 via BGP generates an mLDP FEC with recursive opaque, represented as:

  Leaf FEC: <Root=ASBR-3, opaque-value=<Root=Root-1, <opaque-value = S1,G1>>>

This FEC is 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 is changed to:

  ASBR-3 FEC: <Root=ASBR-1, opaque-value=<Root=Root-1, <opaque-value = S1,G1>>>

This FEC is routed from ASBR-3 to ASBR-1. ASBR-1 is colocated 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.

3.31.5.3 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.

The following table lists all supported opaque values in SR OS.

Table 14: 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>>

3.31.5.4 Optimized Option C and basic FEC generation for inter-AS

Not all leaf nodes can support labeled route or recursive opaque, so recursive opaque functionality can be transferred from the leaf to the ASBR, as shown in Figure 37: Optimized Option C — leaf router not responsible for recursive FEC.

Figure 37: Optimized Option C — leaf router not responsible for recursive FEC
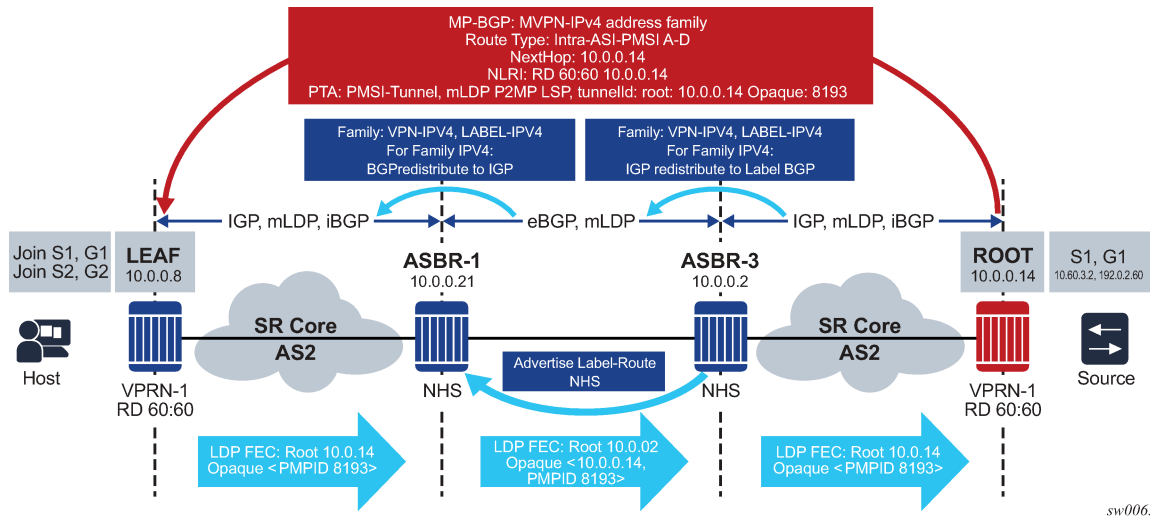
In Figure 37: Optimized Option C — leaf router not responsible for recursive FEC, the root advertises its unicast routes to ASBR-3 using IGP, and the ASBR-3 advertises these routes to ASBR-1 using label-BGP. ASBR-1 can redistribute these routes to IGP with next-hop ASBR-1. The leaf resolves the actual root 10.0.0.14 using IGP and creates a type 1 opaque value <Root 10.0.0.14, Opaque <8193>> to ASBR-1. In addition, all P routers in AS 2 know how to resolve the actual root because of BGP-to-IGP redistribution within AS 2.

ASBR-1 attempts to resolve the 10.0.0.14 actual route via BGP, and creates a recursive type 7 opaque value <Root 10.0.0.2, Opaque <10.0.0.14, 8193>>.

3.31.5.5 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 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 38: Example AS shows an example of this situation.

Figure 38: Example AS
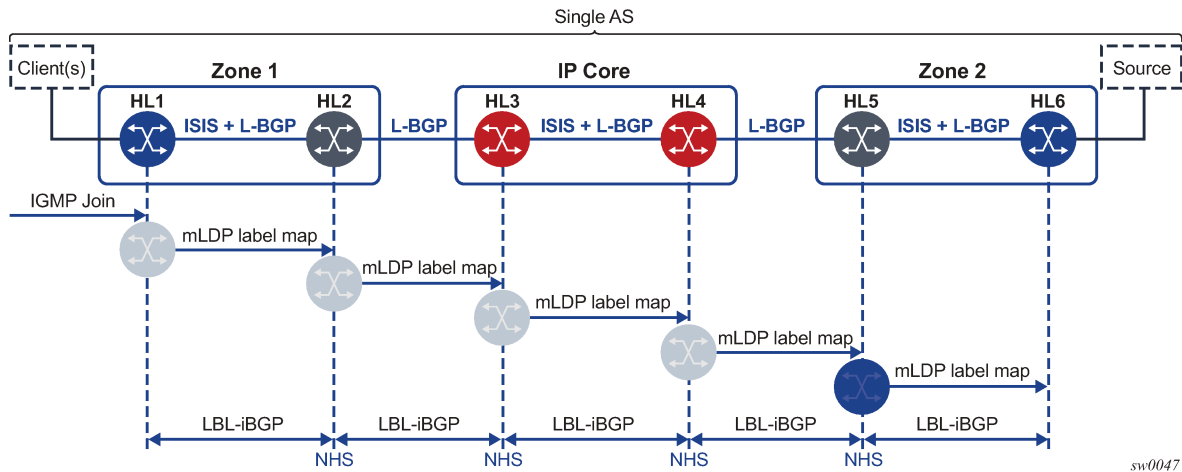
In Figure 38: Example AS, the leaf HL1 is directly attached to ABR HL2, and ABR HL2 is directly attached to ABR HL3. In this case, it is possible to generate a non-recursive opaque simply because there is no P router that cannot resolve the root PE loopback IP address in between any of the elements. 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 following global command disables recursive opaque FEC generation when the provider needs basic opaque FEC generation on the node.

```
configure router ldp generate-basic-fec-only
```

In Figure 38: Example AS, 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 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 generates 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 is not resolved and forwarded upstream, and the solution fails.

3.31.5.5.1 Leaf and ABR behavior

When the following command is enabled on a leaf node, LDP generates a basic opaque type 1 FEC.

```
configure router ldp generate-basic-fec-only
```

When **generate-basic-fec-only** is enabled on the ABR, LDP accepts 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:

1. For recursive FEC with type 7 opaque, the ABR stitches the lower FEC to an upper FEC with basic opaque type 1.

2. For any FEC type other than a recursive FEC with type 7 opaque or a non-recursive FEC with type 1 basic opaque, ABR processes the packet in the same manner as when **generate-basic-fec-only** is disabled.

3.31.5.5.2 Intra-AS support

ABR uses IBGP and peers between the system IP or loopback IP addresses, as shown in Figure 39: ABR and IBGP.

Figure 39: ABR and IBGP
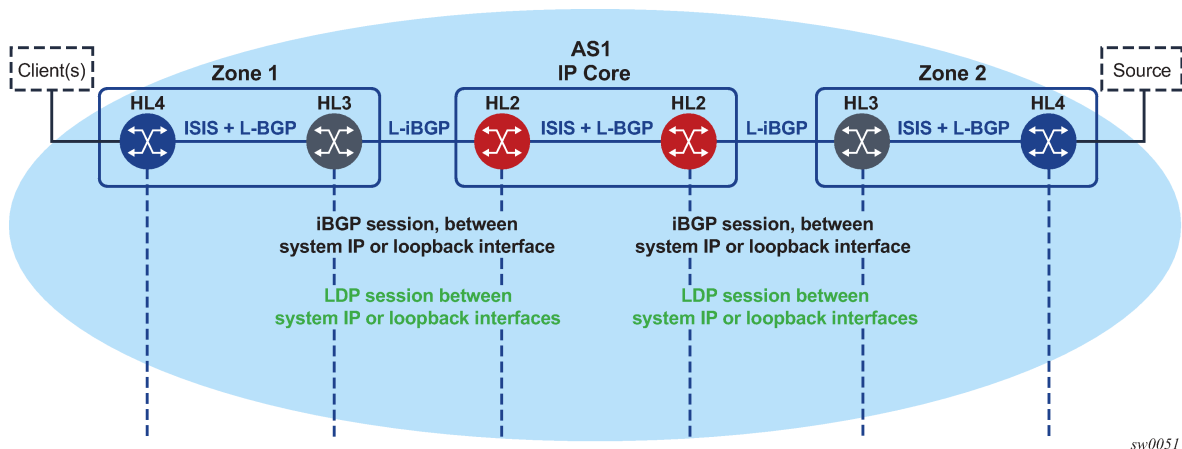
The **generate-basic-fec-only** command is supported on leaf PE and ABR nodes. The **generate-basic-fec-only** command only interoperates with intra-AS as option C, or opaque type 7 with inner opaque type 1. No other opaque type is supported.

3.31.5.5.3 Opaque type behavior with basic FEC generation

Table 15: Opaque type behavior with basic FEC generation describes the behavior of different opaque types when the **generate-basic-fec-only** command is enabled or disabled.

Table 15: 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

3.31.5.5.4 Inter-AS support

In the inter-AS case, the ASBRs use EBGP as shown in Figure 40: ASBR and EBGP.

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 39: ABR and IBGP.

Figure 40: ASBR and EBGP
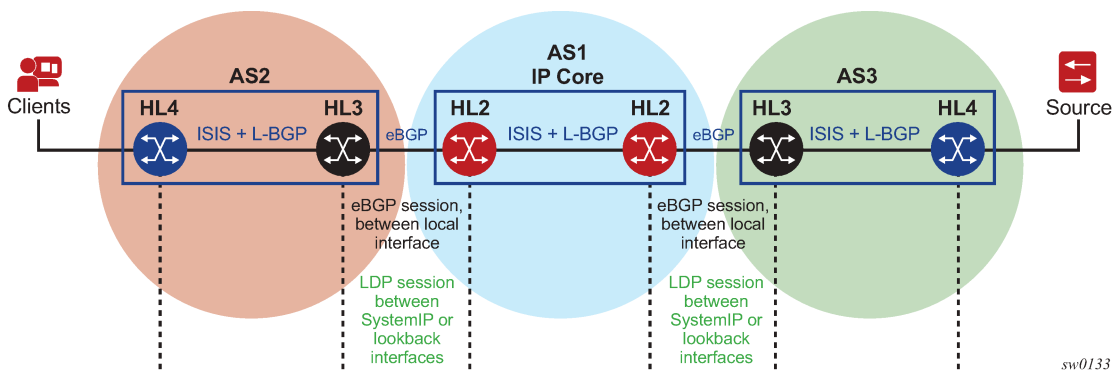
The **generate-basic-fec-only** command is supported on LEAF PE and ASBR nodes in case of inter-AS. The **generate-basic-fec-only** command only interoperates with inter-AS as option C and opaque type 7 with inner opaque type 1.

3.31.5.6 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 use EBGP without IGP, as shown in the following figure.

Figure 41: ASBRs using EBGP without IGP
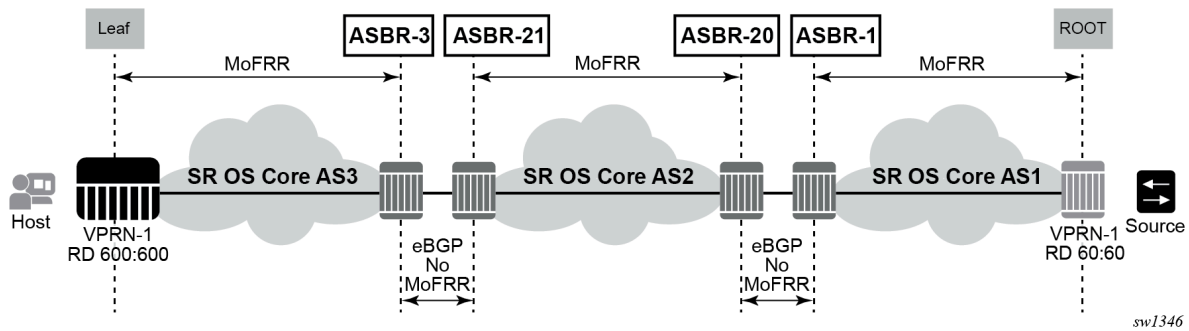
3.31.5.7 ASBR physical connection

Non-segmented mLDP functions with ASBRs directly connected or connected via an IGP domain, as shown in Figure 41: ASBRs using EBGP without IGP.

3.31.5.8 OAM

> **Note:** The **oam p2mp-lsp-ping** command only applies to the classic CLI.

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 **oam p2mp-lsp-ping** command, 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, as shown in the following figure.

Figure 42: ECHO request target FEC Stack TLV
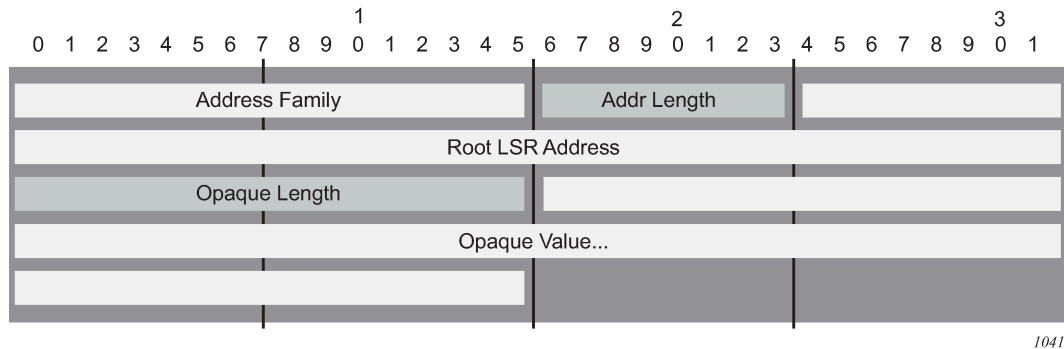
The same concept applies to inter-AS and non-segmented mLDP. The leafs in the remote AS should be able to resolve the root via GRT routing. This is possible for inter-AS Option C where the root is usually in the leaf RTM, which is a next-hop ASBR.

For inter-AS Option B where the root is not present in the leaf RTM, the echo reply cannot be forwarded to the root using the GRT. 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 using VPRN.

Figure 43: MVPN inter-AS Option B OAM
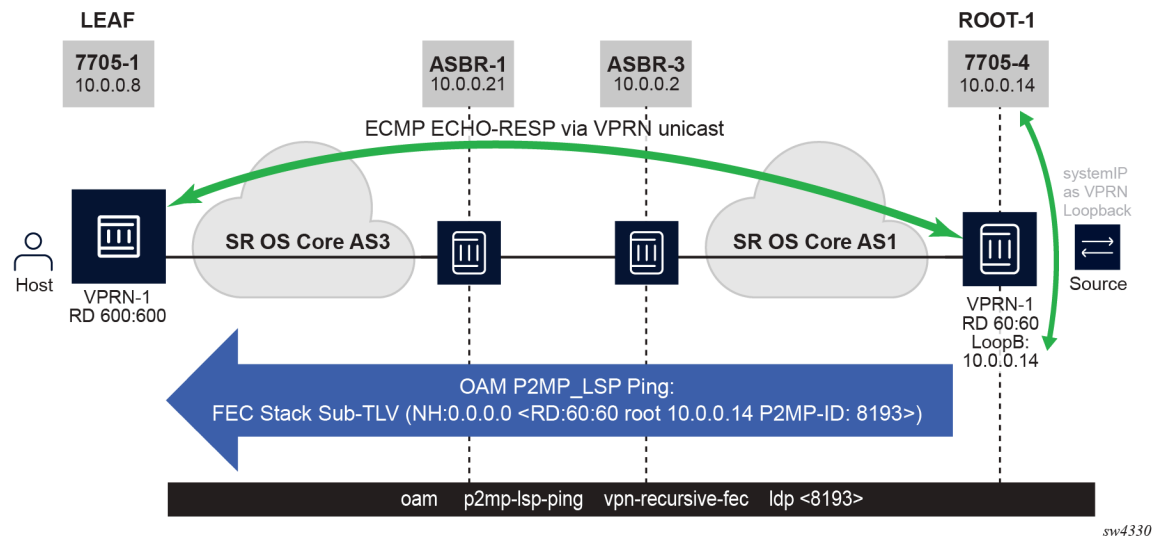
Note: The **vpn-recursive-fec** command option only applies to the classic CLI.

As shown in the preceding figure, the echo request for VPN recursive FEC is generated from the root node by executing the **oam p2mp-lsp-ping** with the **vpn-recursive-fec** command option. When the echo request reaches the leaf, the leaf uses the sub-TLV within the echo request to identify the corresponding VPN using the FEC, which includes the RD, root, and P2MP ID.

After identifying the VPRN, the echo response is sent back via the VPRN and unicast routes. A unicast route (for example, root 10.0.0.14, as shown in Figure 43: MVPN inter-AS Option B OAM) must be 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 16: OAM functionality for Options B and C.

> **Note:**
> - For SR OS, all P2MP mLDP FEC types respond to the **vpn-recursive-fec** echo request. Leafs in the local-AS and inter-AS Option C respond to the recursive-FEC TLV echo request in addition to the leafs in the inter-AS Option B.
>
>   For non inter-AS Option B where the root system IP is visible through the GRT, the echo reply is sent via the GRT, that is, not via the VPRN.
>
> - This **vpn-recursive-fec** is a Nokia proprietary implementation, and therefore third-party routers do not recognize the recursive FEC and do not generate an echo respond.
>
>   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.

> **Note:** The information in the following table only applies to the classic CLI.

Table 16: 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			

3.31.5.9 ECMP support

In the following figure, the leaf discovers the ROOT-1 from all three ASBRs (ASBR-3, ASBR-4 and ASBR-5).

Figure 44: ECMP support

The leaf uses the following process to choose the ASBR used for the multicast stream:

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 (**configure router ecmp**). For example, if the ECMP value is set to 2, only two of the ASBRs are 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 **configure 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 44: ECMP support, 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 44: ECMP support, ASBR-5 goes 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 is 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 specific LSP.

3.31.5.9.1 ECMP hash algorithm

The ECMP hash algorithm requires the opaque value of the FEC (see ECMP hash algorithm) 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 and "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.

3.31.5.10 Dynamic mLDP and static mLDP coexisting on the same node

When creating a static mLDP tunnel, use the commands in the following context to configure the P2MP tunnel ID.

```
configure router tunnel-interface
```

This P2MP ID can coincide with a dynamic mLDP P2MP ID. The dynamic mLDP is created by the PIM automatically without manual configuration. If the node has a static and dynamic mLDP with same label and P2MP ID, there are collisions and OAM errors.

Do not use a static and dynamic mLDP on the 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 is not forwarded downstream. A static mLDP FEC is only 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 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.

The following figure shows 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 45: Static and dynamic mLDP interaction
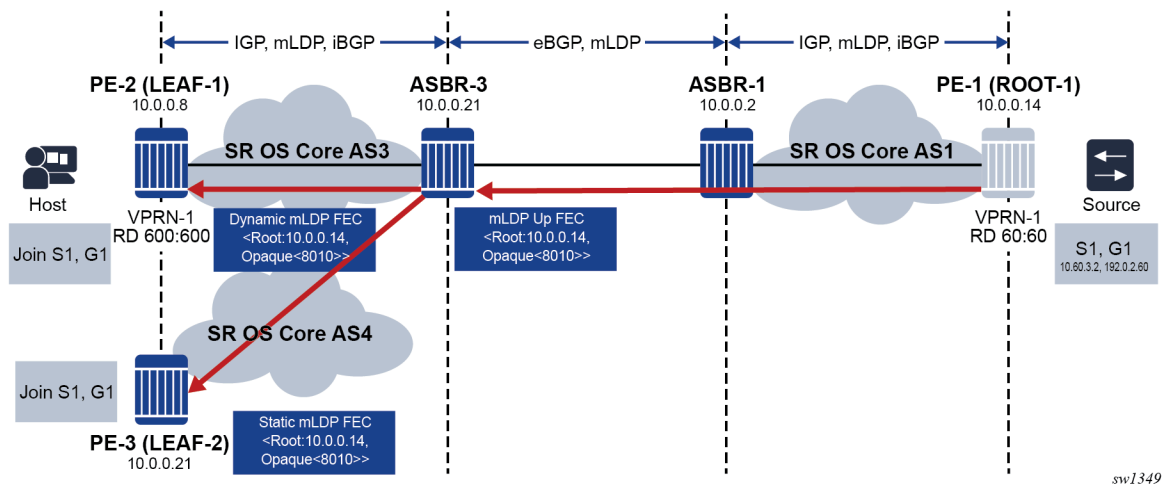
In the preceding figure, both FECs generated by LEAF-1 and LEAF-2 are identical, and the ASBR-3 merges the FECs into a single upper FEC. Any traffic arriving from ROOT-1 to ASBR-3 over VPRN-1 is forked to LEAF-1 and LEAF-2, even if the tunnels were signaled for different services.

3.31.6 Intra-AS non-segmented mLDP

Non-segmented mLDP intra-AS (inter-area) is supported on option B and C only. Figure 46: Intra-AS non-segmented topology 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 labeled routes, which requires recursive FEC.

Figure 46: Intra-AS non-segmented topology
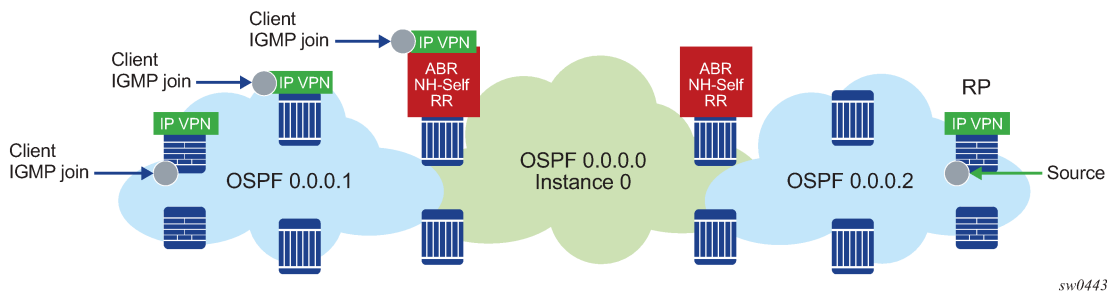
For option B, the ABR routers change the next hop of the MVPN AD routes to the ABR system IP or Loopback IP. The following commands for BGP do not change the next hop of the MVPN AD routes.

```
configure router bgp group next-hop-self
configure router bgp group neighbor next-hop-self
configure service vprn bgp group next-hop-self
configure service vprn bgp group neighbor next-hop-self
```

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 next hop at the ABR.

3.31.6.1 ABR MoFRR for intra-AS

With ABR MoFRR in the intra-AS environment, the leaf chooses 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 supports 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.

3.31.6.2 Interaction with an inter-AS non-segmented mLDP solution

Intra-AS option C is supported in conjunction to inter-AS option B or C. Intra-AS option C with inter-AS option B is not supported.

3.31.6.3 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 following command.

```
configure router ldp import-pmsi-routes mvpn
```

When enabled, LDP requests BGP to provide the LDP task with all of the MVPN PMSI AD routes and LDP caches these routes internally. If **import-pmsi-routes mvpn** is disabled, MVPN discards 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.

3.31.7 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 47: BGP neighboring for MoFRR) and a backup LSP signaled from the leaf node to the second local root (ASBR-2 in Figure 47: BGP neighboring for MoFRR) through the best IGP path in the AS.

Using Figure 47: BGP neighboring for MoFRR 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 47: BGP neighboring for MoFRR
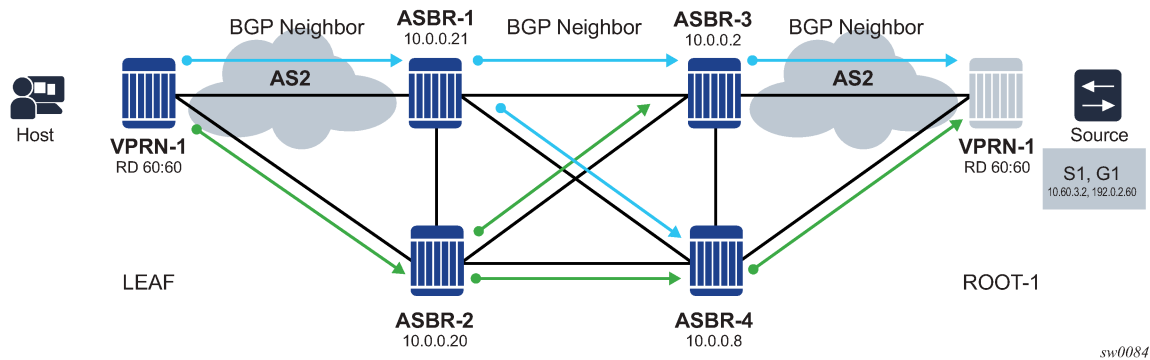
In Figure 47: BGP neighboring for MoFRR, ASBR-2 is a disjointed ASBR; with the AS spanning from the leaf to the local root, which is the ASBR selected in the AS, the traditional IGP MoFRR is used. ASBR MoFRR is used from the leaf node to the local root, and IGP MoFRR is used for any P router that connects the leaf node to the local root.

3.31.7.1 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 48: ASBR node acting as local leaf.

Figure 48: ASBR node acting as local leaf
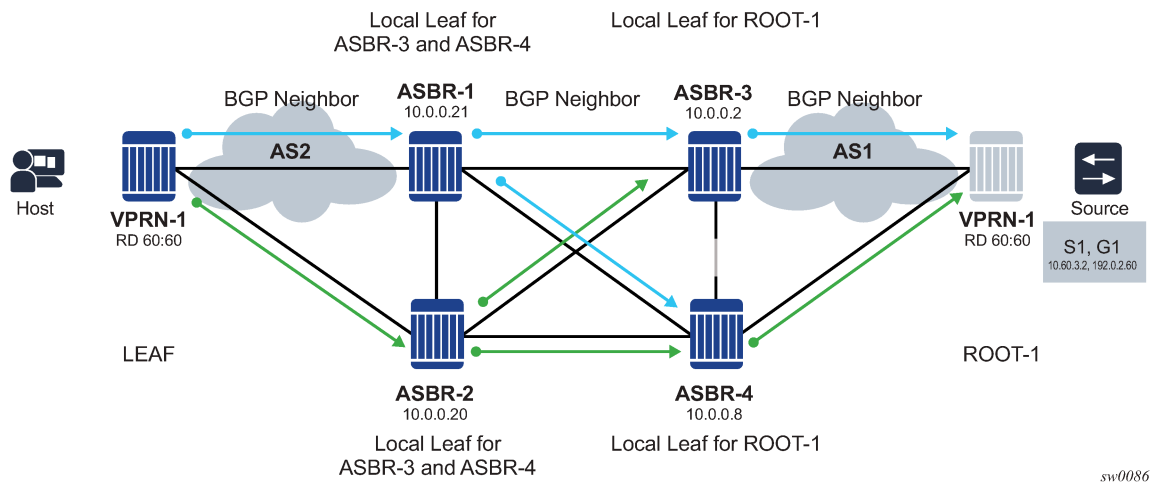
Two types of MoFRR can exist in a unique AS:

•   **IGP MoFRR**

When the following command is enabled for LDP, the local leaf selects a single local root, either ASBR or actual, and creates a FEC toward two different upstream LSRs using LFA/ECMP for the ASBR route.

```
configure router ldp mcast-upstream-frr
```

If there are multiple ASBRs directed toward the actual root, the local leaf only selects a single ASBR; for example, ASBR-1 in Figure 49: IGP MoFRR. 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 49: IGP MoFRR
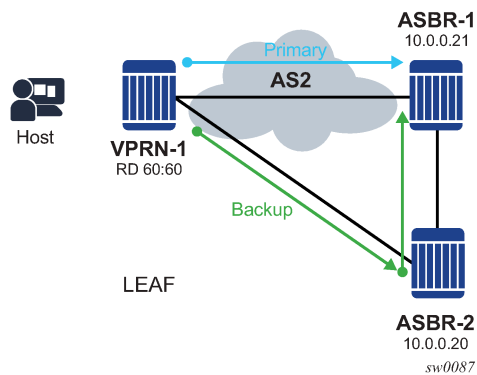
- **ASBR MoFRR**

  When the following command is enabled for LDP, and the **mcast-upstream-frr** command is not enabled, the local leaf selects a single ASBR as the primary ASBR and another ASBR as the backup ASBR.

  ```
  configure router ldp mcast-upstream-asbr-frr
  ```

  The primary and backup LSPs are set to these two ASBRs, as shown in Figure 50: ASBR MoFRR. Because the **mcast-upstream-frr** command is not configured, IGP MoFRR is not enabled in the AS2, and therefore none of the P routers 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 50: ASBR MoFRR) and its routes from the routing table so that the leaf can switch to the backup LSP and backup ASBR.

  Figure 50: ASBR MoFRR
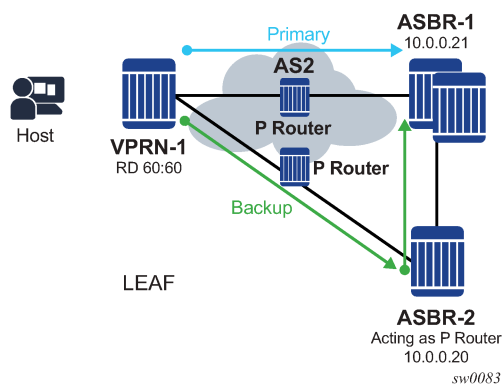
The **mcast-upstream-frr** and **mcast-upstream-asbr-frr** commands can be configured together on the local leaf of each AS to create a high-resilience MoFRR solution. When both commands are enabled, the local leaf sets up ASBR MoFRR first and sets up a primary LSP to one ASBR (ASBR-1 in Figure 51: ASBR MoFRR and IGP MoFRR) and a backup LSP to another ASBR (ASBR-2 in Figure 51: ASBR MoFRR and IGP MoFRR). In addition, the local leaf protects each LSP using IGP MoFRR through the P routers in that AS.

Figure 51: ASBR MoFRR and IGP MoFRR
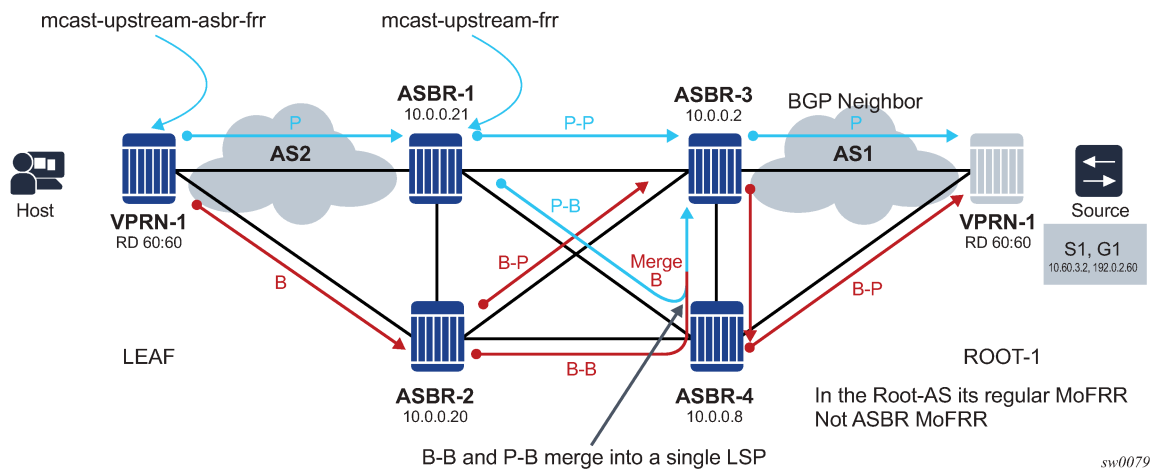
Note: Enabling both the **mcast-upstream-frr** and **mcast-upstream-asbr-frr** commands can cause extra multicast traffic to be created. Ensure that the network is designed and the appropriate commands are enabled to meet network resiliency needs.

At each AS, either command can be configured; for example, in Figure 51: ASBR MoFRR and IGP MoFRR, the leaf is configured with **mcast-upstream-asbr-frr** enabled and sets 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 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 is 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 needed 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 is selected and IGP MoFRR can provide local AS resiliency.

In the actual root AS, only the **mcast-upstream-frr** command needs to be configured.

3.31.7.2 ASBR MoFRR leaf behavior

With inter-AS MoFRR at the leaf, the leaf selects a primary ASBR and a backup ASBR. These ASBRs are disjointed ASBRs.

The primary and backup LSPs is set up using the primary and backup ASBRs, as illustrated in Figure 52: ASBR MoFRR leaf behavior.

Figure 52: ASBR MoFRR leaf behavior
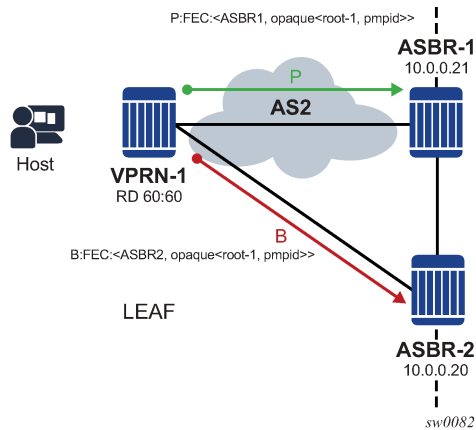
Note: Using Figure 52: ASBR MoFRR leaf behavior as a reference, ensure that the paths to ASBR-1 and ASBR-2 are disjointed from the leaf. MLDP does not support TE and cannot create two disjointed LSPs from the leaf to ASBR-1 and ASBR-2. The operator and IGP architect must define the disjointed paths.

3.31.7.3 ASBR MoFRR ASBR behavior

Each LSP at the ASBR creates its own primary and backup LSPs.

As shown in Figure 53: ASBR MoFRR ASBR behavior, the primary LSP from the leaf to ASBR-1 generates 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 merge.

Figure 53: ASBR MoFRR ASBR behavior

© 2025 Nokia.
Use subject to Terms available at: www.nokia.com/terms.
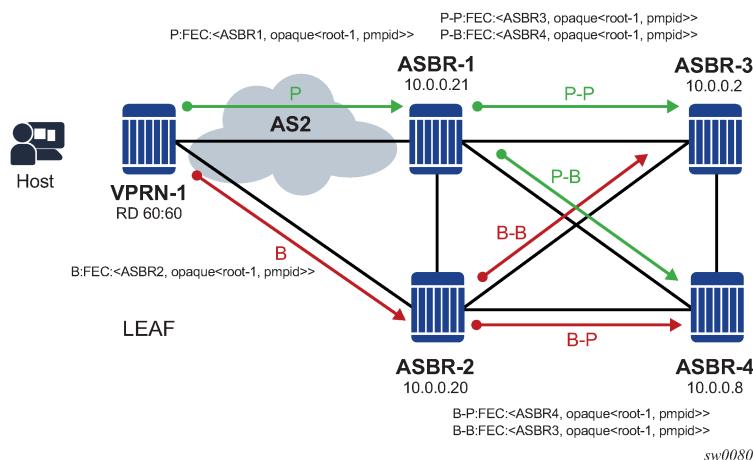
3.31.7.4 MoFRR root AS behavior

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 54: MoFRR root AS behavior.

Figure 54: MoFRR root AS behavior
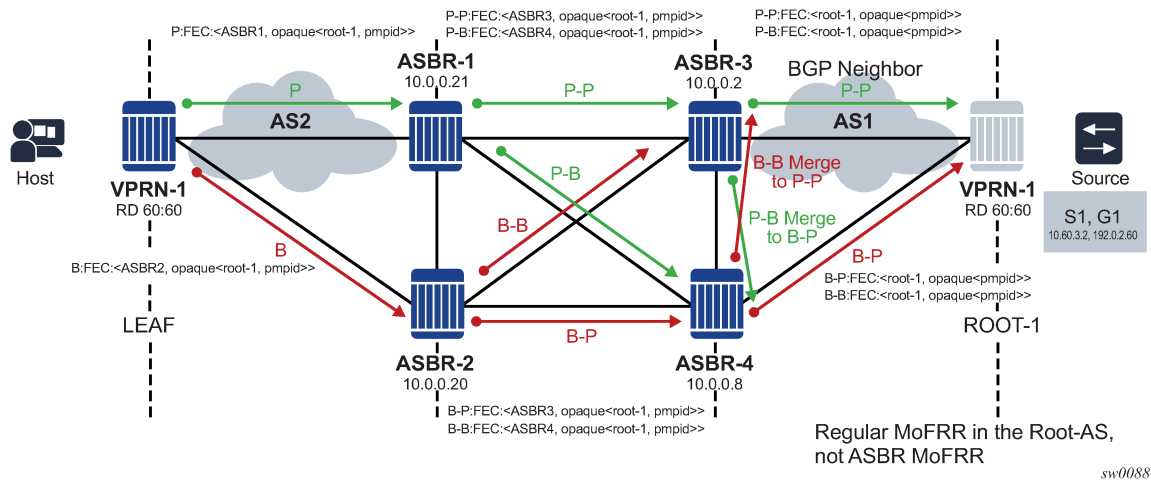
3.31.7.5 Traffic flow

Figure 55: Traffic flow illustrates traffic flow based on the LSP setup. The backup LSPs of the primary and backup LSPs (B-B, P-B) are blocked in the non-leaf AS.

Figure 55: Traffic flow
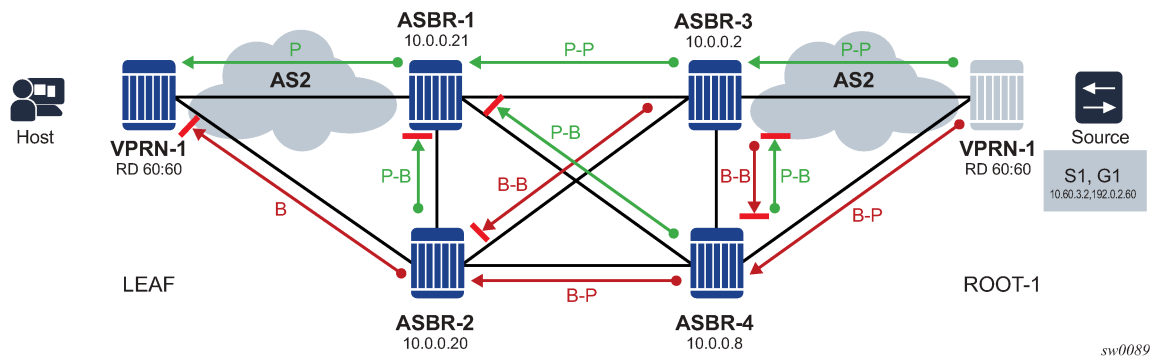
3.31.7.6 Failure detection and handling

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 multihop BFD for ASBR failure is recommended.

– Each local leaf attempts to calculate a primary ASBR or backup ASBR. The local leaf sets 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 removes the primary ASBR from the next-hop list and allows 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.

3.31.7.7 Failure scenario

As shown in Figure 56: Failure scenario 1, when ASBR-3 fails, ASBR-1 detects the failure using ASBR MoFRR and enables 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 56: Failure scenario 1
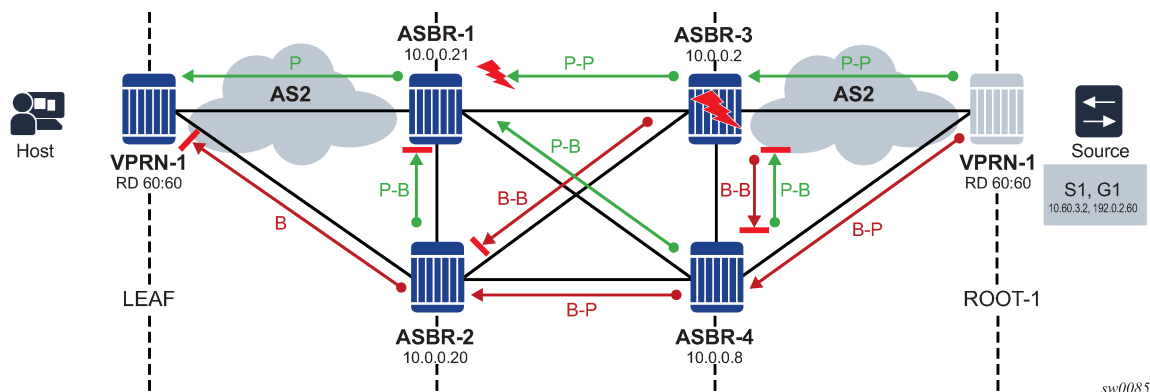
In another example, as shown in Figure 57: Failure scenario 2, failure on ASBR-1 causes the attached P router to generate a route update to the leaf, removing the ASBR-1 from the routing table and causing an ASBR-MoFRR on the leaf node.

Figure 57: Failure scenario 2
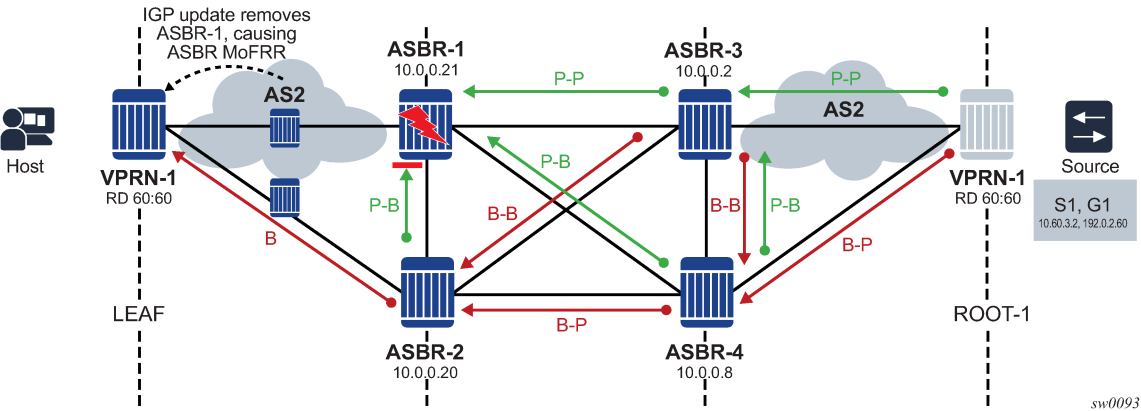
3.31.7.8 ASBR MoFRR consideration

As illustrated in Figure 58: Resolution via ASBR-3, 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 58: Resolution via ASBR-3
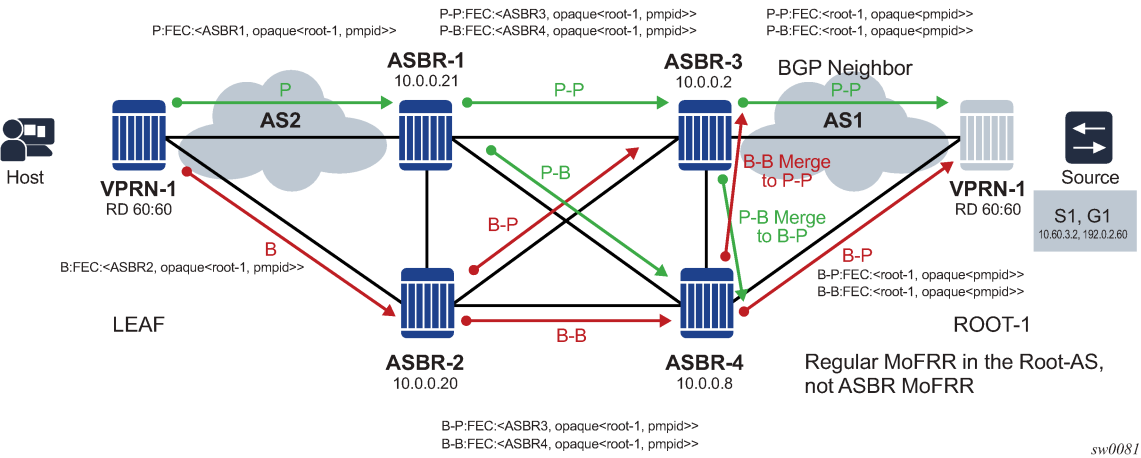
In this case, both the backup-primary LSP and primary-primary LSP are affected when a failure occurs on ASBR-3, as illustrated in Figure 59: ASBR-3 failure.

Figure 59: ASBR-3 failure
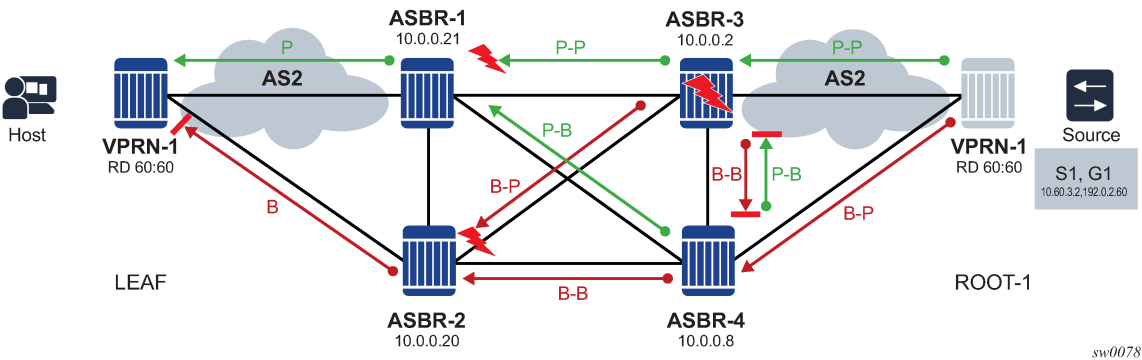
In Figure 59: ASBR-3 failure, the MoFRR can switch to the primary-backup LSP between ASBR-4 and ASBR-1 by detecting BGP MoFRR failure on ASBR-3.

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 are disjointed, but does not guarantee that the path from the local leaf to the local ASBR are 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 are not disjointed. Ensure that 2 disjointed paths are created to the primary and backup ASBRs.

3.31.7.9 ASBR MoFRR opaque support

Table 17: ASBR MoFRR opaque support lists the FEC opaque types that are supported by ASBR MoFRR.

Table 17: 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	

3.31.8 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 occurs and the primary LSP is signaled. After the primary LSP is successfully re-established, MoFRR switches from the backup LSP to the primary LSP.

MBB is performed from the leaf node to the root node, and therefore 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 does not signal any new LSP and uses this backup LSP as the primary.

3.31.9 Add-paths for route reflectors

If the ASBRs and the local leaf are connected by a route reflector, the following BGP **add-paths** command must be enabled on the route reflector.

```
configure router bgp add-paths
```

This allows for the configuration of the following commands.

```
configure router bgp add-paths mcast-vpn-ipv4
configure router bgp add-paths mcast-vpn-ipv6
configure router bgp add-paths label-ipv4 (if Option C is used)
```

The **add-paths** command forces the route reflector to advertise all ASBRs to the local leaf as the next hop for the actual root.

If the **add-paths** command is not enabled for the route reflector, only a single ASBR is advertised to the local root, and ASBR MoFRR is not available.

3.32 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 because of 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*.

3.32.1 Feature configuration

The user enables the mLDP fast upstream switchover feature by configuring the following option in CLI.

```
configure 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 then starts accepting packets.

To make use of the ECMP next hop, the user must configure the following command value in the system to two or more.

```
configure router ecmp max-ecmp-routes
```

To make use of the LFA next hop, the user must enable LFA using the following commands:

- **MD-CLI**

```
configure router isis loopfree-alternate
configure router ospf loopfree-alternate
```

- **classic CLI**

```
configure router isis loopfree-alternates
configure router ospf loopfree-alternates
```

Enabling IP FRR or LDP FRR using the following commands is not strictly required because 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, these features and the mLDP fast upstream switchover can be enabled concurrently using the following commands:

- **MD-CLI**

```
configure routing-options ip-fast-reroute
configure router ldp fast-reroute
```

- **classic CLI**

```
configure router ip-fast-reroute
configure 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. Nokia strongly recommends 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 brings down immediately the LDP session which causes 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 the following command option on all interfaces to the upstream LSR nodes.

```
configure router interface ldp-sync-timer
```

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 goes out of sync. This means the backup ILM remains activated until the next time SPF is rerun by IGP. By enabling IGP-LDP synchronization feature, the advertised link metric is changed to max value as soon as the LDP session goes down. This in turn triggers an SPF and LDP likely downloads a new set of primary and backup ILMs.

3.32.2 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. 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 because of 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 60: mLDP LSP with backup upstream LSR nodes
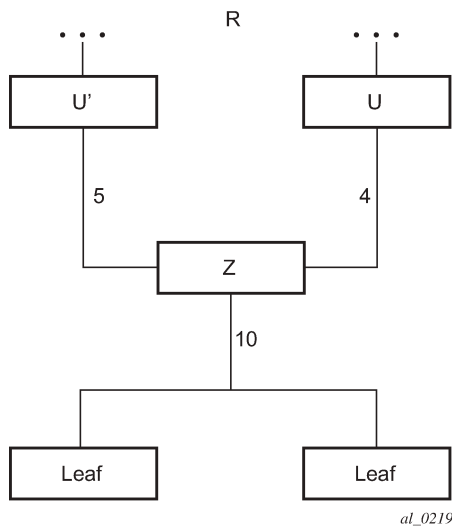
Upstream LSR U in Figure 60: mLDP LSP with backup upstream LSR nodes is the primary next hop for the root LSR *R* of the P2MP FEC. This is also referred to as primary upstream LSR. Upstream LSR *U'* is an ECMP or LFA backup next hop for the root LSR *R* of the same P2MP FEC. This is referred to as backup 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* therefore attracts traffic from both of them.

However, LSR *Z* blocks the ILM on the interface to LSR *U'* and only accepts 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* detects it and reverses the ILM blocking state and immediately starts 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* updates the primary and backup ILMs in the datapath.

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 ext hop are however mutually exclusive for a specified 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:

- The candidate upstream LSRs are numbered from lower to higher IP address.

- 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.

- 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 no ECMP or LFA next hop is found. In this case, LDP programs the primary ILM only.

3.32.3 Uniform failover from primary to backup ILM

When LDP programs the primary ILM record in the datapath, it provides the IOM with the Protect-Group Identifier (PG-ID) associated with this ILM and which identifies which upstream LSR is protected.

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. Therefore, this PG-ID is different from the one associated with unicast FECs and which is assigned to each downstream LDP interface and next hop. However, if 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 therefore the uniform FRR procedures are triggered for both the unicast LDP FECs forwarding packets toward the upstream LSR and the mLDP FECs receiving packets from the same upstream LSR.

When the mLDP FEC is programmed in the datapath, the primary and backup ILM records therefore 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 accepts 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 causes 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 brings 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.

3.33 Multi-area and multi-instance extensions to LDP

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. 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 does 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). The ABR is 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)*.

3.33.1 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 7705 SAR Gen 2. 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 are automatically 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.

See the *7705 SAR Gen 2 Unicast Routing Protocols Guide* for details on the use of LDP FEC and RSVP LSP for BGP Next-Hop Resolution.

3.33.2 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 are 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 has the shortcut next hop as the best route. If such a LDP FEC does not exist, then the routing table has the regular IP next hop and regular IP forwarding is 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.

3.33.2.1 LDP shortcut configuration

The user enables the use of LDP shortcut for resolving IGP routes by entering the following global command:

- **MD-CLI**

```
configure router ldp ldp-shortcut
```

- **classic CLI**

```
configure 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.

3.33.2.2 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 specified 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 specified 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 are 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 are 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 following command.

```
configure router ldp fec-originate
```

At the ingress LER, LDP resolves the FEC by checking in RTM that an exact match exists for this prefix. After LDP activated the FEC, it programs the NHLFE in the egress datapath and the LDP tunnel information in the ingress datapath tunnel table.

Next, LDP provides the shortcut route to RTM which associates it with the same /24 prefix. There are 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 still does 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 that succeed a longest prefix-match against the /16 route entry use the IP next hop.

3.33.2.3 LDP shortcut forwarding plane

After LDP activated a FEC for a specified 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 results 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.

3.33.3 ECMP considerations

When ECMP is enabled and multiple equal-cost next hops exist 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 is performed across the multiple next hops for the LDP FEC. The FEC next hops can either be direct link LDP neighbors or T-LDP 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 is performed across regular IP next hops for the prefix.

3.33.4 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.

```
configure router ldp shortcut-transit-ttl-propagate
configure router ldp 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 following command is configured, TTL propagation is disabled on all locally generated IP packets, including ICMP ping, traceroute, and OAM packets that are destined for a route that is resolved to the LSP shortcut:

- **MD-CLI**

```
configure router ldp shortcut-local-ttl-propagate false
```

- **classic CLI**

```
configure router ldp no shortcut-local-ttl-propagate
```

In this case, a TTL of 255 is programmed onto the pushed label stack. This is referred to as pipe mode.

Similarly, when the following command is configured, TTL propagation is disabled on all IP packets received on any IES interface and destined for a route that is resolved to the LSP shortcut:

- **MD-CLI**

```
configure router ldp shortcut-transit-ttl-propagate false
```

- **classic CLI**

```
configure router ldp no shortcut-transit-ttl-propagate
```

In this case, a TTL of 255 is programmed onto the pushed label stack.

3.34 LDP graceful handling of resource exhaustion

This feature enhances the behavior of LDP when a datapath 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 following command:

- **MD-CLI**

```
configure router ldp admin-state enable
```

- **classic CLI**

```
configure router ldp no shutdown
```

3.34.1 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 shut down. If LDP tries to resolve a FEC over a link or a targeted LDP session and it runs out of datapath or CPM resources, it brings down that interface or targeted peer which brings 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 is 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 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, is 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 are shut down. No deprogramming of datapath is required because 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 that sent a Label Mapping message associated with this ILM are shut down.

After the user has taken action to free resources up, the user must 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 datapath resources can trigger this mechanism:

- NHLFE
- ILM
- Label-to-NHLFE (LTN)
- Tunnel Index
- P2MP Index

The Label allocation CPM resource can trigger this mechanism:

3.34.2 LDP enhanced graceful handling of resources

This feature is an enhanced graceful handling capability that 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 datapath or CPM resources, it puts the LDP/T-LDP session into overload state. As a result, it releases to its LDP peer the labels of the FECs which it could not resolve and also sends 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 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 continues to forward traffic normally.

After the user has taken action to free resources up, the overload state of the LDP/T-LDP sessions toward its peers must be manually cleared.

The enhanced mechanism is 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 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 of the overload notification feature and therefore 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.

3.34.2.1 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.

3.34.2.2 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.
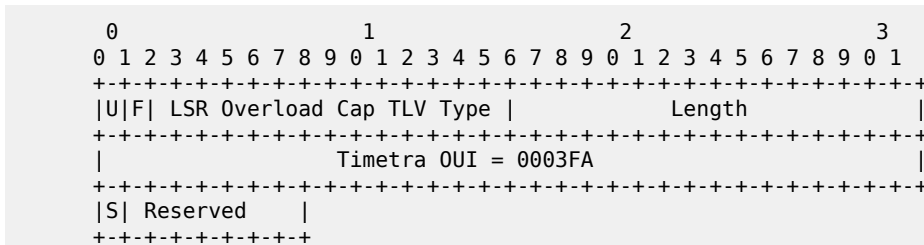
```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |U|F| LSR Overload Cap TLV Type |          Length             |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                   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).
```

3.34.2.3 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:

> **Note:** 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.

1. 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.

2. 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 specified watermark.

3. 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.

4. 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.

5. 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.

6. 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_Resources as defined in RFC 5036.

3.34.3 User guidelines and troubleshooting procedures

3.34.3.1 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 by using the following command.

```
show router ldp status
```

Output example

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

3.34.3.2 Base resource handling procedures

Procedure

Step 1. If the peer or the local LSR does not support the Overload Protection Capability, it means that the associated adjacency [interface/peer] is brought down as part of the base resource handling mechanism.

The user can determine which interface or targeted peer was administratively disabled, by applying the following commands.

```
show router ldp interface resource-failures
show router ldp targ-peer resource-failures
```

Example

```
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
```

Example

```
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 or local, or both FECs have been set in overload by the resource base resource exhaustion mechanism using the **tools dump router ldp instance** command.

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 occurs:

```
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 that are 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 are restored after performing the following commands if the LSR activates resource exhaustion again.

Use one of the following commands to clear the overload state:

-
```
clear router ldp resource-failures
```

- clears the overload state and attempt to restore adjacency and session for LDP interfaces and peers
- clears the overload state for the local FECs

-
```
clear router ldp interface
```

or

```
clear router ldp peer
```

- 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

3.34.3.3 Enhanced resource handling procedures

Procedure

Step 1. If the peer and the local LSR do support the Overload Protection Capability it means that the LSR signals the overload state for the FEC type that caused the resource exhaustion as part of the enhanced resource handling mechanism.

To verify if the local router has received or sent the overload status TLV, use the following command.

Example

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

Example

```
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 or local, or both 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 occurs:

```
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 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 are cleared after performing the following commands if the LSR activates overload again.

One of the following commands can be used depending on whether the user wants to clear all sessions in one step or one session at a time:

-
```
clear router ldp resource-failures
```

  – clears the overload state for the affected sessions and FEC types

  – clears the overload state for the local FECs

-
```
clear router ldp session ip-address overload fec-type
```

  – clears the overload state for the specified session and FEC type

  – clears the overload state for the local FECs

3.35 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:

- **MD-CLI**

```
configure router interface ldp-sync-timer seconds seconds
```

- **classic CLI**

```
configure router interface 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 causes 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 announces a new best next hop and LDP uses 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 is only advertised after the timer expires. The new cost value is also advertised after the user executes any of the following commands:

- **MD-CLI**

```
configure router isis ldp-sync false
configure router isis ldp-sync false
configure route ldp ldp-sync-timer delete seconds
tools perform router isis ldp-sync-exit
tools perform router ospf ldp-sync-exit
```

- **classic CLI**

```
configure router isis disable-ldp-sync
configure router isis disable-ldp-sync
configure router interface no ldp-sync-timer
tools perform router isis ldp-sync-exit
```

```
tools perform router ospf ldp-sync-exit
```

If the user changes the value of the LDP synchronization timer command option, the new value takes effect at the next synchronization event. If the timer is still running, it continues 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 only considers 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 ignores 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 as follows:

- **MD-CLI**

```
configure router interface ldp-sync-timer seconds seconds
configure router interface ldp-sync-timer end-of-lib
```

- **classic CLI**

```
configure router interface ldp-sync-timer seconds end-of-lib
```

If the **end-of lib** command 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** command option is configured, then the system behaves 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 specified 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 drops any unexpected End of LIB messages.

If the **end-of-lib** command option is configured, then the system also sends out an End of LIB message for prefix and P2MP FECs after all FECs are sent for all peers that have advertised the Unrecognized Notification Capability TLV.

3.36 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 resolves the root using these IGP shortcuts. Currently MLDP cannot be resolved using IGP shortcuts. To avoid this, MLDP does a lookup in the multicast RTM. IGP

shortcuts are not installed in MRTM. The following command forces MLDP do next-hop lookups in the RTM or MRTM.

```
configure router ldp resolve-root-using {ucast-rtm | mcast-rtm}
```

By default, the **resolve-root-using** command is set to the **ucast-rtm** command option 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 does 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 **resolve-root-using mcast-rtm** command option. 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 61: Recursive FEC behavior.

- When **resolve-root-using ucast-rtm** is set, MLDP uses the unicast RTM to resolve the FEC and does 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 61: 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
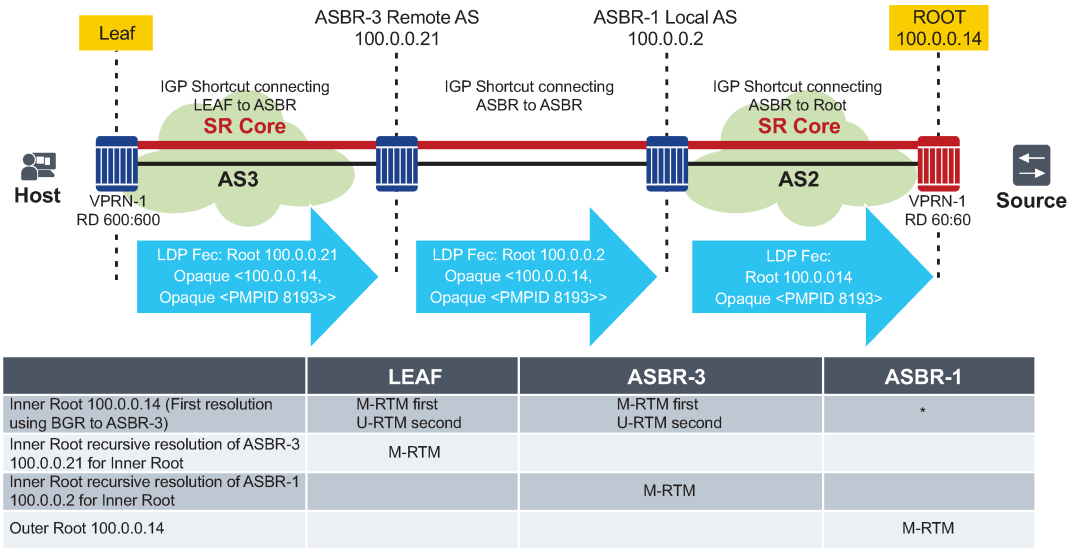
sw0442

In Figure 61: Recursive FEC behavior, the FEC between LEAF and ASBR-3 is resolved using an IGP shortcut. When the **resolve-root-using mcast-rtm** is set, the inner Root 100.0.0.14 is 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, therefore ASBR-3 100.0.0.21 is resolved only using multicast RTM when **mcast-rtm** is enabled.

3.36.1 Other considerations for multicast RTM MLDP resolution

When the **configure ldp resolve-root-using** command 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 used. 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 use MRTM and not RTM.

If **mcast-rtm** is enabled, all MLDP functionality is based on MRTM. This includes MoFRR, ASBR-MoFRR, policy-based SPMSI, and non-segmented inter-AS.

3.37 LDP process overview

Figure 62: LDP configuration and implementation displays the process to provision basic LDP.

Figure 62: LDP configuration and implementation
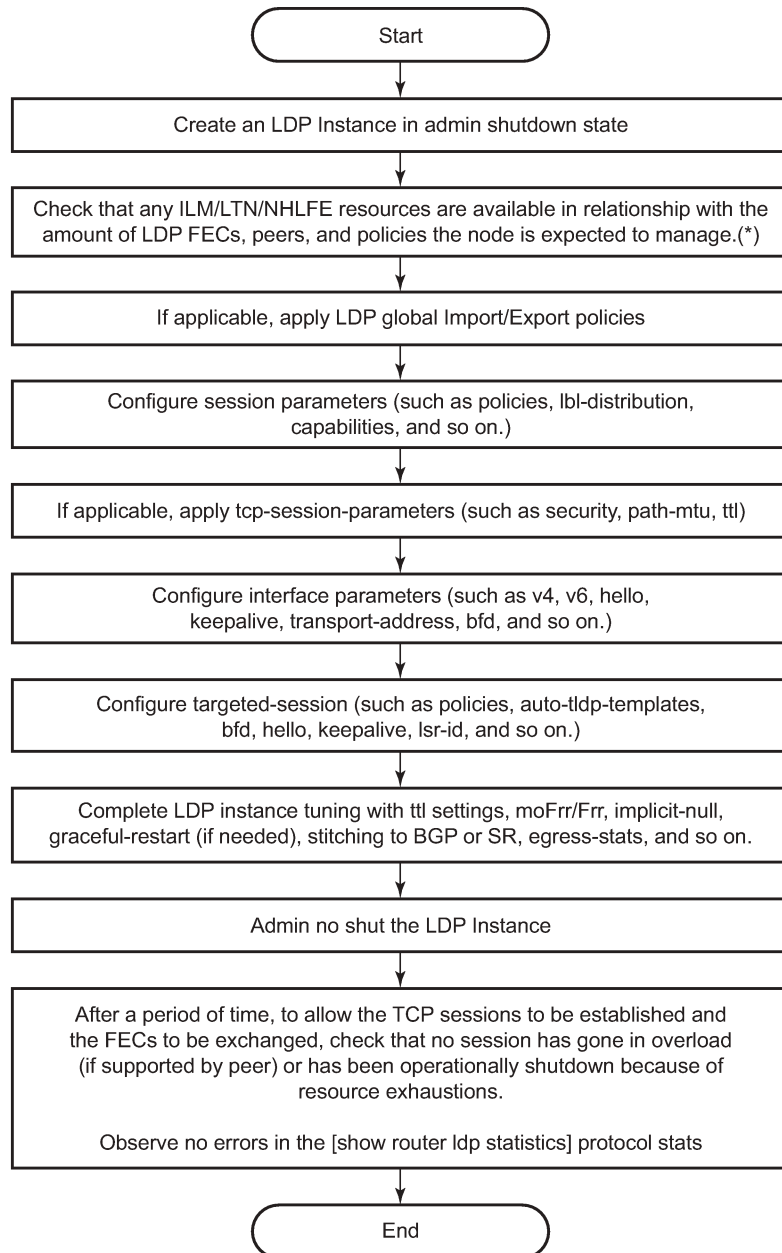
(*) if some of the needed resources are not available consider implementing stricter import-policies
    and/or enabling the per-peer fec-limit functionality.

MPLS_01

3.38 Configuring LDP with CLI

This section provides information about how to configure LDP using the command line interface.

3.38.1 LDP configuration overview

When 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.

3.38.2 Basic LDP configuration

Use this section to configure LDP and remove configuration examples of common configuration tasks.

The LDP protocol instance is created in the enabled state.

The following example shows the default LDP configuration.

Example: MD-CLI

```
[ex:/configure router "Base" ldp]
A:admin@node-2# info detail
...
    import-pmsi-routes {
        mvpn false
        mvpn-no-export-community false
    }
 ## fec-originate
    egress-statistics {
     ## fec-prefix
    }
 ## lsp-bfd
    session-parameters {
     ## peer
    }
    tcp-session-parameters {
     ## authentication-keychain
     ## authentication-key
     ## peer-transport
    }
    interface-parameters {
        ipv4 {
            transport-address system
            hello {
                timeout 15
                factor 3
            }
            keepalive {
                timeout 30
                factor 3
            }
        }
        ipv6 {
            transport-address system
            hello {
                timeout 15
                factor 3
            }
            keepalive {
```

```
                        timeout 30
                        factor 3
                    }
                }
             ## interface
            }
            targeted-session {
                sdp-auto-targeted-session true
             ## export-prefixes
             ## import-prefixes
                resolve-v6-prefix-over-shortcut false
                ipv4 {
                    hello {
                        timeout 45
                        factor 3
                    }
                    keepalive {
                        timeout 40
                        factor 4
                    }
                    hello-reduction {
                        admin-state disable
                        factor 3
                    }
                }
                ipv6 {
                    hello {
                        timeout 45
                        factor 3
                    }
                    keepalive {
                        timeout 40
                        factor 4
                    }
                    hello-reduction {
                        admin-state disable
                        factor 3
                    }
                }
          ...
```

Example: classic CLI

```
A:node-2>config>router>ldp$ info detail
----------------------------------------------
...
            import-pmsi-routes
                no mvpn
                no mvpn-no-export-community
            exit
            tcp-session-parameters
                no auth-keychain
                no authentication-key
            exit
            interface-parameters
                ipv4
                    no hello
                    no keepalive
                    no transport-address
                exit
                ipv6
                    no hello
                    no keepalive
```

```
                    no transport-address
                exit
            exit
            targeted-session
                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
                auto-tx
                    ipv4
                        shutdown
                        no tunneling
                    exit
                exit
                auto-rx
                    ipv4
                        shutdown
                        no tunneling
                    exit
                exit
                no resolve-v6-prefix-over-shortcut
            exit
            no shutdown
----------------------------------------------
```

3.38.3 Common configuration tasks

This section provides information about common LDP configuration tasks.

3.38.3.1 Enabling LDP

Note: This section applies for the classic CLI.

LDP must be enabled in order for the protocol to be active. MPLS is enabled in the **configure router mpls** context.

Use the following command to enable LDP on a router:

```
configure router ldp
```

The following example shows the enabled LDP configuration.

Example: classic CLI

```
A:node-2>config>router# info
...
#------------------------------------------------
```

```
echo "LDP Configuration"
#--------------------------------------------------
        ldp
            import-pmsi-routes
            exit
            tcp-session-parameters
            exit
            interface-parameters
            exit
            targeted-session
            exit
            no shutdown
        exit
--------------------------------------------------
```

3.38.3.2 Configuring FEC originate

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 commands in the following context to configure FEC originate.

```
configure router ldp fec-originate
```

The following example shows a FEC originate configuration.

Example: MD-CLI

```
[ex:/configure router "Base" ldp]
A:admin@node-2# info
    fec-originate 10.1.1.1/32 {
        pop true
    }
    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
    }
```

Example: classic CLI

```
A:node-2>config>router# info

//#--------------------------------------------------
echo "LDP Configuration"
#--------------------------------------------------
        ldp
            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
                import-pmsi-routes
                exit
                tcp-session-parameters
                exit
                interface-parameters
                exit
                targeted-session
                exit
                no shutdown
            exit
     ---------------------------------------------
```

3.38.3.3 Configuring graceful-restart helper

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. Graceful restart used to help third-party routers without a self-healing capability to recover.

The maximum recovery time is the time (in seconds) the sender of the TLV needs the receiver to wait, after detecting the failure of LDP communication with the sender.

The neighbor liveness time is the time (in seconds) the LSR retains its MPLS forwarding state. The time should be long enough to allow the neighboring LSRs to re-sync all the LSPs in a gracefully, without creating congestion in the LDP control plane.

Use the commands in the following context to configure graceful-restart.

```
configure router ldp graceful-restart
```

3.38.3.4 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** – matches on bindings received from the specified peer
- **prefix-list** – matches on bindings with the specified prefix or 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 commands in the following contexts to apply import and export policies.

```
configure router ldp export
configure router ldp import
```

The following example shows the export and import policy configuration.

Example: MD-CLI

```
[ex:/configure router "Base"]
A:admin@node-2# info
  ldp {
        import-policy ["LDP-import"]
        export-policy ["LDP-export"]
        fec-originate 192.168.2.1/32 {
            advertised-label 1000
            next-hop 10.10.1.2
        }
        fec-originate 192.168.1.1/32 {
            pop true
        }
    }
```

Example: classic CLI

```
A:node-2>config>router# info
#--------------------------------------------------
echo "LDP Configuration"
#--------------------------------------------------
        ldp
            export "LDP-export"
            import "LDP-import"
            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-pmsi-routes
            exit
            tcp-session-parameters
            exit
            interface-parameters
            exit
            targeted-session
            exit
            no shutdown
        exit
```

3.38.3.5 Targeted session command options

Use the commands in the following context to specify **targeted-session** command options.

```
configure router ldp targeted-session
```

The following example shows an LDP configuration.

Example: MD-CLI

```
[ex:/configure router "Base" ldp]
A:admin@node-2# info
    targeted-session {
```

```
        ipv4 {
            hello {
                timeout 120
            }
            keepalive {
                timeout 120
                factor 3
            }
        }
        peer 10.10.10.104 {
            hello {
                timeout 240
                factor 3
            }
            keepalive {
                timeout 240
                factor 3
            }
        }
    }
```

Example: classic CLI

```
A:node-2>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
----------------------------------------------
```

3.38.3.6 Configuring the LDP interface

Use the commands in the following context to configure the interface.

```
configure router ldp interface-parameters
```

The following example shows an LDP interface configuration.

Example: MD-CLI

```
[ex:/configure router "Base" ldp]
A:admin@node-2# info
 ...
    interface-parameters {
        interface "to-DUT1" {
            ipv4 {
                hello {
                    timeout 240
                    factor 3
                }
                keepalive {
```

```
                                    timeout 240
                                    factor 3
                            }
                    }
                }
            }
```

Example: classic CLI

```
A:node-2>config>router>ldp# info
----------------------------------------------
 ...
            interface-parameters
                interface "to-DUT1" dual-stack
                    ipv4
                        hello 240 3
                        keepalive 240 3
                        no shutdown
                    exit
                    no shutdown
                exit
            exit
----------------------------------------------
```

3.38.3.7 Configuring the LDP session parameters

Use the commands in the following contexts to specify session parameters.

```
configure router ldp session-parameters
configure router ldp tcp-session-parameters
```

The following example displays an LDP session parameter configuration.

Example: MD-CLI

```
[ex:/configure router "Base" ldp]
A:admin@node-2# info
    session-parameters {
        peer 10.1.1.1 {
        }
        peer 10.10.10.104 {
        }
    }
    tcp-session-parameters {
        peer-transport 10.10.10.104 {
            authentication-key "McTNkSePNJMVFysxyZa4yw8iLZbb7ys= hash2"
        }
    }
```

Example: classic CLI

```
A:node-2>config>router>ldp# info
----------------------------------------------
            import-pmsi-routes
            exit
            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 "McTNkSePNJMVFysxyZa4yw8iLZbb7ys=" hash2
                    exit
                exit
                interface-parameters
                exit
                targeted-session
                exit
                no shutdown
    -----------------------------------------------
```

3.38.3.8 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 targeted LDP (T-LDP).

```
configure service sdp signaling tldp
```

The service SDP uses the targeted-session configured in the following context.

```
configure router ldp targeted-session
```

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.

For more information about configuring SDPs, see the *7705 SAR Gen 2 Services Overview Guide*.

Use the commands in the following contexts to configure LDP on an MPLS SDP.

```
configure service sdp ldp
configure service sdp signaling
```

The following example displays an SDP configuration showing the signaling default tldp enabled.

Example: MD-CLI

```
[ex:/configure service sdp 1]
A:admin@node-2# info detail
 ...
    description "MPLS: to-99"
    path-mtu 4462
    signaling tldp
    far-end {
        ip-address 10.10.10.99
    }
 ...
```

Example: classic CLI

In the classic CLI, you must remove the LSP from the configuration using the **no lsp lsp-name** command to enable LDP on the SDP when an LSP is already specified.

```
A:node-2>config>service>sdp# info detail
----------------------------------------------
...
            description "MPLS: to-99"
            far-end 10.10.10.99
            signaling tldp
            path-mtu 4462
...
               ----------------------------------------------
```

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

Example: MD-CLI

```
[ex:/configure router "Base" ldp]
A:admin@node-2# info
    prefer-tunnel-in-tunnel false
    interface-parameters {
        interface "LDP-test" {
        }
    }
    targeted-session {
        peer 10.51.0.1 {
            admin-state disable
            tunneling {
                lsp "to_P_1" { }
            }
        }
        peer 10.51.0.17 {
            admin-state disable
            tunneling {
                lsp "to_P_6" { }
            }
        }
    }
```

Example: classic CLI

```
A:node-2>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
        ----------------------------------------------
```

Example 2 — Tunnels

Example: MD-CLI

```
[ex:/configure router "Base" interface "LDP-test" if-attribute]
A:admin@node-2# info
    admin-group ["1" "2"]

[ex:/configure router "Base" mpls]
A:admin@node-2# info
    admin-state enable
    resignal-timer 30
    path "dyn" {
        admin-state enable
    }
    lsp "to_P_1" {
        admin-state enable
        type p2p-rsvp
        to 10.51.0.1
        fast-reroute {
            frr-method facility
        }
        primary "dyn" {
        }
    }
    lsp "to_P_6" {
        admin-state enable
        type p2p-rsvp
        to 10.51.0.17
        fast-reroute {
            frr-method facility
        }
        primary "dyn" {
        }
    }
```

Example: classic CLI

```
A:node-2>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
    ---------------------------------------------
```

3.39 LDP configuration management tasks

This section provides information about LDP configuration management tasks.

3.39.1 Disabling LDP

The following command disables the LDP protocol on the router. All command options revert to the default settings.

Use the following commands to disable LDP:

- **MD-CLI**

```
configure router ldp admin-state disable
```

- **classic CLI**
  In the classic CLI, LDP must be shut down before it can be disabled.

```
configure router ldp shutdown
configure router no ldp
```

3.39.2 Modifying targeted session command options

The modification of LDP targeted session command options does not take effect until the next time the session goes down and is re-established. Individual command options cannot be deleted. Different defaults can be configured for IPv4 and IPv6 LDP targeted Hello adjacencies.

The following example displays the default values.

Example: MD-CLI

```
[ex:/configure router "Base" ldp targeted-session]
A:admin@node-2# info detail
    sdp-auto-targeted-session true
 ## export-prefixes
 ## import-prefixes
    resolve-v6-prefix-over-shortcut false
    ipv4 {
        hello {
            timeout 45
            factor 3
        }
        keepalive {
            timeout 40
            factor 4
        }
        hello-reduction {
            admin-state disable
            factor 3
        }
    }
    ipv6 {
        hello {
            timeout 45
            factor 3
        }
        keepalive {
            timeout 40
            factor 4
        }
        hello-reduction {
            admin-state disable
            factor 3
        }
    }
 ## peer
 ...
```

Example: classic CLI

```
A:node-2>config>router>ldp>targ-session# 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
                ...
----------------------------------------------
```

3.39.3 Modifying interface parameters

Individual parameters cannot be deleted. 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 example displays the default values.

Example: MD-CLI

```
!*[pr:/configure router "Base" ldp interface-parameters]
A:admin@node-2# info detail
    ipv4 {
        transport-address system
        hello {
            timeout 15
            factor 3
        }
        keepalive {
            timeout 30
            factor 3
        }
    }
    ipv6 {
        transport-address system
        hello {
            timeout 15
            factor 3
        }
        keepalive {
            timeout 30
            factor 3
        }
    }
    interface "LDP-test" {
     ## apply-groups
     ## apply-groups-exclude
        admin-state enable
     ## load-balancing-weight
        bfd-liveness {
            ipv4 false
            ipv6 false
        }
     ## ipv4
     ## ipv6
    }
```

Example: classic CLI

In the classic CLI, the **no** form of an **interface-parameters interface** command reverts modified values back to the defaults.

```
A:node-2>config>router>ldp>if-params>if$ info detail
---------------------------------------------
                    no bfd-enable
                    no load-balancing-weight
                    ipv4
                        no hello
                        no keepalive
                        no local-lsr-id
                        fec-type-capability
                            prefix-ipv4 enable
```

```
                        prefix-ipv6 enable
                        p2mp-ipv4 enable
                        p2mp-ipv6 enable
                    exit
                    no transport-address
                    no shutdown
            exit
            no shutdown
----------------------------------------------
```

4 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.

4.1 Bidirectional Forwarding Detection (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)*

4.2 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-gr-notification-01, *Notification Message support for BGP Graceful Restart*

draft-ietf-idr-bgp-ls-app-specific-attr-16, *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-link-bandwidth-03, *BGP Link Bandwidth 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 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 5668, *4-Octet AS Specific BGP Extended Community*

RFC 6286, *Autonomous-System-Wide Unique BGP Identifier for BGP-4*

RFC 6368, *Internal BGP as the Provider/Customer Edge Protocol for BGP/MPLS IP Virtual Private Networks (VPNs)*

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 7606, *Revised Error Handling for BGP UPDATE Messages*

RFC 7607, *Codification of AS 0 Processing*

RFC 7752, *North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP*

RFC 7911, *Advertisement of Multiple Paths in BGP*

RFC 7999, *BLACKHOLE Community*

RFC 8092, *BGP Large Communities Attribute*

RFC 8097, *BGP Prefix Origin Validation State Extended Community*

RFC 8212, *Default External BGP (EBGP) Route Propagation Behavior without Policies*

RFC 8277, *Using BGP to Bind MPLS Labels to Address Prefixes*

RFC 9494, *Long-Lived Graceful Restart for BGP*

4.3 Bridging and management

IEEE 802.1AB, *Station and Media Access Control Connectivity Discovery*

IEEE 802.1ad, *Provider Bridges*

IEEE 802.1AX, *Link Aggregation*

IEEE 802.1Q, *Virtual LANs*

4.4 Certificate management

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 5280, *Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile*

RFC 6712, *Internet X.509 Public Key Infrastructure -- HTTP Transfer for the Certificate Management Protocol (CMP)*

RFC 7030, *Enrollment over Secure Transport*

RFC 7468, *Textual Encodings of PKIX, PKCS, and CMS Structures*

4.5 Ethernet VPN (EVPN)

draft-ietf-bess-bgp-srv6-args-00, *SRv6 Argument Signaling for BGP Services*

draft-ietf-bess-evpn-ipvpn-interworking-06, *EVPN Interworking with IPVPN*

draft-sr-bess-evpn-vpws-gateway-03, *Ethernet VPN Virtual Private Wire Services Gateway Solution*

RFC 7432, *BGP MPLS-Based Ethernet VPN*

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 9047, *Propagation of ARP/ND Flags in an Ethernet Virtual Private Network (EVPN)*

RFC 9135, *Integrated Routing and Bridging in Ethernet VPN (EVPN)*

RFC 9136, *IP Prefix Advertisement in Ethernet VPN (EVPN)*

RFC 9161, *Operational Aspects of Proxy ARP/ND in Ethernet Virtual Private Networks*

RFC 9251, *Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Proxies for Ethernet VPN (EVPN)*

4.6 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 0.8.0, *gRPC Network Management Interface (gNMI) Service Specification*

gnmi_ext.proto version 0.1.0, *gNMI Commit Confirmed Extension*

system.proto version 1.0.0, *gRPC Network Operations Interface (gNOI) System Service*

PROTOCOL-HTTP2, *gRPC over HTTP2*

4.7 Intermediate System to Intermediate System (IS-IS)

draft-ietf-isis-mi-02, *IS-IS Multi-Instance*

draft-ietf-lsr-igp-ureach-prefix-announce-01, *IGP Unreachable Prefix Announcement* – without U-Flag and UP-Flag

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 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 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 6213, *IS-IS BFD-Enabled TLV*

RFC 6232, *Purge Originator Identification TLV for IS-IS*

RFC 6233, *IS-IS Registry Extension for Purges*

RFC 7775, *IS-IS Route Preference for Extended IP and IPv6 Reachability*

RFC 7794, *IS-IS Prefix Attributes for Extended IPv4 and IPv6 Reachability* – sections 2.1 and 2.3

RFC 7981, *IS-IS Extensions for Advertising Router Information*

RFC 7987, *IS-IS Minimum Remaining Lifetime*

RFC 8202, *IS-IS Multi-Instance* – single topology

RFC 8570, *IS-IS Traffic Engineering (TE) Metric Extensions* – Min/Max Unidirectional Link Delay metric for flex-algo, RSVP, SR-TE

RFC 8919, *IS-IS Application-Specific Link Attributes*

4.8 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 2784, *Generic Routing Encapsulation (GRE)*

RFC 3164, *The BSD syslog Protocol*

RFC 4632, *Classless Inter-domain Routing (CIDR): The Internet Address Assignment and Aggregation Plan*

RFC 5082, *The Generalized TTL Security Mechanism (GTSM)*

4.9 Internet Protocol (IP) multicast

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 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 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 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 7761, *Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)*

RFC 8487, *Mtrace Version 2: Traceroute Facility for IP Multicast*

4.10 Internet Protocol (IP) version 4

RFC 791, *Internet Protocol*

RFC 792, *Internet Control Message Protocol*

RFC 826, *An Ethernet Address Resolution Protocol*

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 1812, *Requirements for IPv4 Routers*

RFC 1918, *Address Allocation for Private Internets*

RFC 2131, *Dynamic Host Configuration Protocol*; Relay only

RFC 2132, *DHCP Options and BOOTP Vendor Extensions* – DHCP

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

4.11 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 3736, *Stateless Dynamic Host Configuration Protocol (DHCP) Service for IPv6*

RFC 3971, *SEcure Neighbor Discovery (SEND)*

RFC 4007, *IPv6 Scoped Address Architecture*

RFC 4191, *Default Router Preferences and More-Specific Routes* – Default Router Preference

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 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 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*

4.12 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 2560, *X.509 Internet Public Key Infrastructure Online Certificate Status Protocol - OCSP*

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 4109, *Algorithms for Internet Key Exchange version 1 (IKEv1)*

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 4754, *IKE and IKEv2 Authentication Using the Elliptic Curve Digital Signature Algorithm (ECDSA)*

RFC 4835, *Cryptographic Algorithm Implementation Requirements for Encapsulating Security Payload (ESP) and Authentication Header (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 5282, *Using Authenticated Encryption Algorithms with the Encrypted Payload of the IKEv2 Protocol*

RFC 5903, *ECP Groups for IKE and IKEv2*

RFC 5996, *Internet Key Exchange Protocol Version 2 (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 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 8784, *Mixing Preshared Keys in the Internet Key Exchange Protocol Version 2 (IKEv2) for Post-quantum Security*

4.13 Label Distribution Protocol (LDP)

draft-pdutta-mpls-ldp-adj-capability-00, *LDP Adjacency Capabilities*

draft-pdutta-mpls-ldp-v2-00, *LDP Version 2*

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*

4.14 Multiprotocol Label Switching (MPLS)

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 5332, *MPLS Multicast Encapsulations*

RFC 5884, *Bidirectional Forwarding Detection (BFD) for MPLS Label Switched Paths (LSPs)*

RFC 6424, *Mechanism for Performing Label Switched Path Ping (LSP Ping) over MPLS Tunnels*

RFC 7308, *Extended Administrative Groups in MPLS Traffic Engineering (MPLS-TE)*

RFC 7746, *Label Switched Path (LSP) Self-Ping*

4.15 Network Address Translation (NAT)

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*

4.16 Open Shortest Path First (OSPF)

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 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 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 5642, *Dynamic Hostname Exchange Mechanism for OSPF*

RFC 6549, *OSPFv2 Multi-Instance Extensions*

RFC 6987, *OSPF Stub Router Advertisement*

RFC 7471, *OSPF Traffic Engineering (TE) Metric Extensions* – Min/Max Unidirectional Link Delay metric for flex-algo, RSVP, SR-TE

RFC 7684, *OSPFv2 Prefix/Link Attribute Advertisement*

RFC 7770, *Extensions to OSPF for Advertising Optional Router Capabilities*

RFC 8920, *OSPF Application-Specific Link Attributes*

4.17 Path Computation Element Protocol (PCEP)

draft-alvarez-pce-path-profiles-04, *PCE Path Profiles*

draft-ietf-pce-binding-label-sid-15, *Carrying Binding Label/Segment Identifier (SID) in PCE-based Networks.* – MPLS binding SIDs

RFC 5440, *Path Computation Element (PCE) Communication Protocol (PCEP)*

RFC 8231, *Path Computation Element Communication Protocol (PCEP) Extensions for Stateful PCE*

RFC 8253, *PCEPS: Usage of TLS to Provide a Secure Transport for the Path Computation Element Communication Protocol (PCEP)*

RFC 8281, *PCEP Extensions for PCE-initiated LSP Setup in a Stateful PCE Model*

RFC 8408, *Conveying Path Setup Type in PCE Communication Protocol (PCEP) Messages*

RFC 8664, *Path Computation Element Communication Protocol (PCEP) Extensions for Segment Routing*

4.18 Pseudowire (PW)

draft-ietf-l2vpn-vpws-iw-oam-04, *OAM Procedures for VPWS 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 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

RFC 8395, *Extensions to BGP-Signaled Pseudowires to Support Flow-Aware Transport Labels*

4.19 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)*

4.20 Remote Authentication Dial In User Service (RADIUS)

draft-oscca-cfrg-sm3-02, *The SM3 Cryptographic Hash Function*

RFC 2865, *Remote Authentication Dial In User Service (RADIUS)*

RFC 2866, *RADIUS Accounting*

RFC 3162, *RADIUS and IPv6*

RFC 6613, *RADIUS over TCP* – with TLS

RFC 6614, *Transport Layer Security (TLS) Encryption for RADIUS*

RFC 6929, *Remote Authentication Dial-In User Service (RADIUS) Protocol Extensions*

4.21 Resource Reservation Protocol - Traffic Engineering (RSVP-TE)

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 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 5817, *Graceful Shutdown in MPLS and Generalized MPLS Traffic Engineering Networks*

4.22 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*

4.23 Segment Routing (SR)

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 8426, *Recommendations for RSVP-TE and Segment Routing (SR) Label Switched Path (LSP) Coexistence*

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 8665, *OSPF Extensions for Segment Routing*

RFC 8667, *IS-IS Extensions for Segment Routing*

RFC 8669, *Segment Routing Prefix Segment Identifier Extensions for BGP*

RFC 9256, *Segment Routing Policy Architecture*

RFC 9350, *IGP Flexible Algorithm*

4.24 Simple Network Management Protocol (SNMP)

draft-blumenthal-aes-usm-04, *The AES Cipher Algorithm in the SNMP's User-based Security Model –* CFB128-AES-192 and CFB128-AES-256

draft-ietf-isis-wg-mib-06, *Management Information Base for Intermediate System to Intermediate System (IS-IS)*

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

ESO-CONSORTIUM-MIB revision 200406230000Z, *esoConsortiumMIB*

IANA-ADDRESS-FAMILY-NUMBERS-MIB revision 200203140000Z, *ianaAddressFamilyNumbers*

IANAifType-MIB revision 200505270000Z, *ianaifType*

IANA-RTPROTO-MIB revision 200009260000Z, *ianaRtProtoMIB*

IEEE8021-CFM-MIB revision 200706100000Z, *ieee8021CfmMib*

IEEE8021-PAE-MIB revision 200101160000Z, *ieee8021paeMIB*

IEEE8023-LAG-MIB revision 200006270000Z, *lagMIB*

RFC 1157, *A Simple Network Management Protocol (SNMP)*

RFC 1212, *Concise MIB Definitions*

RFC 1215, *A Convention for Defining Traps for use with the SNMP*

RFC 1724, *RIP Version 2 MIB Extension*

RFC 1901, *Introduction to Community-based SNMPv2*

RFC 2206, *RSVP Management Information Base using SMIv2*

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 3273, *Remote Network Monitoring Management Information Base for High Capacity Networks*

RFC 3410, *Introduction and Applicability Statements for Internet Standard Management Framework*

RFC 3430, *Simple Network Management Protocol (SNMP) over Transmission Control Protocol (TCP) Transport Mapping*

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 3418, *Management Information Base (MIB) for the Simple Network Management Protocol (SNMP)*

RFC 3419, *Textual Conventions for Transport Addresses*

RFC 3584, *Coexistence between Version 1, Version 2, and Version 3 of the Internet-standard Network Management Framework*

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 3826, *The Advanced Encryption Standard (AES) Cipher Algorithm in the SNMP User-based Security Model*

RFC 3877, *Alarm Management Information Base (MIB)*

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 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 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*

RFC 7630, *HMAC-SHA-2 Authentication Protocols in the User-based Security Model (USM) for SNMPv3*

4.25 Timing

RFC 3339, *Date and Time on the Internet: Timestamps*

RFC 5905, *Network Time Protocol Version 4: Protocol and Algorithms Specification*

RFC 8573, *Message Authentication Code for the Network Time Protocol*

4.26 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

RFC 8972, *Simple Two-Way Active Measurement Protocol Optional Extensions* – unauthenticated

RFC 9503, *Simple Two-Way Active Measurement Protocol (STAMP) Extensions for Segment Routing Networks* – excluding Sections 3, 4.1.2 and 4.1.3

4.27 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 7117, *Multicast in Virtual Private LAN Service (VPLS)*

4.28 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*

Customer document and product support

Customer documentation
Customer documentation welcome page

Technical support
Product support portal

Documentation feedback
Customer documentation feedback