Segment Routing Configuration Guide for Cisco 8000 Series Routers, IOS XR Release 7.3.x

First Published: 2021-02-26

Last Modified: 2024-01-31

Americas Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
http://www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 527-0883

CONTENTS

Preface

Note This product has reached end-of-life status. For more information, see the End-of-Life and End-of-Sale Notices.

The *Segment Routing Configuration Guide for Cisco 8000 Series Router* preface contains these sections:

Changes to This Document

This table lists the changes made to this document since it was first printed.

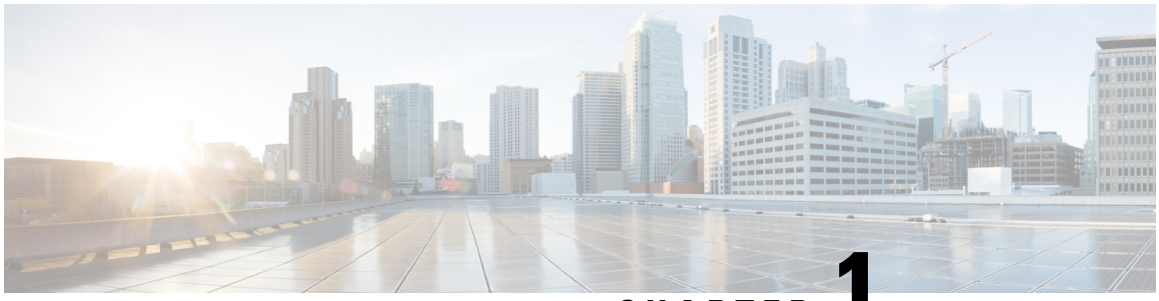
Date	Change Summary
January 2024	Republished for Release 7.3.6
August 2023	Republished for Release 7.3.5
October 2021	Republished for Release 7.3.2
February 2021	Initial release of this document

Communications, Services, and Additional Information

- To receive timely, relevant information from Cisco, sign up at Cisco Profile Manager.
- To get the business impact you're looking for with the technologies that matter, visit Cisco Services.
- To submit a service request, visit Cisco Support.
- To discover and browse secure, validated enterprise-class apps, products, solutions and services, visit Cisco Marketplace.
- To obtain general networking, training, and certification titles, visit Cisco Press.
- To find warranty information for a specific product or product family, access Cisco Warranty Finder.

Cisco Bug Search Tool

Cisco Bug Search Tool (BST) is a web-based tool that acts as a gateway to the Cisco bug tracking system that maintains a comprehensive list of defects and vulnerabilities in Cisco products and software. BST provides you with detailed defect information about your products and software.



CHAPTER 1

New and Changed Information for Segment Routing Features

This table summarizes the new and changed feature information for the *Segment Routing Configuration Guide for Cisco 8000 Series Routers*, and lists where they are documented.

New and Changed Segment Routing Features

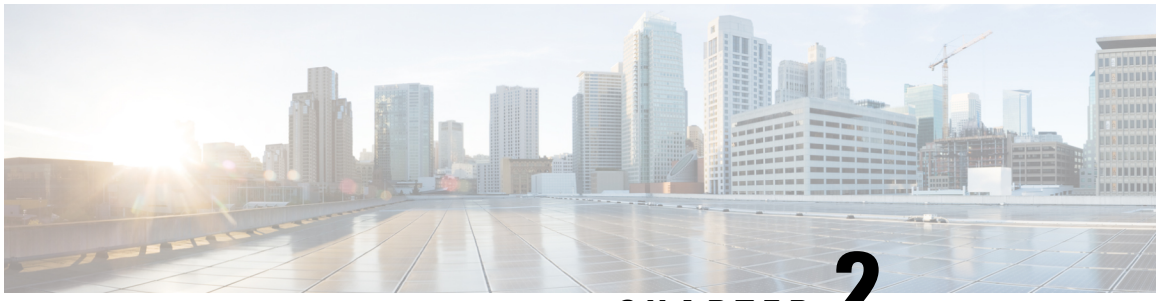
Segment Routing Features Added or Modified in IOS XR Release 7.3.x

Feature	Description	Introduced/Changed in Release	Where Documented
Data Plane Validation for SR-MPLS IPv6-based Controller Instantiated LSPs	You can now verify the network configuration and paths and policies set up, without interrupting or potentially disrupting live network traffic, for SR-MPLS (Segment Routing over Multiprotocol Label Switching) IPv6-based Label Switched-Paths (LSPs). With this feature, you can validate controller instantiated LSPs programmed directly into the forwarding hardware.	Release 7.3.6	Data Plane Validation Support for SR-MPLS IPv6-based LSPs, on page 168
Disable Penultimate Hop Popping	You can now disable the penultimate hop popping (PHP) without adding an explicit-Null label.	Release 7.3.5	Configuring a Prefix-SID on the IS-IS Enabled Loopback Interface, on page 18

Feature	Description	Introduced/Changed in Release	Where Documented
Overriding MPLS Imposition (IP2MPLS) via Service Layer API (SL-API)	In scenarios where sr-prefer is enabled, this feature allows you to specify SR prefixes through an Access Control List where their imposition forwarding entry (IP-to-MPLS) gives preference to SL-API, instead of the SR native LSP.	Release 7.3.5	Overriding MPLS Imposition (IP-to-MPLS) via Service Layer API (SL-API), on page 21
BGP Best Path Computation using SR Policy Paths	You can now disable the penultimate hop popping (PHP) without adding an explicit-Null label.	Release 7.3.4	BGP Best Path Computation using SR Policy Paths, on page 93
BGP Proxy Prefix SID	This feature is a BGP extension to signal BGP prefix-SIDs. This feature allows you to attach BGP prefix SID attributes for remote prefixes learned over BGP labeled unicast (LU) sessions and propagate them as SR prefixes using BGP LU. This allows an LSP towards non-SR endpoints to use segment routing global block in the SR domain.	Release 7.3.2	Configure BGP Proxy Prefix SID, on page 90
BFDv6-triggered TI-LFA	BFDv6-triggered TI-LFA allows you to obtain link, node, and SRLG protection using the Bidirectional Forwarding Detection (BFD) over IPv6 protocol.	Release 7.3.2	Behaviors and Limitations of TI-LFA, on page 125

Feature	Description	Introduced/Changed in Release	Where Documented
Segment Routing Flexible Algorithm	This feature allows for user-defined algorithms where the IGP computes paths based on a combination of metric type and constraint. An operator can assign custom SR prefix-SIDs to realize forwarding beyond link-cost-based SPF. As a result, this feature provides a traffic-engineered path computed automatically by the IGP to any destination reachable by the IGP.	Release 7.3.1	Enabling Segment Routing Flexible Algorithm, on page 99
Link Delay Measurement using TWAMP Light Encoding	The PM for link delay uses the IP/UDP packet format defined in RFC 5357 (TWAMP-Light) for probes. Two-Way Active Measurement Protocol (TWAMP) adds two-way or round-trip measurement capabilities. TWAMP employs time stamps applied at the echo destination (reflector) to enable greater accuracy.	Release 7.3.1	Configure Performance Measurement, on page 107
SR OAM for SR Policy (Policy Name / Binding SID / Custom label stack)	This feature extends SR OAM ping and traceroute function for an SR policy (or binding SID)-LSP end-point combination. This addresses the limitations of the Nil-FEC LSP Ping and Traceroute function which cannot perform a ping operation to a segment list that is not associated with an installed SR policy. Also, it cannot validate egress device-specific SR policies.	Release 7.3.1	Segment Routing Ping and Traceroute, on page 156

Feature	Description	Introduced/Changed in Release	Where Documented
Segment Routing Data Plane Monitoring	Unreported traffic drops in MPLS networks could be difficult to detect and isolate. They can be caused by user configuration, out-of-sync neighbors, or incorrect data-plane programming. Segment Routing Data Plane Monitoring (SR DPM) provides a scalable solution to address data-plane consistency verification and unreported traffic drops. SR DPM validates the actual data plane status of all FIB entries associated with SR IGP prefix SIDs.	Release 7.3.1	Segment Routing Data Plane Monitoring , on page 163



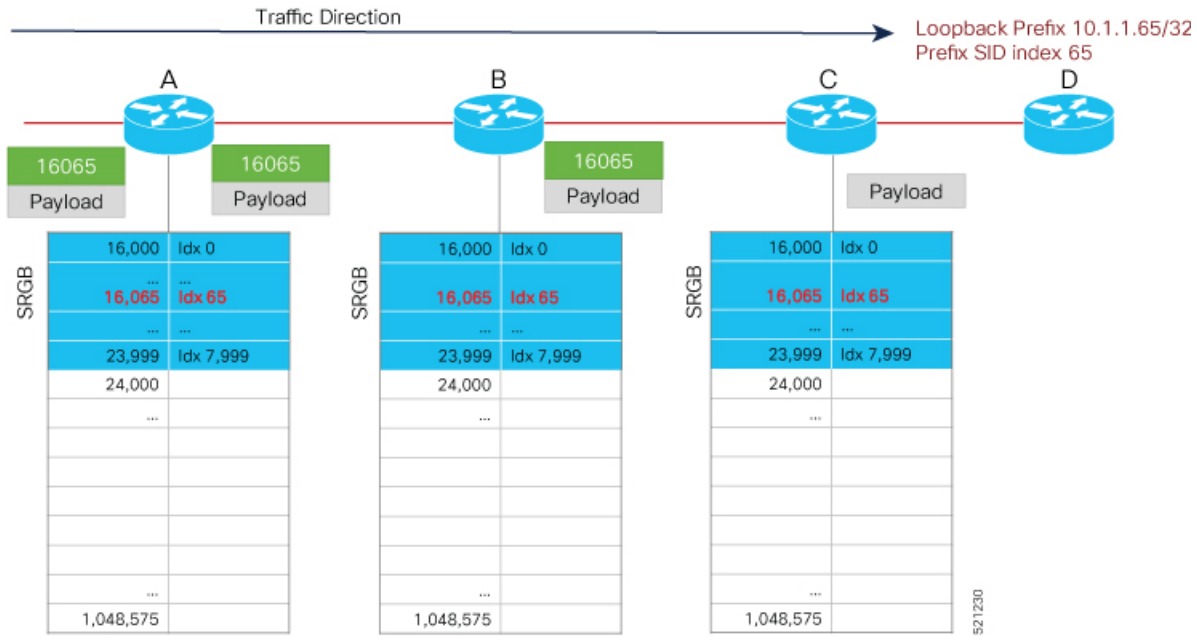
CHAPTER 2

Configure Segment Routing Global Block and Segment Routing Local Block

Local label allocation is managed by the label switching database (LSD). The Segment Routing Global Block (SRGB) and Segment Routing Local Block (SRLB) are label values preserved for segment routing in the LSD.

About the Segment Routing Global Block

The Segment Routing Global Block (SRGB) is a range of labels reserved for Segment Routing global segments. A prefix-SID is advertised as a domain-wide unique index. The prefix-SID index points to a unique label within the SRGB range. The index is zero-based, meaning that the first index is 0. The MPLS label assigned to a prefix is derived from the Prefix-SID index plus the SRGB base. For example, considering an SRGB range of 16,000 to 23,999, a prefix 1.1.1.65/32 with prefix-SID index of **65** is assigned the label value of **16065**.



To keep the configuration simple and straightforward, we strongly recommended that you use a homogenous SRGB (meaning, the same SRGB range across all nodes). Using a heterogenous SRGB (meaning, a different SRGB range of the same size across nodes) is also supported but is not recommended.

Behaviors and Limitations

- The default SRGB in IOS XR has a size of 8000 starting from label value 16000. The default range is 16000 to 23,999. With this size, and assuming one loopback prefix per router, an operator can assign prefix SIDs to a network with 8000 routers.
- There are instances when you might need to define a different SRGB range. For example:
  - Non-IOS XR nodes with a SRGB range that is different than the default IOS XR SRGB range.
  - The default SRGB range is not large enough to accommodate all required prefix SIDs.
- A non-default SRGB can be configured following these guidelines:
  - The SRGB starting value can be configured anywhere in the dynamic label range space (16,000 to 1,048,575).
  - The SRGB can be configured to any size value that fits within the dynamic label range space.
- Allocating an SRGB label range does not mean that all the labels in this range are programmed in the forwarding table. The label range is just reserved for SR and not available for other purposes. Furthermore, a platform may limit the number of local labels that can be programmed.
- We recommend that the non-default SRGB be configured under the **segment-routing** global configuration mode. By default, all IGP instances and BGP use this SRGB.
- You can also configure a non-default SRGB under the IGP, but it is not recommended.

SRGB Label Conflicts

When you define a non-default SRGB range, there might be a label conflict (for example, if labels are already allocated, statically or dynamically, in the new SRGB range). The following system log message indicates a label conflict:

```
%ROUTING-ISIS-4-SRGB_ALLOC_FAIL : SRGB allocation failed: 'SRGB reservation not
successful for [16000,80000], SRGB (16000 80000, SRGB_ALLOC_CONFIG_PENDING, 0x2)
(So far 16 attempts). Make sure label range is free'
```

To remove this conflict, you must reload the router to release the currently allocated labels and to allocate the new SRGB.

After the system reloads, LSD does not accept any dynamic label allocation before IS-IS/OSPF/BGP have registered with LSD. Upon IS-IS/OSPF/BGP registration, LSD allocates the requested SRGB (either the default range or the customized range).

After IS-IS/OSPF/BGP have registered and their SRGB is allocated, LSD starts serving dynamic label requests from other clients.

Note To avoid a potential router reload due to label conflicts, and assuming that the default SRGB size is large enough, we recommend that you use the default IOS XR SRGB range.

Note Allocating a non-default SRGB in the upper part of the MPLS label space increases the chance that the labels are available and a reload can be avoided.

Caution Modifying a SRGB configuration is disruptive for traffic and may require a reboot if the new SRGB is not available entirely.

About the Segment Routing Local Block

A local segment is automatically assigned an MPLS label from the dynamic label range. In most cases, such as TI-LFA backup paths and SR-TE explicit paths defined with IP addresses, this dynamic label allocation is sufficient. However, in some scenarios, it could be beneficial to allocate manually local segment label values to maintain label persistency. For example, an SR-TE policy with a manual binding SID that is performing traffic steering based on incoming label traffic with the binding SID.

The Segment Routing Local Block (SRLB) is a range of label values preserved for the manual allocation of local segments, such as adjacency segment identifiers (adj-SIDs), Layer 2 adj-SIDs, and binding SIDs (BSIDs). These labels are locally significant and are only valid on the nodes that allocate the labels.

Behaviors and Limitations

- The default SRLB has a size of 1000 starting from label value 15000; therefore, the default SRLB range goes from 15000 to 15,999.

- A non-default SRLB can be configured following these guidelines:
  - The SRLB starting value can be configured anywhere in the dynamic label range space (16,000 to 1,048,575).
  - The SRLB can be configured to any size value that fits within the dynamic label range space.

SRLB Label Conflicts

When you define a non-default SRLB range, there might be a label conflict (for example, if labels are already allocated, statically or dynamically, in the new SRLB range). In this case, the new SRLB range will be accepted, but not applied (pending state). The previous SRLB range (active) will continue to be in use.

To remove this conflict, you must reload the router to release the currently allocated labels and to allocate the new SRLB.

Caution You can use the **clear segment-routing local-block discrepancy all** command to clear label conflicts. However, using this command is disruptive for traffic since it forces all other MPLS applications with conflicting labels to allocate new labels.

Note To avoid a potential router reload due to label conflicts, and assuming that the default SRGB size is large enough, we recommend that you use the default IOS XR SRLB range.

Note Allocating a non-default SRLB in the upper part of the MPLS label space increases the chance that the labels are available and a reload can be avoided.

Understanding Segment Routing Label Allocation

In IOS XR, local label allocation is managed by the Label Switching Database (LSD). MPLS applications must register as a client with the LSD to allocate labels. Most MPLS applications (for example: LDP, RSVP, L2VPN, BGP [LU, VPN], IS-IS and OSPF [Adj-SID], SR-TE [Binding-SID]) use labels allocated dynamically by LSD.

With Segment Routing-capable IOS XR software releases, the LSD *preserves* the default SRLB label range (15,000 to 15,999) and default SRGB label range (16,000 to 23,999), even if Segment Routing is not enabled.

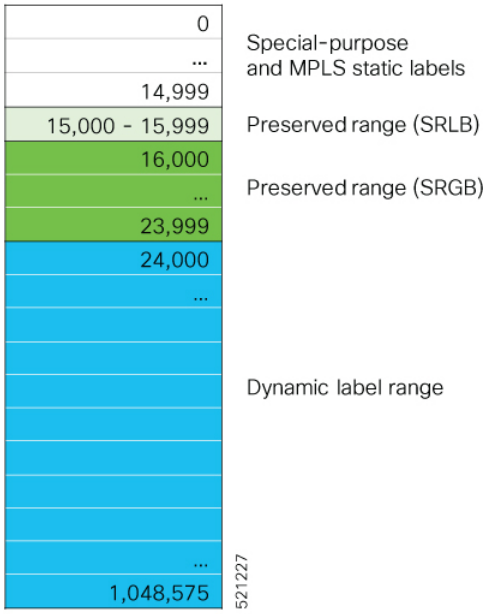
This preservation of the default SRLB/SRGB label range makes future Segment Routing activation possible without a reboot. No labels are allocated from this preserved range. When you enable Segment Routing with the default SRLB/SRGB in the future, these label ranges will be available and ready for use.

The LSD allocates dynamic labels starting from 24,000.

Note If an MPLS label range is configured and it overlaps with the default SRLB/SRGB label ranges (for example, **mpls label range 15000 1048575**), then the default SRLB/SRGB preservation is disabled.

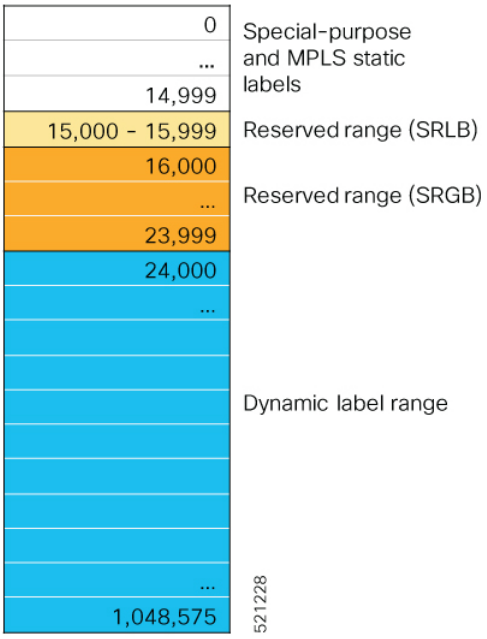
Example 1: LSD Label Allocation When SR is not Configured

- Special use: 0-15
- MPLS static: 16 to 14,999
- SRLB (preserved): 15,000 to 15,999
- SRGB (preserved): 16,000 to 23,999
- Dynamic: 24,000 to max



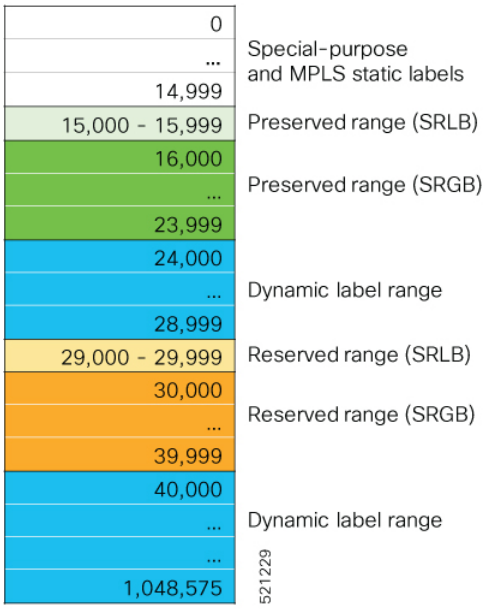
Example 2: LSD Label Allocation When SR is Configured with Default SRGB and Default SRLB

- Special use: 0-15
- MPLS static: 16 to 14,999
- SRLB (reserved): 15,000 to 15,999
- SRGB (reserved): 16,000 to 23,999
- Dynamic: 24,000 to max



Example 3: LSD Label Allocation When SR is Configured with Non-default SRGB and Non-default SRLB

- Special use: 0-15
- MPLS static: 16 to 14,999
- SRLB (preserved): 15,000 to 15,999
- SRGB (preserved): 16,000 to 23,999
- Dynamic: 24000 to 28,999
- SRLB (reserved): 29,000 to 29,999
- SRGB (reserved): 30,000 to 39,999
- Dynamic: 40,000 to max



Setup a Non-Default Segment Routing Global Block Range

This task explains how to configure a non-default SRGB range.

SUMMARY STEPS

1. **configure**
2. **segment-routing global-block** *starting_value ending_value*
3. Use the **commit** or **end** command.

DETAILED STEPS

	Command or Action	Purpose
Step 1	**configure** **Example:** `RP/0/RP0/CPU0:router# configure`	Enters mode.
Step 2	**segment-routing global-block** *starting_value ending_value* **Example:** `RP/0/RP0/CPU0:router(config)# segment-routing global-block 16000 80000`	Enter the lowest value that you want the SRGB range to include as the starting value. Enter the highest value that you want the SRGB range to include as the ending value.
Step 3	Use the **commit** or **end** command.	**commit** —Saves the configuration changes and remains within the configuration session. **end** —Prompts user to take one of these actions:

	Command or Action	Purpose
		• **Yes** — Saves configuration changes and exits the configuration session. • **No** —Exits the configuration session without committing the configuration changes. • **Cancel** —Remains in the configuration session, without committing the configuration changes.

Use the **show mpls label table** [**label** *label-value*] command to verify the SRGB configuration:

```
Router# show mpls label table label 16000 detail
Table Label   Owner                              State  Rewrite
----- ------- -------------------------------- ------ -------
0     16000   ISIS(A):1                        InUse  No
  (Lbl-blk SRGB, vers:0, (start_label=16000, size=64001)
```

What to do next

Configure prefix SIDs and enable segment routing.

Setup a Non-Default Segment Routing Local Block Range

This task explains how to configure a non-default SRLB range.

SUMMARY STEPS

1. **configure**
2. **segment-routing local-block** *starting_value ending_value*
3. Use the **commit** or **end** command.

DETAILED STEPS

	Command or Action	Purpose
Step 1	**configure** **Example:** `RP/0/RP0/CPU0:router# configure`	Enters mode.
Step 2	**segment-routing local-block** *starting_value ending_value* **Example:** `RP/0/RP0/CPU0:router(config)# segment-routing local-block 30000 30999`	Enter the lowest value that you want the SRLB range to include as the starting value. Enter the highest value that you want the SRLB range to include as the ending value.

	Command or Action	Purpose
Step 3	Use the **commit** or **end** command.	**commit** —Saves the configuration changes and remains within the configuration session. **end** —Prompts user to take one of these actions: • **Yes** — Saves configuration changes and exits the configuration session. • **No** —Exits the configuration session without committing the configuration changes. • **Cancel** —Remains in the configuration session, without committing the configuration changes.

Use the **show mpls label table** [**label** *label-value*] [**detail**] command to verify the SRLB configuration:

```
Router# show mpls label table label 30000 detail

Table Label   Owner                           State  Rewrite
----- ------- ------------------------------- ------ -------
0     30000   LSD(A)                          InUse  No
  (Lbl-blk SRLB, vers:0, (start_label=30000, size=1000, app_notify=0)

Router# show segment-routing local-block inconsistencies

No inconsistencies
```

The following example shows an SRLB label conflict in the range of 30000 and 30999. Note that the default SRLB is active and the configured SRLB is pending:

```
Router(config)# segment-routing local-block 30000 30999

%ROUTING-MPLS_LSD-3-ERR_SRLB_RANGE : SRLB allocation failed: 'SRLB reservation not successfull

for [30000,30999]. Use with caution 'clear segment-routing local-block discrepancy all'
command
to force srlb allocation'
```

Caution You can use the **clear segment-routing local-block discrepancy all** command to clear label conflicts. However, using this command is disruptive for traffic since it forces all other MPLS applications with conflicting labels to allocate new labels.

```
Router# show mpls label table label 30000 detail

Table Label   Owner                           State  Rewrite
----- ------- ------------------------------- ------ -------
0     30000   LSD(A)                          InUse  No
  (Lbl-blk SRLB, vers:0, (start_label=30000, size=1000, app_notify=0)

Router# show segment-routing local-block inconsistencies
SRLB inconsistencies range: Start/End: 30000/30999
```

```
Router# show mpls lsd private | i SRLB

SRLB Lbl Mgr:
   Current Active SRLB block      = [15000, 15999]
   Configured Pending SRLB block  = [30000, 30999]
```

Reload the router to release the currently allocated labels and to allocate the new SRLB:

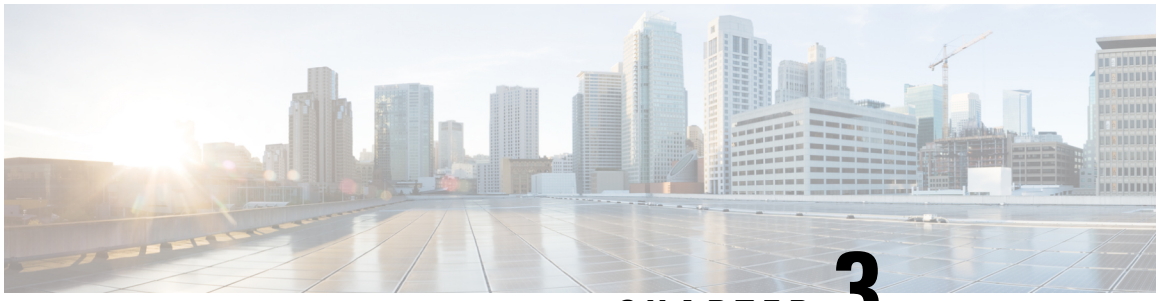
```
Router# reload

Proceed with reload? [confirm]yes
```

After the system is brought back up, verify that there are no label conflicts with the SRLB configuration:

```
Router# show mpls lsd private | i SRLB

SRLB Lbl Mgr:
   Current Active SRLB block      = [30000, 30999]
   Configured Pending SRLB block  = [0, 0]

Router# show segment-routing local-block inconsistencies

No inconsistencies
```

What to do next

Configure adjacency SIDs and enable segment routing.



CHAPTER 3

Configure Segment Routing for IS-IS Protocol

Integrated Intermediate System-to-Intermediate System (IS-IS), Internet Protocol Version 4 (IPv4), is a standards-based Interior Gateway Protocol (IGP). The Cisco IOS XR software implements the IP routing capabilities described in International Organization for Standardization (ISO)/International Engineering Consortium (IEC) 10589 and RFC 1995, and adds the standard extensions for single topology and multitopology IS-IS for IP Version 6 (IPv6).

This module provides the configuration information used to enable segment routing for IS-IS.

Note For additional information on implementing IS-IS on your Cisco 8000 Series Router, see the *Implementing IS-IS* module in the *Routing Configuration Guide for Cisco 8000 Series Routers*.

Enabling Segment Routing for IS-IS Protocol

Segment routing on the IS-IS control plane supports the following:

- IPv4 and IPv6 control plane
- Level 1, level 2, and multi-level routing
- Prefix SIDs for host prefixes on loopback interfaces
- Adjacency SIDs for adjacencies
- MPLS penultimate hop popping (PHP) and explicit-null signaling

This task explains how to enable segment routing for IS-IS.

Before you begin

Your network must support the MPLS Cisco IOS XR software feature before you enable segment routing for IS-IS on your router.

Note You must enter the commands in the following task list on every IS-IS router in the traffic-engineered portion of your network.

SUMMARY STEPS

1. **configure**
2. **router isis** *instance-id*
3. **address-family** { **ipv4** | **ipv6** } [**unicast**]
4. **metric-style wide** [**level** { **1** | **2** }]
5. **router-id loopback** *loopback interface used for prefix-sid*
6. **segment-routing mpls** [**sr-prefer**]
7. **exit**
8. Use the **commit** or **end** command.

DETAILED STEPS

	Command or Action	Purpose
Step 1	**configure** **Example:** `RP/0/RP0/CPU0:router# configure`	Enters mode.
Step 2	**router isis** *instance-id* **Example:** `RP/0/RP0/CPU0:router(config)# router isis isp`	Enables IS-IS routing for the specified routing instance, and places the router in router configuration mode. **Note** You can change the level of routing to be performed by a particular routing instance by using the **is-type** router configuration command.
Step 3	**address-family** { **ipv4**	**ipv6** } [**unicast**] **Example:** `RP/0/RP0/CPU0:router(config-isis)# address-family ipv4 unicast`	Specifies the IPv4 or IPv6 address family, and enters router address family configuration mode.
Step 4	**metric-style wide** [**level** { **1**	**2** }] **Example:** `RP/0/RP0/CPU0:router(config-isis-af)# metric-style wide level 1`	Configures a router to generate and accept only wide link metrics in the Level 1 area.
Step 5	**router-id loopback** *loopback interface used for prefix-sid* **Example:** `RP/0/(config-isis-af)#router-id loopback0`	Configures router ID for each address-family (ipv4/ipv6).
Step 6	**segment-routing mpls** [**sr-prefer**]	Segment routing is enabled by the following actions:

	Command or Action	Purpose

	Example:
	
	`RP/0/RP0/CPU0:router(config-isis-af)#`
	`segment-routing mpls`
		• MPLS forwarding is enabled on all interfaces where IS-IS is active.
		• All known prefix-SIDs in the forwarding plain are programmed, with the prefix-SIDs advertised by remote routers or learned through local or remote mapping server.
		• The prefix-SIDs locally configured are advertised.
		
		Use the **sr-prefer** keyword to set the preference of segment routing (SR) labels over label distribution protocol (LDP) labels.

Step 7	**exit**
	Example:
	
	`RP/0/RP0/CPU0:router(config-isis-af)# exit`
	`RP/0/RP0/CPU0:router(config-isis)# exit`	

Step 8	Use the **commit** or **end** command.	**commit** —Saves the configuration changes and remains within the configuration session. **end** —Prompts user to take one of these actions: • **Yes** — Saves configuration changes and exits the configuration session. • **No** —Exits the configuration session without committing the configuration changes. • **Cancel** —Remains in the configuration session, without committing the configuration changes.

What to do next

Configure the prefix SID.

Configuring a Prefix-SID on the IS-IS Enabled Loopback Interface

Table 1: Feature History Table

Feature Name	Release Information	Feature Description
Disable Penultimate Hop Popping	Release 7.3.5	You can now disable the penultimate hop popping (PHP) without adding an explicit-Null label. In earlier releases, you could disable PHP only by adding an explicit-Null label using the explicit-null keyword. The feature introduces the **php-disable** keyword under the **prefix-sid** command.

A prefix segment identifier (SID) is associated with an IP prefix. The prefix SID is manually configured from the segment routing global block (SRGB) range of labels. A prefix SID is configured under the loopback interface with the loopback address of the node as the prefix. The prefix segment steers the traffic along the shortest path to its destination.

A prefix SID can be a node SID or an Anycast SID. A node SID is a type of prefix SID that identifies a specific node. An Anycast SID is a type of prefix SID that identifies a set of nodes, and is configured with n-flag clear. The set of nodes (Anycast group) is configured to advertise a shared prefix address and prefix SID. Anycast routing enables the steering of traffic toward multiple advertising nodes. Packets addressed to an Anycast address are forwarded to the topologically nearest nodes.

Strict-SPF SIDs are used to forward traffic strictly along the SPF path. IS-IS advertises the SR Algorithm sub Type Length Value (TLV) (in the SR Router Capability SubTLV) to include both algorithm 0 (SPF) and algorithm 1 (Strict-SPF). Strict-SPF SIDs are also used to program the backup paths for prefixes, node SIDs, and adjacency SIDs.

The prefix SID is globally unique within the segment routing domain.

This task explains how to configure prefix segment identifier (SID) index or absolute value on the IS-IS enabled Loopback interface.

Before you begin

Ensure that segment routing is enabled on the corresponding address family.

SUMMARY STEPS

1. **configure**
2. **router isis** *instance-id*
3. **interface Loopback** *instance*
4. **address-family** { **ipv4** | **ipv6** } [**unicast**]
5. **prefix-sid** [**algorithm** *algorithm-number*] {**index** *SID-index* | **absolute** *SID-value*} [**n-flag-clear**] [**explicit-null**]

6. Use the **commit** or **end** command.

DETAILED STEPS

	Command or Action	Purpose
Step 1	**configure** **Example:** `RP/0/RP0/CPU0:router# configure`	Enters XR Config mode.
Step 2	**router isis** *instance-id* **Example:** `RP/0/RP0/CPU0:router(config)# router isis 1`	Enables IS-IS routing for the specified routing instance, and places the router in router configuration mode. • You can change the level of routing to be performed by a particular routing instance by using the **is-type** router configuration command.
Step 3	**interface Loopback** *instance* **Example:** `RP/0/RP0/CPU0:router(config-isis)# interface Loopback0`	Specifies the loopback interface and instance.
Step 4	**address-family** { **ipv4** | **ipv6** } [**unicast**] **Example:** The following is an example for ipv4 address family: `RP/0/RP0/CPU0:router(config-isis-if)# address-family ipv4 unicast`	Specifies the IPv4 or IPv6 address family, and enters router address family configuration mode.

Step 5	**prefix-sid** [**algorithm** *algorithm-number*] {**index** *SID-index* | **absolute** *SID-value*} [**n-flag-clear**] [**explicit-null**]
	Example:
	`RP/0/RP0/CPU0:router(config-isis-if-af)# prefix-sid index 1001`
	`RP/0/RP0/CPU0:router(config-isis-if-af)# prefix-sid absolute 17001`
		Configures the prefix-SID index or absolute value for the interface.
		Specify **algorithm** *algorithm-number* to configure SR Flexible Algorithm. See Enabling Segment Routing Flexible Algorithm, on page 99.
		Specify **index** *SID-index* for each node to create a prefix SID based on the lower boundary of the SRGB + the index.
		Specify **absolute** *SID-value* for each node to create a specific prefix SID within the SRGB.
		By default, the n-flag is set on the prefix-SID, indicating that it is a node SID. For specific prefix-SID (for example, Anycast prefix-SID), enter the **n-flag-clear** keyword. IS-IS does not set the N flag in the prefix-SID sub Type Length Value (TLV).
		To disable penultimate-hop-popping (PHP) and add explicit-Null label, enter **explicit-null** keyword. IS-IS sets the E flag in the prefix-SID sub TLV. Any upstream

	Command or Action	Purpose
		neighbor of the Prefix-SID originator replaces the Prefix-SID with a Prefix-SID having an Explicit NULL value.
Step 6	Use the **commit** or **end** command.	**commit** —Saves the configuration changes and remains within the configuration session. **end** —Prompts user to take one of these actions: • **Yes** — Saves configuration changes and exits the configuration session. • **No** —Exits the configuration session without committing the configuration changes. • **Cancel** —Remains in the configuration session, without committing the configuration changes.

Verify the prefix-SID configuration:

```
RP/0/RP0/CPU0:router# show isis database verbose

IS-IS 1 (Level-2) Link State Database
LSPID                 LSP Seq Num  LSP Checksum  LSP Holdtime  ATT/P/OL
router.00-00        * 0x0000039b   0xfc27        1079            0/0/0
  Area Address: 49.0001
  NLPID:        0xcc
  NLPID:        0x8e
  MT:           Standard (IPv4 Unicast)
  MT:           IPv6 Unicast                                     0/0/0
  Hostname:     router
  IP Address:   10.0.0.1
  IPv6 Address: 2001:0db8:1234::0a00:0001
  Router Cap:   10.0.0.1, D:0, S:0
    Segment Routing: I:1 V:1, SRGB Base: 16000 Range: 8000
    SR Algorithm:
      Algorithm: 0
      Algorithm: 1
<...>
  Metric: 0          IP-Extended 10.0.0.1/32
    Prefix-SID Index: 1001, Algorithm:0, R:0 N:1 P:0 E:0 V:0 L:0
    Prefix-SID Index: 101, Algorithm:1, R:0 N:1 P:0 E:0 V:0 L:0
<...>
```

Overriding MPLS Imposition (IP-to-MPLS) via Service Layer API (SL-API)

Table 2: Feature History Table

Feature Name	Release Information	Feature Description
Overriding MPLS Imposition (IP-to-MPLS) via Service Layer API (SL-API)	Release 24.2.1	In scenarios where SR-prefer is enabled, this feature allows you to specify SR prefixes through an Access Control List where their imposition forwarding entry (IP-to-MPLS) gives preference to SL-API, instead of the SR native LSP. The labeled forwarding entries (MPLS-to-MPLS or MPLS-to-IP) continue to follow the SR native LSP. This feature introduces the following command under Router RIB AF configuration mode: **segment-routing mpls preserve-label-forwarding access-list***acl_name* **[apply-inverse]**

Note For detailed information about Service Layer API (SL-API), refer to "Use Service Layer API to Bring your Controller on Cisco IOS XR Router" of the *Programmability Configuration Guide for Cisco 8000 Series Routers*.

Usage Guidelines and Limitations

The following usage guidelines and limitations apply:

- This feature is applicable when an SR prefix destination is also programmed via SL-API and “sr-prefer” is also enabled due to the presence of other prefixes with both SR and LDP LSPs.
- If the feature is configured for selected (allowed) prefixes, the “sr-prefer” configuration is ignored and the imposition forwarding entry follows the SL-API path instead of the SR native LSP.
- If the feature is not configured, or if a prefix is not allowed for SL-API steering, the “sr-prefer” configuration is honored and the imposition forwarding entry follows the SR native LSP.
- When there is a single source of programming for a destination (SR or SL-API), this feature has no impact on the forwarding.
- This feature is supported for programming of IPv4 SR prefixes.
- This feature is supported for programming of IPv6 SR prefixes.

- This feature does not support forwarding of traffic to IPv4 destinations recursing onto IPv6 next-hops steered over SL-API paths (BGPv4 over SRMPLS-v6).
- Redistribution of SL-API imposition route into another protocol is not supported.
- SR native and SL-API paths must always be labelled.
- The set of prefixes allowed for SR and LDP must be disjointed from the set of prefixes allowed for SR and SL-API. SR/SL-API and SR/LDP can co-exist across different SR prefixes.

Use Case

Assume a node is part of a network with SR and LDP enabled concurrently (ships-in-the-night) with preference to SR over LDP when both LSPs are present (sr-prefer).

The network operator relies on a controller to program a desired traffic-engineered path for specific prefix destinations using SL-API. The following forwarding behaviors are expected at the node programmed via SL-API:

- Imposition forwarding entry (IP-to-MPLS) gives preference to the SL-API LSP
- Labeled forwarding entries (MPLS-to-MPLS or MPLS-to-IP) follow the SR native LSP

Transport Without SL-API Injection

Consider the following :

- A network with SR and LDP enabled concurrently
- Nodes are configured with SR-prefer enabled
- Prefix 10.1.1.2/32 (SR prefix SID 16002) is programmed with SR native LSP
- Prefix 10.1.1.3/32 (SR prefix SID 16003) is programmed with SR native LSP
- Prefix 10.1.1.4/32 (SR prefix SID 16004) is programmed with both SR native and LDP LSPs
- When required, a controller is used to program a desired traffic-engineered path for allowed destination prefixes via an SL-API:
  - Allowed prefixes for controller steering: 10.1.1.2/32, 10.1.1.3/32
  - Not allowed prefix for controller steering: 10.1.1.4/32

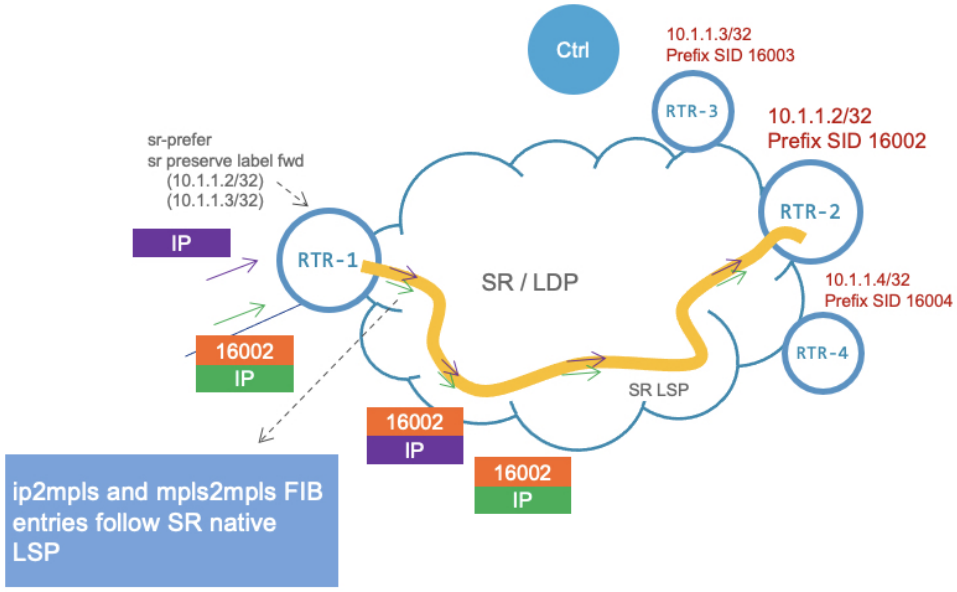
When the controller does not trigger an SL-API path for allowed prefixes, the imposition forwarding entry follows the SR native LSP instead of the LDP LSP (as a result of configuring sr-prefer). The Swap/Pop forwarding entries are programmed to follow both SR native and LDP LSPs.

For SR prefixes 10.1.1.2/32 and 10.1.1.3/32:

- Imposition (IP2MPLS):
  - SR Prefix SID push forwarding entry programmed by IGP
- Swap/Pop (MPLS2MPLS/MPLS2IP):
  - SR prefix SID local label swap/pop forwarding entry programmed by IGP

For SR prefix 10.1.1.4/32:

- Imposition (IP2MPLS):
  - SR Prefix SID push forwarding entry programmed by IGP
- Swap/Pop (MPLS2MPLS/MPLS2IP):
  - SR prefix SID local label swap/pop forwarding entry programmed by IGP
  - LDP local label swap/pop forwarding entry programmed by LDP



Transport After SL-API Injection

When the controller triggers an SL-API path for an allowed destination prefix (for example 10.1.1.2), the imposition forwarding entry will follow the SL-API LSP instead of the SR native LSP.

The imposition forwarding entry for allowed prefixes but not programmed by SL-API (for example 10.1.1.3/32), or not allowed prefixes (for example, 10.1.1.4), will follow the SR native LSP.

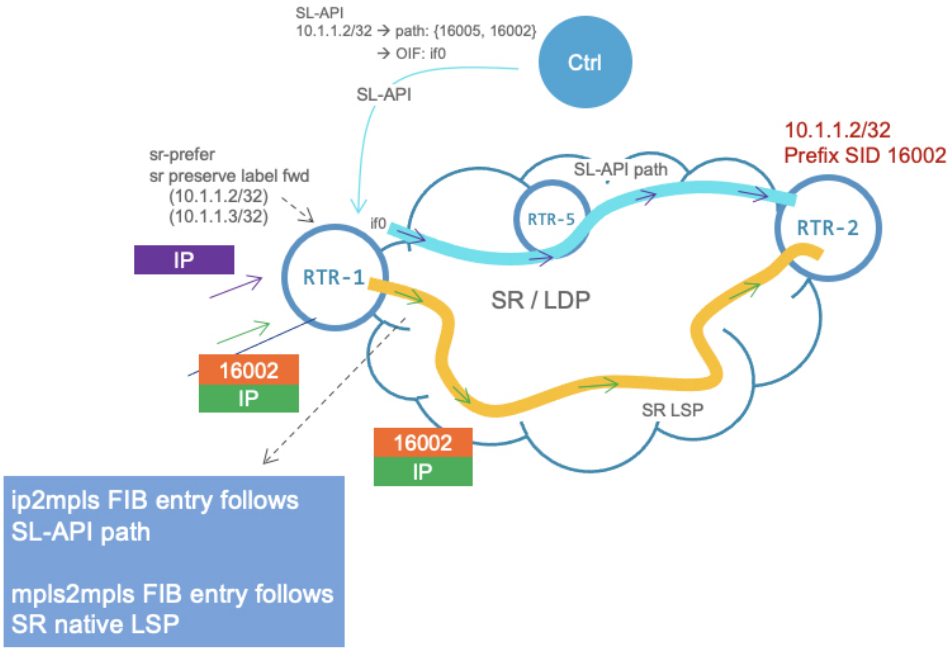
The Swap/Pop forwarding entries for the SR-only prefixes (10.1.1.2/32, 10.1.1.3/32) continue to be programmed by SR.

The Swap/Pop forwarding entries for the SR/LDP prefix (10.1.1.4/32) continue to be programmed by both SR and LDP.

For Prefix 10.1.1.2/32:

- Imposition (IP2MPLS):
  - Label stack push forwarding entry programed by SL-API
- Swap/Pop (MPLS2MPLS/MPLS2IP):
  - SR prefix SID local label swap/pop forwarding entry programmed by IGP

There are no changes to the forwarding entries for prefixes 10.1.1.3/32 and 10.1.1.4/32.



Configuration

To enable overriding of MPLS label imposition via SL-API, use the **segment-routing mpls preserve-label-forwarding access-list** *acl_name* **[apply-inverse]** command under **router rib address-family** {**ipv4**|**ipv6**} configuration mode.

Example

The following example shows how to allow IPv4 SR prefix 10.1.1.2/32 to have it MPLS imposition forwarding entry to be override via SL-API:

```
RP/0/RP0/CPU0:ios(config)# ipv4 access-list SL-API-PREFER-ALLOWED-PFX
RP/0/RP0/CPU0:ios(config-ipv4-acl)# 10 permit 10.1.1.2
RP/0/RP0/CPU0:ios(config-ipv4-acl)# exit

RP/0/RP0/CPU0:ios(config)# router rib address-family ipv4
RP/0/RP0/CPU0:ios(config-rib-afi)# segment-routing mpls preserve-label-forwarding access-list
 SL-API-PREFER-ALLOWED-PFX
```

The following example shows how to allow any IPv4 SR prefix except 10.1.1.2/32 to have it MPLS imposition forwarding entry to be override via SL-API:

```
RP/0/RP0/CPU0:ios(config)# ipv4 access-list SL-API-PREFER-DISALLOWED-PFX
RP/0/RP0/CPU0:ios(config-ipv4-acl)# 10 permit 10.1.1.2
RP/0/RP0/CPU0:ios(config-ipv4-acl)# exit

RP/0/RP0/CPU0:ios(config)# router rib address-family ipv4
RP/0/RP0/CPU0:ios(config-rib-afi)# segment-routing mpls preserve-label-forwarding access-list
 SL-API-PREFER-DISALLOWED-PFX apply-inverse
```

Verification

Consider the following SR prefix as allowed to be steered over an SL-API path:

- Prefix: 10.1.1.1/32
- Prefix SID: 20000

The SR native LSP programmed at a node in the network is as follows:

- Local label: 20000
- ECMP:
  - Path0 – out label: 20000; out int: Bundle-Ether20131
  - Path1 – out label: 20000; out int: Bundle-Ether20132
  - Path2 – out label: 20000; out int: Bundle-Ether20133
  - Path3 – out label: 20000; out int: Bundle-Ether20134

The SL-API LSP to be programmed by the controller at the same node is as follows:

- Local label: 100051 (dynamically allocated)
- Weighted ECMP:
  - Path0 – out label: 24000, 18001; out int: Bundle-Ether2012; weight: 1
  - Path1 – out label: 24000, 18001; out int: Bundle-Ether2013; weight: 2
  - Path2 – out label: 24001, 18001; out int: Bundle-Ether2014; weight: 4
  - Path3 – out label: 24001, 18001; out int: Bundle-Ether2015; weight: 8

The following sequence of show command outputs can be used to verify the programming of the imposition forwarding entry when an SL-API path is present.

The following output shows the RIB entry for an allowed prefix highlighting the fields to indicate that overriding of MPLS label imposition via SL-API is enabled:

```
Router# show route 10.1.1.1/32 detail

Routing entry for 10.1.1.1/32
  Known via "isis 0", distance 115, metric 500, labeled SR (label forwarding preserve),
type level-2
  Installed Jul 30 21:23:50.539 for 01:43:09
  Routing Descriptor Blocks
    100.201.201.2, from 199.1.0.2, via Bundle-Ether20131
      Route metric is 500
      Label: 0x4e20 (20000)
      Tunnel ID: None
      Binding Label: None
      Extended communities count: 0
      Path id:4       Path ref count:0
      NHID:0x0(Ref:0)
      MPLS eid:0x109c700000001
    101.201.201.2, from 199.1.0.2, via Bundle-Ether20132
      Route metric is 500
      Label: 0x4e20 (20000)
      Tunnel ID: None
```

```
      Binding Label: None
      Extended communities count: 0
      Path id:3       Path ref count:0
      NHID:0x0(Ref:0)
      MPLS eid:0x109c700000001
    102.201.201.2, from 199.1.0.2, via Bundle-Ether20133
      Route metric is 500
      Label: 0x4e20 (20000)
      Tunnel ID: None
      Binding Label: None
      Extended communities count: 0
      Path id:2       Path ref count:0
      NHID:0x0(Ref:0)
      MPLS eid:0x109c700000001
    103.201.201.2, from 199.1.0.2, via Bundle-Ether20134
      Route metric is 500
      Label: 0x4e20 (20000)
      Tunnel ID: None
      Binding Label: None
      Extended communities count: 0
      Path id:1       Path ref count:0
      NHID:0x0(Ref:0)
      MPLS eid:0x109c700000001
  Route version is 0x47 (71)
  Local Label: 0x4e20 (20000)
  IP Precedence: Not Set
  QoS Group ID: Not Set
  Flow-tag: Not Set
  Fwd-class: Not Set
  Route Priority: RIB_PRIORITY_NON_RECURSIVE_MEDIUM (7) SVD Type RIB_SVD_TYPE_LOCAL
  Download Priority 1, Download Version 1793240
  Route eid: 0x109c700000001
  No advertising protos.
```

The following output shows the imposition forwarding entry programmed via SL-API (rewrite owner) in LSD. Observe the programmed paths and their parameters (output interface, output labels, and weight).

```
Router# show mpls lsd forwarding ipv4 detail | begin 10.1.1.1/32

'default':4U, 10.1.1.1/32, (100051)[SR Merge], 4 Paths,
       Owner=Static(A):Service-layer
   1/4: IPv4_STACK, 'default':4U, BE2012, BSID: NO_LABEL, nh=100.201.200.2, lbls={ 24000,
18001 }
            lbl flags= {0x0 0x0} ()}, ext_flags=0x0 path_flags=0x0
            nh-id=0x0, path-id=0, backup-path-id=0, load-metric=32, parent-intf=None,
path-set-id=0, path-priority=0
            Inner Stack Flags:  { 0x0}
            MPLS eid: N/A
   2/4: IPv4_STACK, 'default':4U, BE2013, BSID: NO_LABEL, nh=101.201.200.2, lbls={ 24000,
18001 }
            lbl flags= {0x0 0x0} ()}, ext_flags=0x0 path_flags=0x0
            nh-id=0x0, path-id=0, backup-path-id=0, load-metric=64, parent-intf=None,
path-set-id=0, path-priority=0
            Inner Stack Flags:  { 0x0}
            MPLS eid: N/A
   3/4: IPv4_STACK, 'default':4U, BE2014, BSID: NO_LABEL, nh=102.201.200.2, lbls={ 24001,
18001 }
            lbl flags= {0x0 0x0} ()}, ext_flags=0x0 path_flags=0x0
            nh-id=0x0, path-id=0, backup-path-id=0, load-metric=128, parent-intf=None,
path-set-id=0, path-priority=0
            Inner Stack Flags:  { 0x0}
            MPLS eid: N/A
   4/4: IPv4_STACK, 'default':4U, BE2015, BSID: NO_LABEL, nh=103.201.200.2, lbls={ 24001,
18001 }
```

```
            lbl flags= {0x0 0x0} ()}, ext_flags=0x0 path_flags=0x0
            nh-id=0x0, path-id=0, backup-path-id=0, load-metric=256, parent-intf=None,
path-set-id=0, path-priority=0
            Inner Stack Flags:  { 0x0}
            MPLS eid: N/A
      BCDL priority:1, LSD queue:9, version:178429,
      flags: 0x8, fwd_flags: 0x100 (sr_lbl_fwd_preserve),
      Installed Jul 30 21:26:25.111 (01:42:45 ago)
      Prefix eid: 0x1275100000001

. . .
```

The following output shows that the CEF imposition forwarding entry prefers the SL-API path.

```
Router# show cef 10.1.1.1/32 detail location 0/0/CPU0

10.1.1.1/32, version 178429, internal 0x1000001 0x110 (ptr 0xa0ea1428) [3], 0x0 (0x1182a378),
 0xa28 (0x202c3ba8)
 Updated Jul 30 21:26:25.635
 local adjacency to Bundle-Ether2012

 Prefix Len 32, traffic index 0, precedence n/a, priority 1, encap-id 0x1275100000001
  gateway array (0x1d140ef8) reference count 750, flags 0x68, source lsd (5), 1 backups
                [251 type 5 flags 0x8401 (0x8c18bda0) ext 0x0 (0x0)]
  LW-LDI[type=5, refc=3, ptr=0x1182a378, sh-ldi=0x8c18bda0]
  gateway array update type-time 1 Jul 30 21:26:25.532
 LDI Update time Jul 30 21:26:25.532
 LW-LDI-TS Jul 30 21:26:25.635
   via 100.201.200.2/32, Bundle-Ether2012, 7 dependencies, weight 32, class 0 [flags 0x0]
    path-idx 0 NHID 0x0 [0x90fa8028 0x0]
    next hop 100.201.200.2/32
    local adjacency
     local label 100051      labels imposed {24000 18001}
   via 101.201.200.2/32, Bundle-Ether2013, 7 dependencies, weight 64, class 0 [flags 0x0]
    path-idx 1 NHID 0x0 [0x90fa85c8 0x0]
    next hop 101.201.200.2/32
    local adjacency
     local label 100051      labels imposed {24000 18001}
   via 102.201.200.2/32, Bundle-Ether2014, 7 dependencies, weight 128, class 0 [flags 0x0]
    path-idx 2 NHID 0x0 [0x90fa82f8 0x0]
    next hop 102.201.200.2/32
    local adjacency
     local label 100051      labels imposed {24001 18001}
   via 103.201.200.2/32, Bundle-Ether2015, 7 dependencies, weight 256, class 0 [flags 0x0]
    path-idx 3 NHID 0x0 [0x90fa8190 0x0]
    next hop 103.201.200.2/32
    local adjacency
     local label 100051      labels imposed {24001 18001}

    Weight distribution:
    slot 0, weight 32, normalized_weight 1, class 0
    slot 1, weight 64, normalized_weight 2, class 0
    slot 2, weight 128, normalized_weight 4, class 0
    slot 3, weight 256, normalized_weight 8, class 0
    Load distribution: 0 1 1 2 2 2 2 3 3 3 3 3 3 3 3 (refcount 251)

    Hash  OK  Interface                 Address
    0     Y   Bundle-Ether2012          100.201.200.2
    1     Y   Bundle-Ether2013          101.201.200.2
    2     Y   Bundle-Ether2013          101.201.200.2
    3     Y   Bundle-Ether2014          102.201.200.2
    4     Y   Bundle-Ether2014          102.201.200.2
    5     Y   Bundle-Ether2014          102.201.200.2
```

```
    6     Y   Bundle-Ether2014          102.201.200.2
    7     Y   Bundle-Ether2015          103.201.200.2
    8     Y   Bundle-Ether2015          103.201.200.2
    9     Y   Bundle-Ether2015          103.201.200.2
    10    Y   Bundle-Ether2015          103.201.200.2
    11    Y   Bundle-Ether2015          103.201.200.2
    12    Y   Bundle-Ether2015          103.201.200.2
    13    Y   Bundle-Ether2015          103.201.200.2
    14    Y   Bundle-Ether2015          103.201.200.2
```

The following output shows the backup CEF imposition forwarding entry if the SL-API path is not present. Observe that it follows the SR native LSP.

```
Router# show cef 10.1.1.1/32 backup detail location 0/0/CPU0

10.1.1.1/32, version 1793240, priority 1, flags 0x200000, flags2 0x81, extn_flags 0x2100,
source rib (7), ctx-flags 0xc1
 Updated Jul 30 21:23:50.819
 Prefix Len 32
 Label count = 1, src = 7, label = 20000
   via  BE20131 (0xf013954) 100.201.201.2, weight 0, class 0 [flags 0x0]
    next hop VRF - 'default', table - 0xe0000000
    Output labels {20000}
   via  BE20132 (0xf01395c) 101.201.201.2, weight 0, class 0 [flags 0x0]
    next hop VRF - 'default', table - 0xe0000000
    Output labels {20000}
   via  BE20133 (0xf013964) 102.201.201.2, weight 0, class 0 [flags 0x0]
    next hop VRF - 'default', table - 0xe0000000
    Output labels {20000}
   via  BE20134 (0xf01396c) 103.201.201.2, weight 0, class 0 [flags 0x0]
    next hop VRF - 'default', table - 0xe0000000
    Output labels {20000}
```

The following output shows the labeled forwarding entry (MPLS-to-MPLS) for the SR prefix SID local label (20000). Observe that it follows the SR native LSP.

```
Router# show mpls forwarding labels 20000 location 0/0/CPU0

Local  Outgoing    Prefix             Outgoing     Next Hop        Bytes
Label  Label       or ID              Interface                    Switched
------ ----------- ------------------ ------------ --------------- ------------
20000  20000       SR Pfx (idx 4000)  BE20131      100.201.201.2   0
       20000       SR Pfx (idx 4000)  BE20132      101.201.201.2   0
       20000       SR Pfx (idx 4000)  BE20133      102.201.201.2   0
       20000       SR Pfx (idx 4000)  BE20134      103.201.201.2   0
```

The following output shows the details for the labeled forwarding entry (MPLS-to-MPLS) for the SR prefix SID local label (20000).

```
Router# show mpls forwarding labels 20000 detail location 0/0/CPU0

Local  Outgoing    Prefix             Outgoing     Next Hop        Bytes
Label  Label       or ID              Interface                    Switched
------ ----------- ------------------ ------------ --------------- ------------
20000  20000       SR Pfx (idx 4000)  BE20131      100.201.201.2   0
     Updated: Jul 30 21:26:25.158
     Version: 1793240, Priority: 15
     Label Stack (Top -> Bottom): { 20000 }
     NHID: 0x0, Encap-ID: N/A, Path idx: 0, Backup path idx: 0, Weight: 0
     MAC/Encaps: 14/18, MTU: 1500
     Outgoing Interface: Bundle-Ether20131 (ifhandle 0x0f013954)
```

```
    Packets Switched: 0

      20000       SR Pfx (idx 4000)  BE20132      101.201.201.2   0
    Updated: Jul 30 21:26:25.158
    Version: 1793240, Priority: 15
    Label Stack (Top -> Bottom): { 20000 }
    NHID: 0x0, Encap-ID: N/A, Path idx: 1, Backup path idx: 0, Weight: 0
    MAC/Encaps: 14/18, MTU: 1500
    Outgoing Interface: Bundle-Ether20132 (ifhandle 0x0f01395c)
    Packets Switched: 0

      20000       SR Pfx (idx 4000)  BE20133      102.201.201.2   0
    Updated: Jul 30 21:26:25.158
    Version: 1793240, Priority: 15
    Label Stack (Top -> Bottom): { 20000 }
    NHID: 0x0, Encap-ID: N/A, Path idx: 2, Backup path idx: 0, Weight: 0
    MAC/Encaps: 14/18, MTU: 1500
    Outgoing Interface: Bundle-Ether20133 (ifhandle 0x0f013964)
    Packets Switched: 0

      20000       SR Pfx (idx 4000)  BE20134      103.201.201.2   0
    Updated: Jul 30 21:26:25.158
    Version: 1793240, Priority: 15
    Label Stack (Top -> Bottom): { 20000 }
    NHID: 0x0, Encap-ID: N/A, Path idx: 3, Backup path idx: 0, Weight: 0
    MAC/Encaps: 14/18, MTU: 1500
    Outgoing Interface: Bundle-Ether20134 (ifhandle 0x0f01396c)
    Packets Switched: 0

Traffic-Matrix Packets/Bytes Switched: 0/0
```

Configuring an Adjacency SID

An adjacency SID (Adj-SID) is associated with an adjacency to a neighboring node. The adjacency SID steers the traffic to a specific adjacency. Adjacency SIDs have local significance and are only valid on the node that allocates them.

An adjacency SID can be allocated dynamically from the dynamic label range or configured manually from the segment routing local block (SRLB) range of labels.

Adjacency SIDs that are dynamically allocated do not require any special configuration, however there are some limitations:

- A dynamically allocated Adj-SID value is not known until it has been allocated, and a controller will not know the Adj-SID value until the information is flooded by the IGP.
- Dynamically allocated Adj-SIDs are not persistent and can be reallocated after a reload or a process restart.
- Each link is allocated a unique Adj-SID, so the same Adj-SID cannot be shared by multiple links.

Manually allocated Adj-SIDs are persistent over reloads and restarts. They can be provisioned for multiple adjacencies to the same neighbor or to different neighbors. You can specify that the Adj-SID is protected. If the Adj-SID is protected on the primary interface and a backup path is available, a backup path is installed. By default, manual Adj-SIDs are not protected.

Adjacency SIDs are advertised using the existing IS-IS Adj-SID sub-TLV. The S and P flags are defined for manually allocated Adj-SIDs.

```
 0 1 2 3 4 5 6 7
+-+-+-+-+-+-+-+-+
|F|B|V|L|S|P|   |
+-+-+-+-+-+-+-+-+
```

Table 3: Adjacency Segment Identifier (Adj-SID) Flags Sub-TLV Fields

Field	Description
S (Set)	This flag is set if the same Adj-SID value has been provisioned on multiple interfaces.
P (Persistent)	This flag is set if the Adj-SID is persistent (manually allocated).

Manually allocated Adj-SIDs are supported on point-to-point (P2P) interfaces.

This task explains how to configure an Adj-SID on an interface.

Before you begin

Ensure that segment routing is enabled on the corresponding address family.

Use the **show mpls label table detail** command to verify the SRLB range.

SUMMARY STEPS

1. **configure**
2. **router isis** *instance-id*
3. **interface** *type interface-path-id*
4. **point-to-point**
5. **address-family** { **ipv4** | **ipv6** } [**unicast**]
6. **adjacency-sid** {**index** *adj-SID-index* | **absolute** *adj-SID-value* } [**protected**]
7. Use the **commit** or **end** command.

DETAILED STEPS

	Command or Action	Purpose
Step 1	**configure** **Example:** `RP/0/RP0/CPU0:router# configure`	Enters mode.
Step 2	**router isis** *instance-id* **Example:** `RP/0/RP0/CPU0:router(config)# router isis 1`	Enables IS-IS routing for the specified routing instance, and places the router in router configuration mode. • You can change the level of routing to be performed by a particular routing instance by using the **is-type** router configuration command.
Step 3	**interface** *type interface-path-id* **Example:**	Specifies the interface and enters interface configuration mode.

	Command or Action	Purpose
	`RP/0/RP0/CPU0:router(config-isis)# interface GigabitEthernet0/0/0/7`	
Step 4	**point-to-point** **Example:** `RP/0/RP0/CPU0:router(config-isis-if)# point-to-point`	Specifies the interface is a point-to-point interface.
Step 5	**address-family** { **ipv4** | **ipv6** } [**unicast**] **Example:** The following is an example for ipv4 address family: `RP/0/RP0/CPU0:router(config-isis-if)# address-family ipv4 unicast`	Specifies the IPv4 or IPv6 address family, and enters router address family configuration mode.

Step 6	**adjacency-sid** {**index** *adj-SID-index* | **absolute** *adj-SID-value* } [**protected**]
	Example:
	`RP/0/RP0/CPU0:router(config-isis-if-af)# adjacency-sid index 10`
	`RP/0/RP0/CPU0:router(config-isis-if-af)# adjacency-sid absolute 15010`
		Configures the Adj-SID index or absolute value for the interface.
		Specify **index** *adj-SID-index* for each link to create an Ajd-SID based on the lower boundary of the SRLB + the index.
		Specify **absolute** *adj-SID-value* for each link to create a specific Ajd-SID within the SRLB.
		Specify if the Adj-SID is **protected**. For each primary path, if the Adj-SID is protected on the primary interface and a backup path is available, a backup path is installed. By default, manual Adj-SIDs are not protected.

Step 7	Use the **commit** or **end** command.	**commit** —Saves the configuration changes and remains within the configuration session. **end** —Prompts user to take one of these actions: • **Yes** — Saves configuration changes and exits the configuration session. • **No** —Exits the configuration session without committing the configuration changes. • **Cancel** —Remains in the configuration session, without committing the configuration changes.

Verify the Adj-SID configuration:

```
RP/0/RP0/CPU0:router# show isis segment-routing label adjacency persistent
Mon Jun 12 02:44:07.085 PDT

IS-IS 1 Manual Adjacency SID Table
```

```
15010 AF IPv4
      GigabitEthernet0/0/0/3: IPv4, Protected 1/65/N, Active
      GigabitEthernet0/0/0/7: IPv4, Protected 2/66/N, Active

15100 AF IPv6
      GigabitEthernet0/0/0/3: IPv6, Not protected 255/255/N, Active
```

Verify the labels are added to the MPLS Forwarding Information Base (LFIB):

```
RP/0/RP0/CPU0:router# show mpls forwarding labels 15010
Mon Jun 12 02:50:12.172 PDT
Local  Outgoing    Prefix             Outgoing     Next Hop        Bytes
Label  Label       or ID              Interface                    Switched
------ ----------- ------------------ ------------ --------------- ------------
15010  Pop         SRLB (idx 10)      Gi0/0/0/3    10.0.3.3        0
       Pop         SRLB (idx 10)      Gi0/0/0/7    10.1.0.5        0
       16004       SRLB (idx 10)      Gi0/0/0/7    10.1.0.5        0            (!)
       16004       SRLB (idx 10)      Gi0/0/0/3    10.0.3.3        0            (!)
```

Manually Configure a Layer 2 Adjacency SID

Typically, an adjacency SID (Adj-SID) is associated with a Layer 3 adjacency to a neighboring node, to steer the traffic to a specific adjacency. If you have Layer 3 bundle interfaces, where multiple physical interfaces form a bundle interface, the individual Layer 2 bundle members are not visible to IGP; only the bundle interface is visible.

You can configure a Layer 2 Adj-SID for the individual Layer 2 bundle interfaces. This configuration allows you to track the availability of individual bundle member links and to verify the segment routing forwarding over the individual bundle member links, for Operational Administration and Maintenance (OAM) purposes.

A Layer 2 Adj-SID can be allocated dynamically or configured manually.

- IGP dynamically allocates Layer 2 Adj-SIDs from the dynamic label range for each Layer 2 bundle member. A dynamic Layer 2 Adj-SID is not persistent and can be reallocated as the Layer 3 bundle link goes up and down.
- Manually configured Layer 2 Adj-SIDs are persistent if the Layer 3 bundle link goes up and down. Layer 2 Adj-SIDs are allocated from the Segment Routing Local Block (SRLB) range of labels. However, if the configured value of Layer 2 Adj-SID does not fall within the available SRLB, a Layer 2 Adj-SID will not be programmed into forwarding information base (FIB).

Restrictions

- Adj-SID forwarding requires a next-hop, which can be either an IPv4 address or an IPv6 address, but not both. Therefore, manually configured Layer 2 Adj-SIDs are configured per address-family.
- Manually configured Layer 2 Adj-SID can be associated with only one Layer 2 bundle member link.
- A SID value used for Layer 2 Adj-SID cannot be shared with Layer 3 Adj-SID.
- SR-TE using Layer 2 Adj-SID is not supported.

This task explains how to configure a Layer 2 Adj-SID on an interface.

Before you begin

Ensure that segment routing is enabled on the corresponding address family.

Use the **show mpls label table detail** command to verify the SRLB range.

SUMMARY STEPS

1. **configure**
2. **segment-routing**
3. **adjacency-sid**
4. **interface** *type interface-path-id*
5. **address-family** { **ipv4** | **ipv6** } [**unicast**]
6. **l2-adjacency sid** {**index** *adj-SID-index* | **absolute** *adj-SID-value* } [**next-hop** { *ipv4_address* | *ipv6_address* }]
7. Use the **commit** or **end** command.
8. **end**
9. **router isis** *instance-id*
10. **address-family** { **ipv4** | **ipv6** } [**unicast**]
11. **segment-routing bundle-member-adj-sid**

DETAILED STEPS

	Command or Action	Purpose
Step 1	**configure** **Example:** `RP/0/RP0/CPU0:router# configure`	Enters mode.
Step 2	**segment-routing** **Example:** `RP/0/RP0/CPU0:Router(config)# segment-routing`	Enters segment routing configuration mode.
Step 3	**adjacency-sid** **Example:** `RP/0/RP0/CPU0:Router(config-sr)# adjacency-sid`	Enters adjacency SID configuration mode.
Step 4	**interface** *type interface-path-id* **Example:** `RP/0/RP0/CPU0:Router(config-sr-adj)# interface GigabitEthernet0/0/0/3`	Specifies the interface and enters interface configuration mode.
Step 5	**address-family** { **ipv4** | **ipv6** } [**unicast**] **Example:** `RP/0/RP0/CPU0:Router(config-sr-adj-intf)# address-family ipv4 unicast`	Specifies the IPv4 or IPv6 address family, and enters router address family configuration mode.

	Command or Action	Purpose
Step 6	**l2-adjacency sid** {**index** *adj-SID-index* | **absolute** *adj-SID-value* } [**next-hop** { *ipv4_address* | *ipv6_address* }] **Example:** `RP/0/RP0/CPU0:Router(config-sr-adj-intf-af)# l2-adjacency sid absolute 15015 next-hop 10.1.1.4`	Configures the Adj-SID index or absolute value for the interface. Specify **index** *adj-SID-index* for each link to create an Ajd-SID based on the lower boundary of the SRLB + the index. Specify **absolute** *adj-SID-value* for each link to create a specific Ajd-SID within the SRLB. For point-to-point interfaces, you are not required to specify a next-hop. However, if you do specify the next-hop, the Layer 2 Adj-SID will be used only if the specified next-hop matches the neighbor address. For LAN interfaces, you must configure the next-hop IPv4 or IPv6 address. If you do not configure the next-hop, the Layer 2 Adj-SID will not be used for LAN interface.
Step 7	Use the **commit** or **end** command.	**commit** —Saves the configuration changes and remains within the configuration session. **end** —Prompts user to take one of these actions: • **Yes** — Saves configuration changes and exits the configuration session. • **No** —Exits the configuration session without committing the configuration changes. • **Cancel** —Remains in the configuration session, without committing the configuration changes.
Step 8	**end**	
Step 9	**router isis** *instance-id* **Example:** `RP/0/RP0/CPU0:Router(config)# router isis isp`	Enables IS-IS routing for the specified routing instance, and places the router in router configuration mode.
Step 10	**address-family** { **ipv4** | **ipv6** } [**unicast**] **Example:** `RP/0/RP0/CPU0:Router(config-isis)# address-family ipv4 unicast`	Specifies the IPv4 or IPv6 address family, and enters router address family configuration mode.
Step 11	**segment-routing bundle-member-adj-sid** **Example:**	Programs the dynamic Layer 2 Adj-SIDs, and advertises both manual and dynamic Layer 2 Adj-SIDs.

	Command or Action	Purpose
	`RP/0/RP0/CPU0:Router(config-isis-af)#` **`segment-routing bundle-member-adj-sid`**	**Note** This command is not required to program manual L2 Adj-SID, but is required to program the dynamic Layer 2 Adj-SIDs and to advertise both manual and dynamic Layer 2 Adj-SIDs.

Verify the configuration:

```
Router# show mpls forwarding detail | i "Pop|Outgoing Interface|Physical Interface"
Tue Jun 20 06:53:51.876 PDT
. . .
15001  Pop         SRLB (idx 1)       BE1          10.1.1.4        0
     Outgoing Interface: Bundle-Ether1 (ifhandle 0x000000b0)
     Physical Interface: GigabitEthernet0/0/0/3 (ifhandle 0x000000b0)


Router# show running-config segment-routing
Tue Jun 20 07:14:25.815 PDT
segment-routing
 adjacency-sid
  interface GigabitEthernet0/0/0/3
   address-family ipv4 unicast
    l2-adjacency-sid absolute 15001
   !
  !
 !
!
```

Associated Commands

- l2-adjacency sid
- segment-routing bundle-member-adj-sid

IS-IS Prefix Attributes for Extended IPv4 and IPv6 Reachability

The following sub-TLVs support the advertisement of IPv4 and IPv6 prefix attribute flags and the source router ID of the router that originated a prefix advertisement, as described in RFC 7794.

- Prefix Attribute Flags
- IPv4 and IPv6 Source Router ID

Prefix Attribute Flags

The Prefix Attribute Flag sub-TLV supports the advertisement of attribute flags associated with prefix advertisements. Knowing if an advertised prefix is directly connected to the advertising router helps to determine how labels that are associated with an incoming packet should be processed.

This section describes the behavior of each flag when a prefix advertisement is learned from one level to another.

Note Prefix attributes are only added when wide metric is used.

Prefix Attribute Flags Sub-TLV Format

```
 0 1 2 3 4 5 6 7 ...
+-+-+-+-+-+-+-+-+...
|X|R|N|           ...
+-+-+-+-+-+-+-+-+...
```

Prefix Attribute Flags Sub-TLV Fields

Field	Description
X (External Prefix Flag)	This flag is set if the prefix has been redistributed from another protocol. The value of the flag is preserved when the prefix is propagated to another level.
R (Re-advertisement Flag)	This flag is set to 1 by the Level 1-2 router when the prefix is propagated between IS-IS levels (from Level 1 to Level 2, or from Level 2 to Level 1). This flag is set to 0 when the prefix is connected locally to an IS-IS-enabled interface (regardless of the level configured on the interface).
N (Node Flag)	For prefixes that are propagated from another level: 1. Copy the N-flag from the prefix attribute sub-TLV, if present in the source level. 2. Copy the N-flag from the prefix-SID sub-TLV, if present in the source level. 3. Otherwise, set to 0. For connected prefixes: 1. Set to 0 if **prefix-attributes n-flag-clear** is configured (see Configuring Prefix Attribute N-flag-clear, on page 37). 2. Set to 0 if **prefix-sid** {**index***SID-index* | **absolute** *SID-value*} {**n-flag-clear**] is configured (see Configuring a Prefix-SID on the IS-IS Enabled Loopback Interface, on page 18). 3. Otherwise, set to 1 when the prefix is a host prefix (/32 for IPV4, /128 for IPv6) that is associated with a loopback address. **Note** If the flag is set and the prefix length is not a host prefix, then the flag must be ignored.

IPv4 and IPv6 Source Router ID

The Source Router ID sub-TLV identifies the source of the prefix advertisement. The IPv4 and IPv6 source router ID is displayed in the output of the **show isis database verbose** command.

The Source Router ID sub-TLV is added when the following conditions are met:

1. The prefix is locally connected.
2. The N-flag is set to 1 (when it's a host prefix and the **n-flag-clear** configuration is not used).
3. The router ID is configured in the corresponding address family.

The source router ID is propagated between levels.

Table 4: Source Router Sub-TLV Format

IPv4 Source Router ID	Type: 11 Length: 4 Value: IPv4 Router ID of the source of the prefix advertisement
IPv6 Source Router ID	Type: 12 Length: 16 Value: IPv6 Router ID of the source of the prefix advertisement

Configuring Prefix Attribute N-flag-clear

The N-flag is set to 1 when the prefix is a host prefix (/32 for IPV4, /128 for IPv6) that is associated with a loopback address. The advertising router can be configured to not set this flag. This task explains how to clear the N-flag.

SUMMARY STEPS

1. **configure**
2. **router isis** *instance-id*
3. **interface Loopback** *instance*
4. **prefix-attributes n-flag-clear** [**Level-1** | **Level-2**]
5. Use the **commit** or **end** command.

DETAILED STEPS

	Command or Action	Purpose
Step 1	**configure** **Example:** `RP/0/RP0/CPU0:router# configure`	Enters mode.
Step 2	**router isis** *instance-id* **Example:** `RP/0/RP0/CPU0:router(config)# router isis 1`	

	Command or Action	Purpose
Step 3	**interface Loopback** *instance* **Example:** `RP/0/RP0/CPU0:router(config)# interface Loopback0`	Specifies the loopback interface.
Step 4	**prefix-attributes n-flag-clear** [**Level-1** | **Level-2**] **Example:** `RP/0/RP0/CPU0:router(config-if)# isis prefix-attributes n-flag-clear`	Clears the prefix attribute N-flag explicitly.
Step 5	Use the **commit** or **end** command.	**commit** —Saves the configuration changes and remains within the configuration session. **end** —Prompts user to take one of these actions: • **Yes** — Saves configuration changes and exits the configuration session. • **No** —Exits the configuration session without committing the configuration changes. • **Cancel** —Remains in the configuration session, without committing the configuration changes.

Verify the prefix attribute configuration:

```
RP/0/RP0/CPU0:router# show isis database verbose

IS-IS 1 (Level-2) Link State Database
LSPID                 LSP Seq Num  LSP Checksum  LSP Holdtime  ATT/P/OL
router.00-00        * 0x0000039b   0xfc27        1079            0/0/0
  Area Address: 49.0001
  NLPID:        0xcc
  NLPID:        0x8e
  MT:           Standard (IPv4 Unicast)
  MT:           IPv6 Unicast                                     0/0/0
  Hostname:     router
  IP Address:   10.0.0.1
  IPv6 Address: 2001:0db8:1234::0a00:0001
  Router Cap:   10.0.0.1, D:0, S:0
    Segment Routing: I:1 V:1, SRGB Base: 16000 Range: 8000
    SR Algorithm:
      Algorithm: 0
      Algorithm: 1
<...>
  Metric: 0          IP-Extended 10.0.0.1/32
    Prefix-SID Index: 1001, Algorithm:0, R:1 N:0 P:1 E:0 V:0 L:0
    Prefix Attribute Flags: X:0 R:1 N:0
  Metric: 10         IP-Extended 10.0.0.2/32
    Prefix-SID Index: 1002, Algorithm:0, R:0 N:1 P:0 E:0 V:0 L:0
    Prefix Attribute Flags: X:0 R:0 N:1
    Source Router ID: 10.0.0.2
<...>
```

Conditional Prefix Advertisement

In some situations, it's beneficial to make the IS-IS prefix advertisement conditional. For example, an Area Border Router (ABR) or Autonomous System Boundary Router (ASBR) that has lost its connection to one of the areas or autonomous systems (AS) might keep advertising a prefix. If an ABR or ASBR advertises the Segment Routing (SR) SID with this prefix, the label stack of the traffic routed toward the disconnected area or AS might use this SID, which would result in dropped traffic at the ABR or ASBR.

ABRs or ASBRs are often deployed in pairs for redundancy and advertise a shared Anycast prefix SID. Conditional Prefix Advertisement allows an ABR or an ASBR to advertise its Anycast SID only when connected to a specific area or domain. If an ABR or ASBR becomes disconnected from the particular area or AS, it stops advertising the address for a specified interface (for example, Loopback).

Configure the conditional prefix advertisement under a specific interface. The prefix advertisement on this interface is associated with the route-policy that tracks the presence of a set of prefixes (prefix-set) in the Routing Information Base (RIB).

For faster convergence, the route-policy used for conditional prefix advertisement uses the new event-based **rib-has-route async** condition to notify IS-IS of the following situations:

- When the last prefix from the prefix-set is removed from the RIB.
- When the first prefix from the prefix-set is added to the RIB.

Configuration

To use the conditional prefix advertisement in IS-IS, create a prefix-set to be tracked. Then create a route policy that uses the prefix-set.

```
Router(config)# prefix-set prefix-set-name
Router(config-pfx)# prefix-address-1/length[, prefix-address-2/length,,,
prefix-address-16/length]
Router(config-pfx)# end-set

Router(config)# route-policy rpl-name
Router(config-rpl)# if rib-has-route async prefix-set-name then
Router(config-rpl-if)# pass
Router(config-rpl-if)# endif
Router(config-rpl)# end-policy
```

To advertise the loopback address in IS-IS conditionally, use the **advertise prefix route-policy** command under IS-IS interface address-family configuration sub-mode.

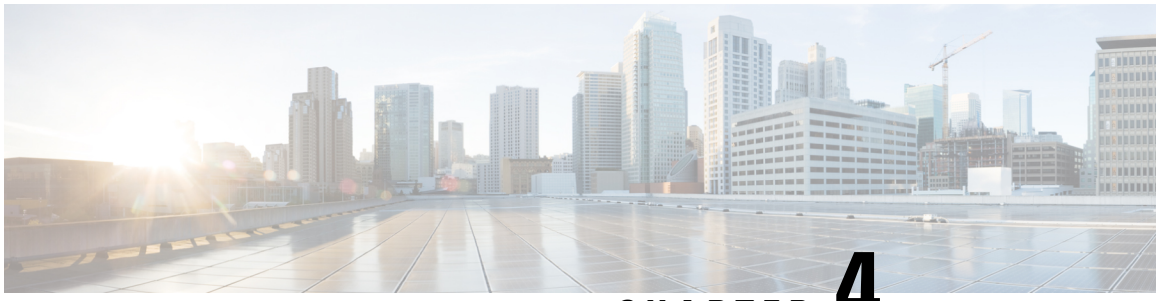
```
Router(config)# router isis 1
Router(config-isis)# interface Loopback0
Router(config-isis-if)# address-family ipv4 unicast
Router(config-isis-if-af)# advertise prefix route-policy rpl-name
Router(config-isis-if-af)# commit
```

Example

```
Router(config)# prefix-set domain_2
Router(config-pfx)# 2.3.3.3/32, 2.4.4.4/32
Router(config-pfx)# end-set
Router(config)# route-policy track_domain_2
Router(config-rpl)# if rib-has-route async domain_2 then
Router(config-rpl-if)# pass
Router(config-rpl-if)# endif
Router(config-rpl)# end-policy
Router(config)# router isis 1
Router(config-isis)# interface Loopback0
Router(config-isis-if)# address-family ipv4 unicast
Router(config-isis-if-af)# advertise prefix route-policy track_domain-2
Router(config-isis-if-af)# commit
```

Running Configuration

```
prefix-set domain_2
  2.3.3.3/32,
  2.4.4.4/32
end-set
!
route-policy track_domain_2
  if rib-has-route async domain_2 then
    pass
  endif
end-policy
!
router isis 1
 interface Loopback0
  address-family ipv4 unicast
   advertise prefix route-policy track_domain_2
  !
 !
!
```



CHAPTER 4

Configure Segment Routing for OSPF Protocol

Open Shortest Path First (OSPF) is an Interior Gateway Protocol (IGP) developed by the OSPF working group of the Internet Engineering Task Force (IETF). Designed expressly for IP networks, OSPF supports IP subnetting and tagging of externally derived routing information. OSPF also allows packet authentication and uses IP multicast when sending and receiving packets.

This module provides the configuration information to enable segment routing for OSPF.

Enabling Segment Routing for OSPF Protocol

Segment routing on the OSPF control plane supports the following:

- OSPFv2 control plane
- Multi-area
- IPv4 prefix SIDs for host prefixes on loopback interfaces
- Adjacency SIDs for adjacencies
- MPLS penultimate hop popping (PHP) and explicit-null signaling

This section describes how to enable segment routing MPLS and MPLS forwarding in OSPF. Segment routing can be configured at the instance, area, or interface level.

Before you begin

Your network must support the MPLS Cisco IOS XR software feature before you enable segment routing for OSPF on your router.

Note You must enter the commands in the following task list on every OSPF router in the traffic-engineered portion of your network.

SUMMARY STEPS

1. **configure**
2. **router ospf** *process-name*
3. **segment-routing mpls**
4. **segment-routing sr-prefer**
5. **area 0**
6. **segment-routing mpls**
7. **exit**
8. Use the **commit** or **end** command.

DETAILED STEPS

	Command or Action	Purpose
Step 1	**configure** **Example:** `RP/0/RP0/CPU0:router# configure`	Enters mode.
Step 2	**router ospf** *process-name* **Example:** `RP/0/RP0/CPU0:router(config)# router ospf 1`	Enables OSPF routing for the specified routing process and places the router in router configuration mode.
Step 3	**segment-routing mpls** **Example:** `RP/0/RP0/CPU0:router(config-ospf)# segment-routing mpls`	Enables segment routing using the MPLS data plane on the routing process and all areas and interfaces in the routing process. Enables segment routing fowarding on all interfaces in the routing process and installs the SIDs received by OSPF in the forwarding table.
Step 4	**segment-routing sr-prefer** **Example:** `RP/0/RP0/CPU0:router(config-ospf)# segment-routing sr-prefer`	Sets the preference of segment routing (SR) labels over label distribution protocol (LDP) labels.
Step 5	**area 0** **Example:** `RP/0/RP0/CPU0:router(config-ospf)# area 0`	Enters area configuration mode.
Step 6	**segment-routing mpls** **Example:** `RP/0/RP0/CPU0:router(config-ospf-ar)# segment-routing mpls`	(Optional) Enables segment routing using the MPLS data plane on the area and all interfaces in the area. Enables segment routing fowarding on all interfaces in the area and installs the SIDs received by OSPF in the forwarding table.
Step 7	**exit** **Example:**	

	Command or Action	Purpose
	`RP/0/RP0/CPU0:router(config-ospf-ar)# exit` `RP/0/RP0/CPU0:router(config-ospf)# exit`	
Step 8	Use the **commit** or **end** command.	**commit** —Saves the configuration changes and remains within the configuration session. **end** —Prompts user to take one of these actions: • **Yes** — Saves configuration changes and exits the configuration session. • **No** —Exits the configuration session without committing the configuration changes. • **Cancel** —Remains in the configuration session, without committing the configuration changes.

What to do next

Configure the prefix SID.

Configuring a Prefix-SID on the OSPF-Enabled Loopback Interface

A prefix segment identifier (SID) is associated with an IP prefix. The prefix SID is manually configured from the segment routing global block (SRGB) range of labels. A prefix SID is configured under the loopback interface with the loopback address of the node as the prefix. The prefix segment steers the traffic along the shortest path to its destination.

A prefix SID can be a node SID or an Anycast SID. A node SID is a type of prefix SID that identifies a specific node. An Anycast SID is a type of prefix SID that identifies a set of nodes, and is configured with n-flag clear. The set of nodes (Anycast group) is configured to advertise a shared prefix address and prefix SID. Anycast routing enables the steering of traffic toward multiple advertising nodes. Packets addressed to an Anycast address are forwarded to the topologically nearest nodes.

The prefix SID is globally unique within the segment routing domain.

This task describes how to configure prefix segment identifier (SID) index or absolute value on the OSPF-enabled Loopback interface.

Before you begin

Ensure that segment routing is enabled on an instance, area, or interface.

SUMMARY STEPS

1. **configure**
2. **router ospf** *process-name*
3. **area** *value*

4. **interface Loopback** *interface-instance*
5. **prefix-sid**{**index** *SID-index* | **absolute** *SID-value* } [**n-flag-clear**] [**explicit-null**]
6. Use the **commit** or **end** command.

DETAILED STEPS

	Command or Action	Purpose
Step 1	**configure** **Example:** `RP/0/RP0/CPU0:router# configure`	Enters mode.
Step 2	**router ospf** *process-name* **Example:** `RP/0/RP0/CPU0:router(config)# router ospf 1`	Enables OSPF routing for the specified routing process, and places the router in router configuration mode.
Step 3	**area** *value* **Example:** `RP/0/RP0/CPU0:router(config-ospf)# area 0`	Enters area configuration mode.
Step 4	**interface Loopback** *interface-instance* **Example:** `RP/0/RP0/CPU0:router(config-ospf-ar)# interface Loopback0 passive`	Specifies the loopback interface and instance.

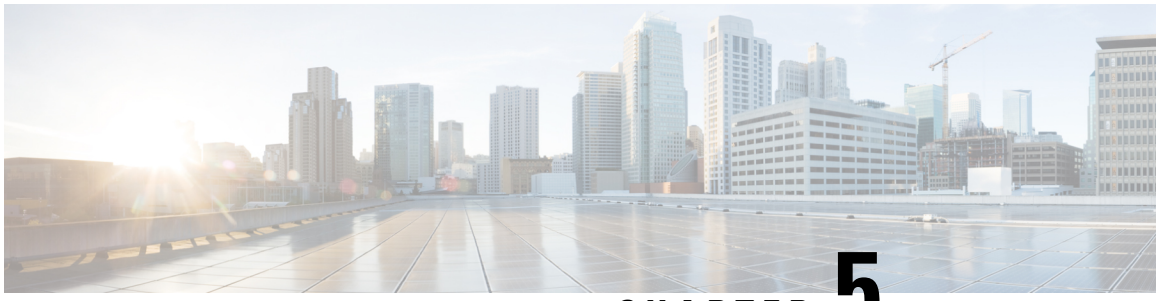
Step 5	**prefix-sid**{**index** *SID-index* | **absolute** *SID-value* } [**n-flag-clear**] [**explicit-null**]
	Example:
	`RP/0/RP0/CPU0:router(config-ospf-ar)# prefix-sid index 1001`
	`RP/0/RP0/CPU0:router(config-ospf-ar)# prefix-sid absolute 17001`
		Configures the prefix-SID index or absolute value for the interface.
		Specify **index** *SID-index* for each node to create a prefix SID based on the lower boundary of the SRGB + the index.
		Specify **absolute** *SID-value* for each node to create a specific prefix SID within the SRGB.
		By default, the n-flag is set on the prefix-SID, indicating that it is a node SID. For specific prefix-SID (for example, Anycast prefix-SID), enter the `n-flag-clear` keyword. OSPF does not set the `N` flag in the prefix-SID sub Type Length Value (TLV).
		To disable penultimate-hop-popping (PHP) and add an explicit-Null label, enter the `explicit-null` keyword. OSPF sets the `E` flag in the prefix-SID sub TLV.

Step 6	Use the **commit** or **end** command.	**commit** —Saves the configuration changes and remains within the configuration session. **end** —Prompts user to take one of these actions:

	Command or Action	Purpose
		• **Yes** — Saves configuration changes and exits the configuration session. • **No** —Exits the configuration session without committing the configuration changes. • **Cancel** —Remains in the configuration session, without committing the configuration changes.

Verify the prefix-SID configuration:

```
RP/0/RP0/CPU0:router# show ospf database opaque-area 7.0.0.1 self-originate
 OSPF Router with ID (10.0.0.1) (Process ID 1)
               Type-10 Opaque Link Area Link States (Area 0)
<...>
    Extended Prefix TLV: Length: 20
      Route-type: 1
      AF        : 0
      Flags     : 0x40
      Prefix    : 10.0.0.1/32

      SID sub-TLV: Length: 8
        Flags     : 0x0
        MTID      : 0
        Algo      : 0
        SID Index : 1001
```



CHAPTER 5

Configure Segment Routing for BGP

Border Gateway Protocol (BGP) is an Exterior Gateway Protocol (EGP) that allows you to create loop-free inter-domain routing between autonomous systems. An autonomous system is a set of routers under a single technical administration. Routers in an autonomous system can use multiple Interior Gateway Protocols (IGPs) to exchange routing information inside the autonomous system and an EGP to route packets outside the autonomous system.

This module provides the configuration information used to enable Segment Routing for BGP.

Note For additional information on implementing BGP on your router, see the *BGP Configuration Guide for Cisco 8000 Series Routers*.

Segment Routing for BGP

In a traditional BGP-based data center (DC) fabric, packets are forwarded hop-by-hop to each node in the autonomous system. Traffic is directed only along the external BGP (eBGP) multipath ECMP. No traffic engineering is possible.

In an MPLS-based DC fabric, the eBGP sessions between the nodes exchange BGP labeled unicast (BGP-LU) network layer reachability information (NLRI). An MPLS-based DC fabric allows any leaf (top-of-rack or border router) in the fabric to communicate with any other leaf using a single label, which results in higher packet forwarding performance and lower encapsulation overhead than traditional BGP-based DC fabric. However, since each label value might be different for each hop, an MPLS-based DC fabric is more difficult to troubleshoot and more complex to configure.

BGP has been extended to carry segment routing prefix-SID index. BGP-LU helps each node learn BGP prefix SIDs of other leaf nodes and can use ECMP between source and destination. Segment routing for BGP simplifies the configuration, operation, and troubleshooting of the fabric. With segment routing for BGP, you can enable traffic steering capabilities in the data center using a BGP prefix SID.

Configure BGP Prefix Segment Identifiers

Segments associated with a BGP prefix are known as BGP prefix SIDs. The BGP prefix SID is global within a segment routing or BGP domain. It identifies an instruction to forward the packet over the ECMP-aware best-path computed by BGP to the related prefix. The BGP prefix SID is manually configured from the segment routing global block (SRGB) range of labels.

Each BGP speaker must be configured with an SRGB using the **segment-routing global-block** command. See the About the Segment Routing Global Block, on page 5 section for information about the SRGB.

Note You must enable SR and explicitly configure the SRGB before configuring SR BGP. The SRGB must be explicitly configured, even if you are using the default range (16000 – 23999). BGP uses the SRGB and the index in the BGP prefix-SID attribute of a learned BGP-LU advertisement to allocate a local label for a given destination.

If SR and the SRGB are enabled after configuring BGP, then BGP is not aware of the SRGB, and therefore it allocates BGP-LU local labels from the dynamic label range instead of from the SRGB. In this case, restart the BGP process in order to allocate BGP-LU local labels from the SRGB.

Note Because the values assigned from the range have domain-wide significance, we recommend that all routers within the domain be configured with the same range of values.

To assign a BGP prefix SID, first create a routing policy using the **set label-index** *index* attribute, then associate the index to the node.

Note A routing policy with the **set label-index** attribute can be attached to a network configuration or redistribute configuration. Other routing policy language (RPL) configurations are possible. For more information on routing policies, refer to the "Implementing Routing Policy" chapter in the *Routing Configuration Guide for Cisco 8000 Series Routers*.

Example

The following example shows how to configure the SRGB, create a BGP route policy using a $SID parameter and **set label-index** attribute, and then associate the prefix-SID index to the node.

```
RP/0/RP0/CPU0:router(config)# segment-routing global-block 16000 23999

RP/0/RP0/CPU0:router(config)# route-policy SID($SID)
RP/0/RP0/CPU0:router(config-rpl)# set label-index $SID
RP/0/RP0/CPU0:router(config-rpl)# end policy

RP/0/RP0/CPU0:router(config)# router bgp 1
RP/0/RP0/CPU0:router(config-bgp)# bgp router-id 1.1.1.1
RP/0/RP0/CPU0:router(config-bgp)# address-family ipv4 unicast
RP/0/RP0/CPU0:router(config-bgp-af)# network 1.1.1.3/32 route-policy SID(3)
RP/0/RP0/CPU0:router(config-bgp-af)# allocate-label all
```

```
RP/0/RP0/CPU0:router(config-bgp-af)# commit
RP/0/RP0/CPU0:router(config-bgp-af)# end


RP/0/RP0/CPU0:router# show bgp 1.1.1.3/32
BGP routing table entry for 1.1.1.3/32
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker                74          74
    Local Label: 16003
Last Modified: Sep 29 19:52:18.155 for 00:07:22
Paths: (1 available, best #1)
  Advertised to update-groups (with more than one peer):
    0.2
  Path #1: Received by speaker 0
  Advertised to update-groups (with more than one peer):
    0.2
  3
    99.3.21.3 from 99.3.21.3 (1.1.1.3)
      Received Label 3
      Origin IGP, metric 0, localpref 100, valid, external, best, group-best
      Received Path ID 0, Local Path ID 1, version 74
      Origin-AS validity: not-found
      Label Index: 3
```

Segment Routing Egress Peer Engineering

Table 5: Feature History Table

Feature Name	Release Information	Feature Description
BGP PeerSet SID	Release 7.3.2	BGP peer SIDs are used to express source-routed interdomain paths and are of two types: Peer Node SIDs and Peer Adjacency SIDs. This release supports a new type of BGP peering SID, called BGP Peer Set SID. It is a group or set of BGP peer SIDs, that can provide load balancing over BGP neighbors (nodes) or links (adjacencies). The BGP peer Set SID can be associated with any combination of Peer Node SIDs or Peer Adjacency SIDs.

Segment routing egress peer engineering (EPE) uses a controller to instruct an ingress provider edge, or a content source (node) within the segment routing domain, to use a specific egress provider edge (node) and a specific external interface to reach a destination. BGP peer SIDs are used to express source-routed inter-domain paths.

Below are the BGP-EPE peering SID types:

- PeerNode SID—To an eBGP peer. Pops the label and forwards the traffic on any interface to the peer.

- PeerAdjacency SID—To an eBGP peer via interface. Pops the label and forwards the traffic on the related interface.
- PeerSet SID—To a set of eBGP peers. Pops the label and forwards the traffic on any interface to the set of peers. All the peers in a set might not be in the same AS.

  Multiple PeerSet SIDs can be associated with any combination of PeerNode SIDs or PeerAdjacency SIDs.

The controller learns the BGP peer SIDs and the external topology of the egress border router through BGP-LS EPE routes. The controller can program an ingress node to steer traffic to a destination through the egress node and peer node using BGP labeled unicast (BGP-LU).

EPE functionality is only required at the EPE egress border router and the EPE controller.

Usage Guidelines and Limitations

- When enabling BGP EPE, you must enable MPLS encapsulation on the egress interface connecting to the eBGP peer. This can be done by enabling either BGP labeled unicast (BGP-LU) address family or MPLS static for the eBGP peer.

  For information about BGP-LU, refer to the Implementing BGP chapter in the *BGP Configuration Guide for Cisco 8000 Series Routers*.

  For information about MPLS static, refer to the Implementing MPLS Static Labeling chapter in the *MPLS Configuration Guide for Cisco 8000 Series Routers*.

- Note the following points related to the IP-lookup backup support for EPEs:
  - This feature works only when you enable the **epe backup enable**, under the Global Address Family ID (AFI).
  - With this feature, an IP-Lookup backup is installed for each Egress Peer Engineering. This means, when all the paths of that EPE go down, the Forwarding Information Base (FIB) table searches in the IP table for the destination IP address in the data packet and forwards them accordingly.
  - The peer-set EPEs have a backup installed only when the mentioned CLI knob is enabled.

Configure Segment Routing Egress Peer Engineering

This task explains how to configure segment routing EPE on the EPE egress node.

SUMMARY STEPS

1. **router bgp** *as-number*
2. **neighbor** *ip-address*
3. **remote-as** *as-number*
4. **egress-engineering**
5. **exit**
6. **mpls static**
7. **interface** *type interface-path-id*

DETAILED STEPS

	Command or Action	Purpose
Step 1	**router bgp** *as-number* **Example:** `RP/0/RP0/CPU0:router(config)# router bgp 1`	Specifies the BGP AS number and enters the BGP configuration mode, allowing you to configure the BGP routing process.
Step 2	**neighbor** *ip-address* **Example:** `RP/0/RP0/CPU0:router(config-bgp)# neighbor 10.10.10.2`	Places the router in neighbor configuration mode for BGP routing and configures the neighbor IP address as a BGP peer.
Step 3	**remote-as** *as-number* **Example:** `RP/0/RP0/CPU0:router(config-bgp-nbr)# remote-as 3`	Creates a neighbor and assigns a remote autonomous system number to it.
Step 4	**egress-engineering** **Example:** `RP/0/RP0/CPU0:router(config-bgp-nbr)# egress-engineering`	Configures the egress node with EPE for the eBGP peer.

Step 5	**exit**
	Example:
	`RP/0/RP0/CPU0:router(config-bgp-nbr)# exit`
	`RP/0/RP0/CPU0:router(config-bgp)# exit`
	`RP/0/RP0/CPU0:router(config)#`	

Step 6	**mpls static** **Example:** `RP/0/RP0/CPU0:router(config)# mpls static`	Configure MPLS static on the egress interface connecting to the eBGP peer.
Step 7	**interface** *type interface-path-id* **Example:** `RP/0/RP0/CPU0:router(config-mpls-static)# interface GigabitEthernet0/0/1/2`	Specifies the egress interface connecting to the eBGP peer.

```
router bgp 1
 neighbor 10.10.10.2
  remote-as 3
  egress-engineering
 !
!
mpls static
 interface GigabitEthernet0/0/1/2
!
```

Configuring Manual BGP-EPE Peering SIDs

Table 6: Feature History Table

Feature Name	Release Information	Feature Description
Manual BGP-EPE Peer SIDs	Release 7.3.2	BGP Peering SIDs that are allocated dynamically are not persistent and can be reallocated after a reload or a process restart. This feature allows you to manually configure BGP Egress Peer Engineering (EPE) Peering SIDs. This functionality provides predictability, consistency, and reliability if there are system reloads or process restarts.

Configuring manual BGP-EPE Peer SIDs allows for persistent EPE label values. Manual BGP-EPE SIDs are advertised through BGP-LS and are allocated from the Segment Routing Local Block (SRLB). See Configure Segment Routing Global Block and Segment Routing Local Block, on page 5 for information about the SRLB.

Each PeerNode SID, PeerAdjacency SID, and PeerSet SID is configured with an index value. This index serves as an offset from the configured SRLB start value and the resulting MPLS label (SRLB start label + index) is assigned to these SIDs. This label is used by CEF to perform load balancing across the individual BGP PeerSet SIDs, BGP PeerNode SID, or ultimately across each first-hop adjacency associated with that BGP PeerNode SID or BGP PeerSet SID.

Configuring Manual PeerNode SID

Each eBGP peer will be associated with a PeerNode SID index that is configuration driven.

```
RP/0/0/CPU0:PE1(config)# router bgp 10
RP/0/0/CPU0:PE1(config-bgp)# neighbor 10.10.10.2
RP/0/0/CPU0:PE1(config-bgp-nbr)# remote-as 20
RP/0/0/CPU0:PE1(config-bgp-nbr)# egress-engineering
RP/0/0/CPU0:PE1(config-bgp-nbr)# peer-node-sid index 600
```

Configuring Manual PeerAdjacency SID

Any first-hop for which an adjacency SID is configured needs to be in the resolution chain of at least one eBGP peer that is configured for egress-peer engineering. Otherwise such a kind of “orphan” first-hop with

regards to BGP has no effect on this feature. This is because BGP only understands next-hops learnt by the BGP protocol itself and in addition only the resolving IGP next-hops for those BGP next-hops.

```
RP/0/0/CPU0:PE1(config)# router bgp 10
RP/0/0/CPU0:PE1(config-bgp)# adjacencies
RP/0/0/CPU0:PE1(config-bgp-adj)# 1.1.1.2
RP/0/0/CPU0:PE1(config-bgp-adj)# adjacency-sid index 500
```

Configuring Manual PeerSet SID

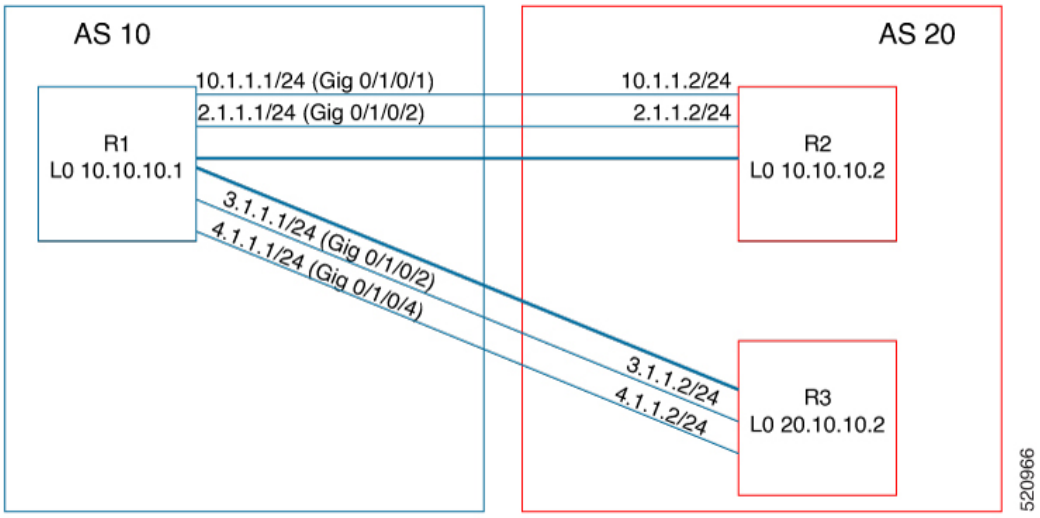
The PeerSet SID is configured under global Address Family. This configuration results in the creation of a Peer-Set SID EPE object.

```
RP/0/0/CPU0:PE1(config)# router bgp 10
RP/0/0/CPU0:PE1(config-bgp)# address-family ipv4 unicast
RP/0/0/CPU0:PE1(config-bgp-afi)# peer-set-id 1
RP/0/0/CPU0:PE1(config-bgp-peer-set)# peer-set-sid 300
```

Example

Topology

The example in this section uses the following topology.



In this example, BGP-EPE peer SIDs are allocated from the default SRLB label range (15000 – 15999). The BGP-EPE peer SIDs are configured as follows:

- PeerNode SIDs to 10.10.10.2 with index 600 (label 15600), and for 20.10.10.2 with index 700 (label 15700)
- PeerAdj SID to link 1.1.1.2 with index 500 (label 15500)
- PeerSet SID 1 to load balance over BGP neighbors 10.10.10.1 and 20.10.10.2 with SID index 300 (label 15300)
- PeerSet SID 2 to load balance over BGP neighbor 20.10.10.2 and link 1.1.1.2 with SID index 400 (label 15400)

Configuration on R1

```
router bgp 10
 address-family ipv4 unicast
  peer-set-id 1
   peer-set-sid index 300
  !
  peer-set-id 2
   peer-set-sid index 400
  !
 !
 adjacencies
  1.1.1.2
   adjacency-sid index 500
   peer-set 2
  !
 !
 neighbor 10.10.10.2
  remote-as 20
  egress-engineering
  peer-node-sid index 600
  peer-set 1
 !
 neighbor 20.10.10.2
  egress-engineering
  peer-node-sid index 700
  peer-set 1
  peer-set 2
 !
```

To further show the load balancing of this example:

- 15600 is load balanced over {1.1.1.1 and 2.1.1.1}
- 15700 is load balanced over {3.1.1.1 and 4.1.1.1}
- 15500 is load balanced over {1.1.1.1}
- 15300 is load balanced over {1.1.1.1, 2.1.1.1, 3.1.1.1 and 4.1.1.1}
- 15400 is load balanced over {1.1.1.1, 3.1.1.1 and 4.1.1.1}

Advertising EPE-Enabled BGP Neighbors via BGP-LU

Table 7: Feature History Table

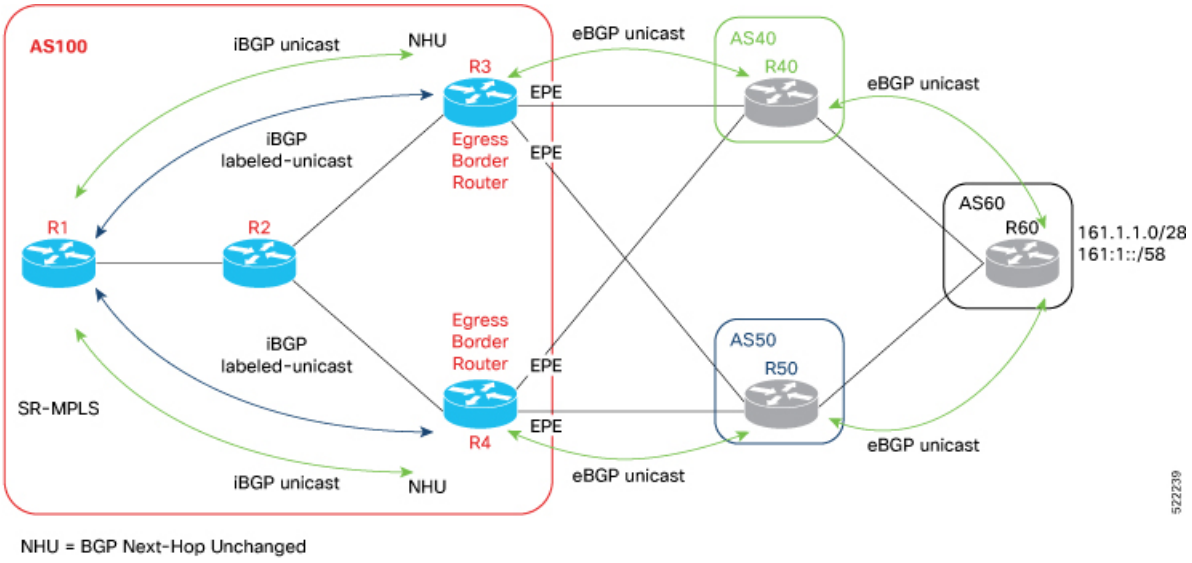
Feature Name	Release	Description
Advertising EPE-Enabled BGP Neighbors via BGP-LU	Release 7.3.3	BGP peering segments/SIDs are part of the Segment Routing Centralized BGP Egress Peer Engineering solution (BGP-EPE). A BGP-EPE-enabled border router allocates and programs BGP peering SIDs (EPE labels) to steer traffic over a specific external interface/BGP neighbor to reach a particular destination. This feature provides an alternate BGP-EPE solution leveraging BGP peering segments. It allows a BGP-EPE-enabled border router to use BGP Labeled Unicast (BGP-LU) to advertise the IP address of a neighbor with an LU label equal to the EPE label assigned to that neighbor.

BGP peering segments/SIDs are part of the Segment Routing Centralized BGP Egress Peer Engineering solution (BGP-EPE), as described in IETF RFC 9087. A BGP-EPE-enabled border router allocates and programs BGP peering SIDs (EPE labels) to steer traffic over a specific external interface/BGP neighbor to reach a particular destination.

This feature provides an alternate BGP-EPE solution leveraging BGP peering segments. It allows a BGP-EPE-enabled border router to use BGP Labeled Unicast (BGP-LU) to advertise the IP address of a neighbor with an LU label equal to the EPE label assigned to that neighbor.

The following figure illustrates a Segment Routing network (AS100) connected to a pair of transit Autonomous Systems. The egress border routers (R3 and R4) have BGP Peering segments (EPE) enabled on their eBGP neighbors in AS40 and AS50. Prefixes are propagated inside AS100 via BGP. R3 and R4 maintain the BGP next-hops unchanged. In addition, BGP labeled unicast is enabled inside AS100 to advertise the IP address of these eBGP neighbors.

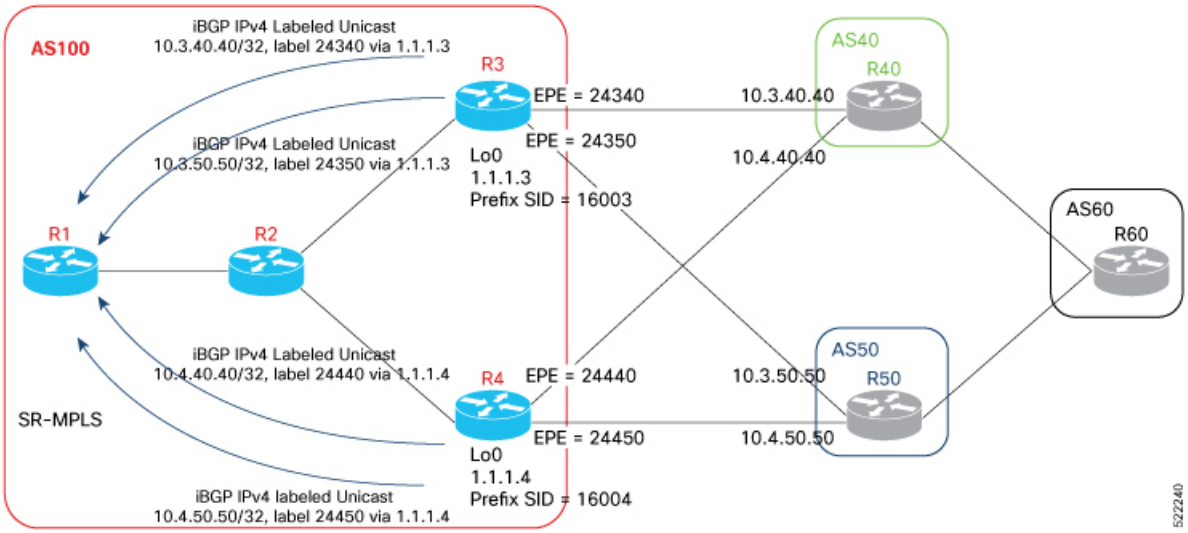
Figure 1: Solution Overview



NHU = BGP Next-Hop Unchanged

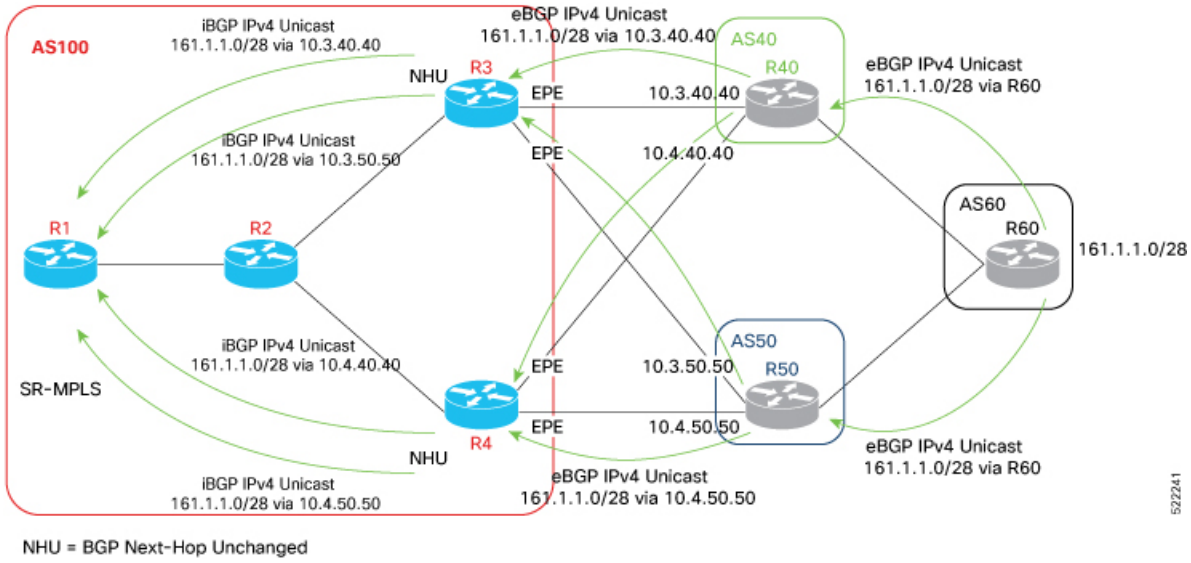
The figure below depicts the BGP-LU advertisements originated by R3 and R4 for the IP addresses of their eBGP neighbors. The figure also indicates the EPE label values assigned to each eBGP neighbor. Note that the local BGP-LU label on the egress border router is equal to the EPE label assigned to that neighbor.

Figure 2: Advertising EPE-Enabled BGP Neighbors via BGP-LU



In the following figure, an overlay prefix 161.1.1.0/28 originating at AS60 is advertised inside AS100. Egress border routers are configured to advertise all of their paths. Note that the BGP next-hops are not modified. In this example, the ingress router in AS100 (R1) learns the overlay prefix via 4 paths (one for each eBGP neighbor).

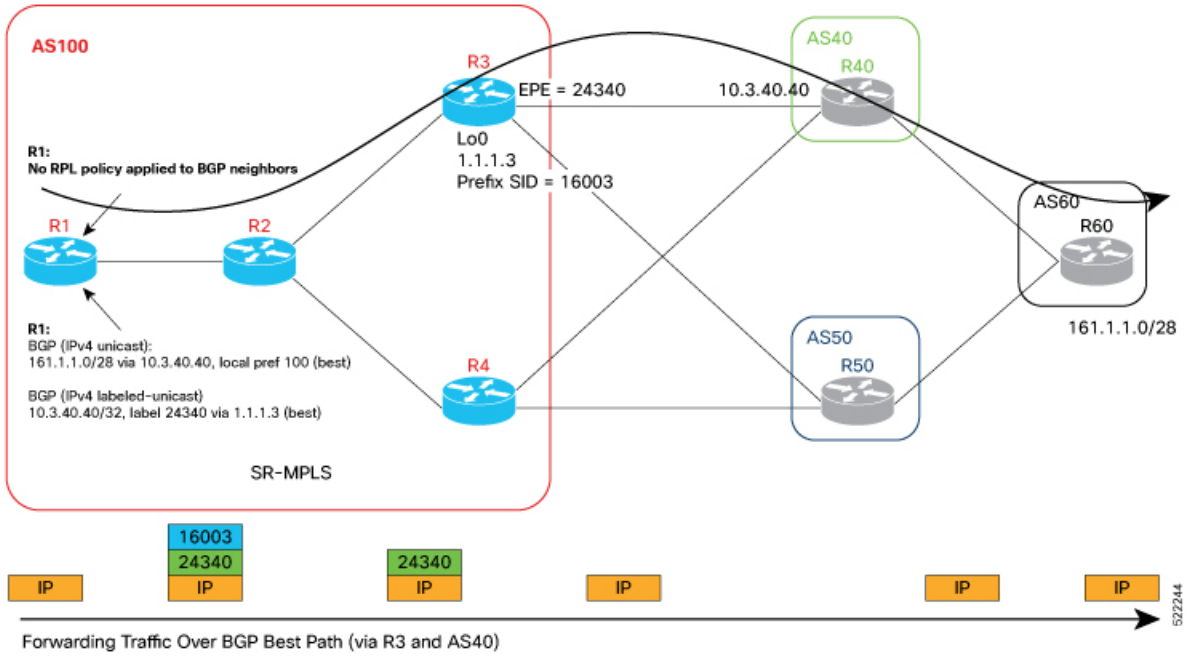
Figure 3: Advertising Overlay Prefixes



NHU = BGP Next-Hop Unchanged

On ingress border router R1, and without any BGP policy modification, assume that BGP selects the path corresponding to AS40 as best path for the overlay prefix 161.1.1.0/28. Its BGP next-hop (10.3.40.40) is learned via BGP-LU from egress border router R3 (1.1.1.3) and with LU label 24340. This label is the EPE label assigned at R3 for the eBGP neighbor to AS40. The EPE local label is programmed as a POP-and-forward toward the interface connecting to AS40. Lastly, R3's loopback (1.1.1.3) and its prefix label (16003) are learned via IS-IS with SR extensions. As a result, incoming traffic matching the 161.1.1.0/28 route is encapsulated at R1 with two MPLS labels (bottom-of-stack label **24340** and top label **16003**) in order to send the traffic to R3 and then to AS40.

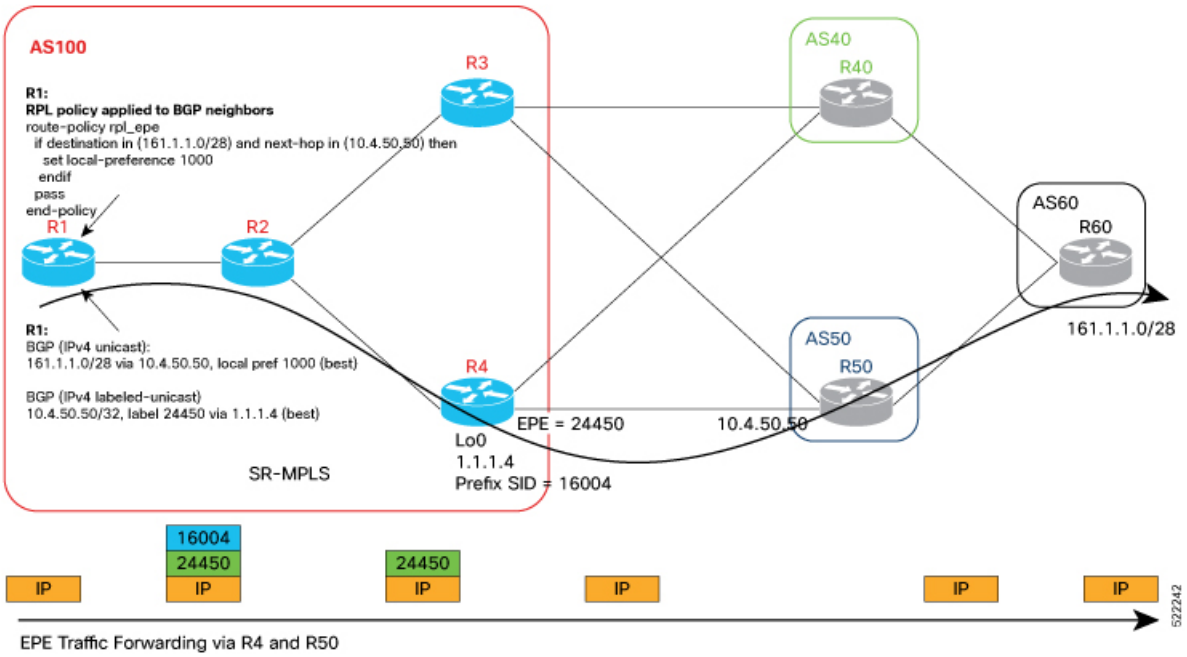
Figure 4: Forwarding Traffic Over BGP Best Path (via R3 and AS40)



Forwarding Traffic Over BGP Best Path (via R3 and AS40)

When the operator wants to modify the exit egress border router and/or an exit AS for a given overlay prefix, a BGP policy can be applied at the ingress border router to influence the best-path selection. In our example, consider that the desired egress path to 161.1.1.0/28 is via R4 and then AS50 (instead of R3 and AS40, as shown in the previous figure). An RPL policy, for example, can be used to assign a higher BGP local preference to the desired path. As a result, incoming traffic matching the 161.1.1.0/28 route is now encapsulated at R1 with two MPLS labels (bottom-of-stack label **24450** and top label **16004**) in order to send the traffic to R4 and then to AS50.

Figure 5: Forwarding Traffic Over EPE Path (via R4 and AS50)



Usage Guidelines and Limitations

The following usage guidelines and limitations apply for this feature:

- BGPv4 and BGPv6 EPE-enabled neighbors are supported.
- BGP peering SIDs (EPE Peer-Node SIDs and Peer-Adjacencies SIDs) allocated dynamically or configured manually can be used as BGP-LU labels when advertising the IP address of an EPE-enabled BGP neighbor via BGP-LU.
- BGP Peer-Set SIDs are not supported.

Enabling Advertisement of EPE-Enabled BGP Neighbors via BGP-LU

To enable advertisement of EPE-enabled BGP neighbors via BGP-LU, use the **advertise epe-bgp labeled-unicast** command in router BGP address family configuration mode.

The following example shows how to enable advertisement of EPE-enabled BGP neighbors via BGP-LU:

```
RP/0/RP0/CPU0:R3(config)# router bgp 100
RP/0/RP0/CPU0:R3(config-bgp)# address-family ipv4 unicast
RP/0/RP0/CPU0:R3(config-bgp-af)# advertise epe-bgp labeled-unicast
```

Running Config

```
router bgp 100
  address-family ipv4 unicast
   advertise epe-bgp labeled-unicast
```

Use Case:

This section provides the router configuration and show command outputs of the scenario described in the overview above

Egress Border Router R3 Configuration

Configure the SRGB:

```
RP/0/RP0/CPU0:R3(config)# segment-routing
RP/0/RP0/CPU0:R3(config-sr)# global-block 16000 23999
RP/0/RP0/CPU0:R3(config-sr)# exit
```

Configure the Loopback address:

```
RP/0/RP0/CPU0:R3(config)# interface Loopback0
RP/0/RP0/CPU0:R3(config-if)# ipv4 address 1.1.1.3 255.255.255.255
RP/0/RP0/CPU0:R3(config-if)# exit
```

Configure MPLS Static on the egress interface connecting to the eBGP peer:

```
RP/0/RP0/CPU0:R3(config)# mpls static
RP/0/RP0/CPU0:R3(config-mpls-static)# interface HundredGigE0/0/0/0
RP/0/RP0/CPU0:R3(config-mpls-static)# exit
```

Enable SR MPLS under IS-IS:

```
RP/0/RP0/CPU0:R3(config)# router isis 1
RP/0/RP0/CPU0:R3(config-isis)# address-family ipv4 unicast
RP/0/RP0/CPU0:R3(config-isis-af)# segment-routing mpls
RP/0/RP0/CPU0:R3(config-isis-af)# metric-style wide
RP/0/RP0/CPU0:R3(config-isis-af)# exit
```

Configure prefix segment identifier (SID) value on the IS-IS enabled Loopback interface:

```
RP/0/RP0/CPU0:R3(config-isis)# interface Loopback0
RP/0/RP0/CPU0:R3(config-isis-if)# address-family ipv4 unicast
RP/0/RP0/CPU0:R3(config-isis-if-af)# prefix-sid absolute 16003
RP/0/RP0/CPU0:R3(config-isis-if-af)# exit
RP/0/RP0/CPU0:R3(config-isis-if)# exit
```

Enable IS-IS in core-facing interface:

```
RP/0/RP0/CPU0:R3(config-isis)# interface HundredGigE0/0/0/0
RP/0/RP0/CPU0:R3(config-isis-if)# point-to-point
RP/0/RP0/CPU0:R3(config-isis-if)# address-family ipv4 unicast
RP/0/RP0/CPU0:R3(config-isis-if-af)# exit
RP/0/RP0/CPU0:R3(config-isis-if)# exit
RP/0/RP0/CPU0:R3(config-isis)# exit
RP/0/RP0/CPU0:R3(config)#
```

Configure a route policy to advertise all BGP paths:

```
RP/0/RP0/CPU0:R3(config)# route-policy rpl_advertise_all_paths
RP/0/RP0/CPU0:R3(config-rpl)# set path-selection all advertise
RP/0/RP0/CPU0:R3(config-rpl)# set path-selection backup 1 install multipath-protect
RP/0/RP0/CPU0:R3(config-rpl)# end-policy
```

Enable advertisement of EPE-enabled BGP neighbors via BGP-LU:

```
RP/0/RP0/CPU0:R3(config)# router bgp 100
RP/0/RP0/CPU0:R3(config-bgp)# bgp router-id 1.1.1.3
RP/0/RP0/CPU0:R3(config-bgp)# ibgp policy out enforce-modifications
RP/0/RP0/CPU0:R3(config-bgp)# address-family ipv4 unicast
RP/0/RP0/CPU0:R3(config-bgp-af)# advertise epe-bgp labeled-unicast
RP/0/RP0/CPU0:R3(config-bgp-af)# additional-paths receive
RP/0/RP0/CPU0:R3(config-bgp-af)# additional-paths send
RP/0/RP0/CPU0:R3(config-bgp-af)# additional-paths selection route-policy
rpl_advertise_all_paths
RP/0/RP0/CPU0:R3(config-bgp-af)# allocate-label all
RP/0/RP0/CPU0:R3(config-bgp-af)# exit
```

Enable IPv4 unicast and IPv4 labeled unicast address families on iBGP peer:

```
RP/0/RP0/CPU0:R3(config-bgp)# neighbor 1.1.1.1
RP/0/RP0/CPU0:R3(config-bgp-nbr)# remote-as 100
RP/0/RP0/CPU0:R3(config-bgp-nbr)# update-source Loopback0
RP/0/RP0/CPU0:R3(config-bgp-nbr)# address-family ipv4 unicast
RP/0/RP0/CPU0:R3(config-bgp-nbr-af)# advertise local-labeled-route disable
RP/0/RP0/CPU0:R3(config-bgp-nbr-af)# exit
RP/0/RP0/CPU0:R3(config-bgp-nbr)# address-family ipv4 labeled-unicast
RP/0/RP0/CPU0:R3(config-bgp-nbr-af)# exit
RP/0/RP0/CPU0:R3(config-bgp-nbr)# exit
```

Enable EPE for the eBGP peers:

```
RP/0/RP0/CPU0:R3(config-bgp)# neighbor 10.3.40.40
RP/0/RP0/CPU0:R3(config-bgp-nbr)# remote-as 40
RP/0/RP0/CPU0:R3(config-bgp-nbr)# egress-engineering
RP/0/RP0/CPU0:R3(config-bgp-nbr)# address-family ipv4 unicast
RP/0/RP0/CPU0:R3(config-bgp-nbr-af)# route-policy pass_all in
RP/0/RP0/CPU0:R3(config-bgp-nbr-af)# route-policy pass_all out
RP/0/RP0/CPU0:R3(config-bgp-nbr-af)# exit
RP/0/RP0/CPU0:R3(config-bgp-nbr)# exit

RP/0/RP0/CPU0:R3(config-bgp)# neighbor 10.3.50.50
RP/0/RP0/CPU0:R3(config-bgp-nbr)# remote-as 50
RP/0/RP0/CPU0:R3(config-bgp-nbr)# egress-engineering
RP/0/RP0/CPU0:R3(config-bgp-nbr)# address-family ipv4 unicast
RP/0/RP0/CPU0:R3(config-bgp-nbr-af)# route-policy pass_all in
RP/0/RP0/CPU0:R3(config-bgp-nbr-af)# route-policy pass_all out
RP/0/RP0/CPU0:R3(config-bgp-nbr-af)# exit
RP/0/RP0/CPU0:R3(config-bgp-nbr)# exit
RP/0/RP0/CPU0:R3(config-bgp)# exit
RP/0/RP0/CPU0:R3(config)# commit
```

Egress Border Router R3 Running Configuration

```
segment-routing
 global-block 16000 23999
!

interface Loopback0
 ipv4 address 1.1.1.3 255.255.255.255
```

```
mpls static
 interface GigabitEthernet0/0/0/0
!

router isis 1
 is-type level-2-only
 net 47.0000.0000.0003.00
 address-family ipv4 unicast
  metric-style wide
  segment-routing mpls
 !
 interface Loopback0
  address-family ipv4 unicast
   prefix-sid absolute 16003
  !
 !
 interface HundredGigE0/0/0/0
  point-to-point
  address-family ipv4 unicast
  !
 !
!

route-policy rpl_advertise_all_paths
  set path-selection all advertise
  set path-selection backup 1 install multipath-protect
end-policy
!

router bgp 100
 bgp router-id 1.1.1.3
 ibgp policy out enforce-modifications
 address-family ipv4 unicast
  advertise epe-bgp labeled-unicast
  additional-paths receive
  additional-paths send
  additional-paths selection route-policy rpl_advertise_all_paths
  allocate-label all
 !
 neighbor 1.1.1.1
  remote-as 100
  update-source Loopback0
  address-family ipv4 unicast
   advertise local-labeled-route disable
  !
  address-family ipv4 labeled-unicast
  !
 !
 neighbor 10.3.40.40
  remote-as 40
  egress-engineering
  address-family ipv4 unicast
   route-policy pass_all in
   route-policy pass_all out
  !
 !
 neighbor 10.3.50.50
  remote-as 50
  egress-engineering
  address-family ipv4 unicast
   route-policy pass_all in
   route-policy pass_all out
  !
```

```
 !
!
```

Egress Border Router R4 Configuration

The configuration of egress border router R4 follows the configuration of R3:

```
RP/0/RP0/CPU0:R4(config)# segment-routing
RP/0/RP0/CPU0:R4(config-sr)# global-block 16000 23999
RP/0/RP0/CPU0:R4(config-sr)# exit

RP/0/RP0/CPU0:R4(config)# interface Loopback0
RP/0/RP0/CPU0:R4(config-if)# ipv4 address 1.1.1.4 255.255.255.255
RP/0/RP0/CPU0:R4(config-if)# exit

RP/0/RP0/CPU0:R4(config)# mpls static
RP/0/RP0/CPU0:R4(config-mpls-static)# interface HundredGigE0/0/0/0
RP/0/RP0/CPU0:R4(config-mpls-static)# exit

RP/0/RP0/CPU0:R4(config)# router isis 1
RP/0/RP0/CPU0:R4(config-isis)# address-family ipv4 unicast
RP/0/RP0/CPU0:R4(config-isis-af)# segment-routing mpls
RP/0/RP0/CPU0:R4(config-isis-af)# metric-style wide
RP/0/RP0/CPU0:R4(config-isis-af)# exit

RP/0/RP0/CPU0:R4(config-isis)# interface Loopback0
RP/0/RP0/CPU0:R4(config-isis-if)# address-family ipv4 unicast
RP/0/RP0/CPU0:R4(config-isis-if-af)# prefix-sid absolute 16004
RP/0/RP0/CPU0:R4(config-isis-if-af)# exit
RP/0/RP0/CPU0:R4(config-isis-if)# exit

RP/0/RP0/CPU0:R4(config-isis)# interface HundredGigE0/0/0/0
RP/0/RP0/CPU0:R4(config-isis-if)# point-to-point
RP/0/RP0/CPU0:R4(config-isis-if)# address-family ipv4 unicast
RP/0/RP0/CPU0:R4(config-isis-if-af)# exit
RP/0/RP0/CPU0:R4(config-isis-if)# exit
RP/0/RP0/CPU0:R4(config-isis)# exit

RP/0/RP0/CPU0:R4(config)# route-policy rpl_advertise_all_paths
RP/0/RP0/CPU0:R4(config-rpl)# set path-selection all advertise
RP/0/RP0/CPU0:R4(config-rpl)# set path-selection backup 1 install multipath-protect
RP/0/RP0/CPU0:R4(config-rpl)# end-policy

RP/0/RP0/CPU0:R4(config)# router bgp 100
RP/0/RP0/CPU0:R4(config-bgp)# bgp router-id 1.1.1.4
RP/0/RP0/CPU0:R4(config-bgp)# ibgp policy out enforce-modifications
RP/0/RP0/CPU0:R4(config-bgp)# address-family ipv4 unicast
RP/0/RP0/CPU0:R4(config-bgp-af)# advertise epe-bgp labeled-unicast
RP/0/RP0/CPU0:R4(config-bgp-af)# additional-paths receive
RP/0/RP0/CPU0:R4(config-bgp-af)# additional-paths send
RP/0/RP0/CPU0:R4(config-bgp-af)# additional-paths selection route-policy
rpl_advertise_all_paths
RP/0/RP0/CPU0:R4(config-bgp-af)# allocate-label all
RP/0/RP0/CPU0:R4(config-bgp-af)# exit

RP/0/RP0/CPU0:R4(config-bgp)# neighbor 1.1.1.1
RP/0/RP0/CPU0:R4(config-bgp-nbr)# remote-as 100
RP/0/RP0/CPU0:R4(config-bgp-nbr)# update-source Loopback0
RP/0/RP0/CPU0:R4(config-bgp-nbr)# address-family ipv4 unicast
RP/0/RP0/CPU0:R4(config-bgp-nbr-af)# advertise local-labeled-route disable
RP/0/RP0/CPU0:R4(config-bgp-nbr-af)# exit
RP/0/RP0/CPU0:R4(config-bgp-nbr)# address-family ipv4 labeled-unicast
```

```
RP/0/RP0/CPU0:R4(config-bgp-nbr-af)# exit
RP/0/RP0/CPU0:R4(config-bgp-nbr)# exit

RP/0/RP0/CPU0:R4(config-bgp)# neighbor 10.4.40.40
RP/0/RP0/CPU0:R4(config-bgp-nbr)# remote-as 40
RP/0/RP0/CPU0:R4(config-bgp-nbr)# egress-engineering
RP/0/RP0/CPU0:R4(config-bgp-nbr)# address-family ipv4 unicast
RP/0/RP0/CPU0:R4(config-bgp-nbr-af)# route-policy pass_all in
RP/0/RP0/CPU0:R4(config-bgp-nbr-af)# route-policy pass_all out
RP/0/RP0/CPU0:R4(config-bgp-nbr-af)# exit
RP/0/RP0/CPU0:R4(config-bgp-nbr)# exit

RP/0/RP0/CPU0:R4(config-bgp)# neighbor 10.4.50.50
RP/0/RP0/CPU0:R4(config-bgp-nbr)# remote-as 50
RP/0/RP0/CPU0:R4(config-bgp-nbr)# egress-engineering
RP/0/RP0/CPU0:R4(config-bgp-nbr)# address-family ipv4 unicast
RP/0/RP0/CPU0:R4(config-bgp-nbr-af)# route-policy pass_all in
RP/0/RP0/CPU0:R4(config-bgp-nbr-af)# route-policy pass_all out
RP/0/RP0/CPU0:R4(config-bgp-nbr-af)# exit
RP/0/RP0/CPU0:R4(config-bgp-nbr)# exit
RP/0/RP0/CPU0:R4(config-bgp)# exit
RP/0/RP0/CPU0:R4(config)# commit
```

Egress Border Router R4 Running Configuration

```
segment-routing
 global-block 16000 23999

interface Loopback0
 ipv4 address 1.1.1.4 255.255.255.255

mpls static
 interface GigabitEthernet0/0/0/0
!

router isis 1
 is-type level-2-only
 net 47.0000.0000.0004.00
 address-family ipv4 unicast
  metric-style wide
  segment-routing mpls
 !

 interface Loopback0
  address-family ipv4 unicast
   prefix-sid absolute 16004
  !
 !
 interface HundredGigE0/0/0/0
  point-to-point
  address-family ipv4 unicast
  !
 !

route-policy rpl_advertise_all_paths
  set path-selection all advertise
  set path-selection backup 1 install multipath-protect
end-policy
!

router bgp 100
 bgp router-id 1.1.1.4
 ibgp policy out enforce-modifications
```

```
 address-family ipv4 unicast
  advertise epe-bgp labeled-unicast
  additional-paths receive
  additional-paths send
  additional-paths selection route-policy rpl_advertise_all_paths
  allocate-label all
 !
 neighbor 1.1.1.1
  remote-as 100
  update-source Loopback0
  address-family ipv4 unicast
   advertise local-labeled-route disable
  !
  address-family ipv4 labeled-unicast
  !
 !
 neighbor 10.4.40.40
  remote-as 40
  egress-engineering
  address-family ipv4 unicast
   route-policy pass_all in
   route-policy pass_all out
  !
 !
 neighbor 10.4.50.50
  remote-as 50
  egress-engineering
  address-family ipv4 unicast
   route-policy pass_all in
   route-policy pass_all out
  !
 !
!
```

Ingress Border Router R1 Configuration

Configure the SRGB:

```
RP/0/RP0/CPU0:R1(config)# segment-routing
RP/0/RP0/CPU0:R1(config-sr)# global-block 16000 23999
RP/0/RP0/CPU0:R1(config-sr)# exit
RP/0/RP0/CPU0:R1(config)#
```

Configure the Loopback addresses. Lo0 is advertised in IS-IS and used a BGP next-hop. Lo100 is advertised in BGP as an overlay prefix:

```
RP/0/RP0/CPU0:R1(config)# interface Loopback0
RP/0/RP0/CPU0:R1(config-if)# ipv4 address 1.1.1.1 255.255.255.255
RP/0/RP0/CPU0:R1(config-if)# exit

RP/0/RP0/CPU0:R1(config)# interface Loopback100
RP/0/RP0/CPU0:R1(config-if)# ipv4 address 151.1.1.1 255.255.255.255
RP/0/RP0/CPU0:R1(config-if)# exit
```

Enable SR MPLS under IS-IS:

```
RP/0/RP0/CPU0:R1(config)# router isis 1
RP/0/RP0/CPU0:R1(config-isis)# address-family ipv4 unicast
RP/0/RP0/CPU0:R1(config-isis-af)# metric-style wide
RP/0/RP0/CPU0:R1(config-isis-af)# segment-routing mpls
RP/0/RP0/CPU0:R1(config-isis-af)# exit
```

Configure prefix segment identifier (SID) value on the IS-IS enabled Loopback interface:

```
RP/0/RP0/CPU0:R1(config-isis)# interface Loopback0 address-family ipv4 unicast
RP/0/RP0/CPU0:R1(config-isis-if-af)# prefix-sid absolute 16001
RP/0/RP0/CPU0:R1(config-isis-if-af)# exit
RP/0/RP0/CPU0:R1(config-isis-if)# exit
```

Enable IS-IS in core-facing interface:

```
RP/0/RP0/CPU0:R1(config-isis)# interface HundredGigE0/0/0/0
RP/0/RP0/CPU0:R1(config-isis-if)# point-to-point
RP/0/RP0/CPU0:R1(config-isis-if)# address-family ipv4 unicast
RP/0/RP0/CPU0:R1(config-isis-if-af)# exit
RP/0/RP0/CPU0:R1(config-isis-if)# exit
RP/0/RP0/CPU0:R1(config-isis)# exit
```

Configure an RPL policy to prevent allocation of local label to overlay prefixes; such as Lo100 151.1.1.1/32:

```
RP/0/RP0/CPU0:R1(config)# prefix-set unlabelled_prefixes
RP/0/RP0/CPU0:R1(config-pfx)# 151.1.1.1/32
RP/0/RP0/CPU0:R1(config-pfx)# end-set
RP/0/RP0/CPU0:R1(config)# route-policy rpl_allocate_label
RP/0/RP0/CPU0:R1(config-rpl)# if destination in unlabelled_prefixes then
RP/0/RP0/CPU0:R1(config-rpl-if)# drop
RP/0/RP0/CPU0:R1(config-rpl-if)# else
RP/0/RP0/CPU0:R1(config-rpl-else)# pass
RP/0/RP0/CPU0:R1(config-rpl-else)# endif
RP/0/RP0/CPU0:R1(config-rpl)# end-policy
RP/0/RP0/CPU0:R1(config)#
```

Configure an RPL policy to influence the best-path selection by assigning a higher BGP local preference to the desired path. In this example, the desired egress exit path for prefix 161.1.1.0/28 is via R4 and then AS 50, and for prefix 161.1.1.1/32 is via R4 and then AS 40. Otherwise, the uninfluenced exit path for these prefixes is via R3:

```
RP/0/RP0/CPU0:R1(config)# route-policy rpl_epe
RP/0/RP0/CPU0:R1(config-rpl)# if destination in (161.1.1.0/28) and next-hop in (10.4.50.50)
 then
RP/0/RP0/CPU0:R1(config-rpl-if)# set local-preference 1000
RP/0/RP0/CPU0:R1(config-rpl-if)# elseif destination in (161.1.1.1/32) and next-hop in
(10.4.40.40) then
RP/0/RP0/CPU0:R1(config-rpl-elseif)# set local-preference 1000
RP/0/RP0/CPU0:R1(config-rpl-elseif)# endif
RP/0/RP0/CPU0:R1(config-rpl)# pass
RP/0/RP0/CPU0:R1(config-rpl)# end-policy
RP/0/RP0/CPU0:R1(config)#
```

Configure an RPL policy to advertise all candidate paths:

```
RP/0/RP0/CPU0:R1(config)# route-policy rpl_advertise_all_paths
RP/0/RP0/CPU0:R1(config-rpl)# set path-selection all advertise
RP/0/RP0/CPU0:R1(config-rpl)# end-policy


RP/0/RP0/CPU0:R1(config)# router bgp 100
RP/0/RP0/CPU0:R1(config-bgp)# bgp router-id 1.1.1.1
RP/0/RP0/CPU0:R1(config-bgp)# address-family ipv4 unicast
RP/0/RP0/CPU0:R1(config-bgp-af)# additional-paths receive
RP/0/RP0/CPU0:R1(config-bgp-af)# additional-paths send
RP/0/RP0/CPU0:R1(config-bgp-af)# additional-paths selection route-policy
rpl_advertise_all_paths
```

```
RP/0/RP0/CPU0:R1(config-bgp-af)# exit


RP/0/RP0/CPU0:R1(config-bgp)# neighbor 1.1.1.3
RP/0/RP0/CPU0:R1(config-bgp-nbr)# remote-as 100
RP/0/RP0/CPU0:R1(config-bgp-nbr)# update-source Loopback0
RP/0/RP0/CPU0:R1(config-bgp-nbr)# address-family ipv4 unicast
RP/0/RP0/CPU0:R1(config-bgp-nbr-af)# advertise local-labeled-route disable
RP/0/RP0/CPU0:R1(config-bgp-nbr-af)# exit
RP/0/RP0/CPU0:R1(config-bgp-nbr)# address-family ipv4 labeled-unicast
RP/0/RP0/CPU0:R1(config-bgp-nbr-af)# exit
RP/0/RP0/CPU0:R1(config-bgp-nbr)# exit

RP/0/RP0/CPU0:R1(config-bgp)# neighbor 1.1.1.4
RP/0/RP0/CPU0:R1(config-bgp-nbr)# remote-as 100
RP/0/RP0/CPU0:R1(config-bgp-nbr)# update-source Loopback0
RP/0/RP0/CPU0:R1(config-bgp-nbr)# address-family ipv4 unicast
RP/0/RP0/CPU0:R1(config-bgp-nbr-af)# advertise local-labeled-route disable
RP/0/RP0/CPU0:R1(config-bgp-nbr-af)# exit
RP/0/RP0/CPU0:R1(config-bgp-nbr)# address-family ipv4 labeled-unicast
RP/0/RP0/CPU0:R1(config-bgp-nbr-af)# exit
RP/0/RP0/CPU0:R1(config-bgp-nbr)# exit
RP/0/RP0/CPU0:R1(config-bgp)# exit
RP/0/RP0/CPU0:R1(config)# commit
```

Ingress Border Router R1 Running Configuration

```
segment-routing
 global-block 16000 23999

interface Loopback0
 ipv4 address 1.1.1.1 255.255.255.255
!
interface Loopback100
 ipv4 address 151.1.1.1 255.255.255.255

router isis 1
 is-type level-2-only
 net 47.0000.0000.0001.00
 address-family ipv4 unicast
  metric-style wide
  segment-routing mpls
 !
 interface Loopback0
  address-family ipv4 unicast
   prefix-sid absolute 16001
  !
 !
 interface HundredGigE0/0/0/0
  point-to-point
  address-family ipv4 unicast
  !
 !
!

prefix-set unlabelled_prefixes
  151.1.1.1/32
end-set
!

route-policy rpl_allocate_label
  if destination in unlabelled_prefixes then
    drop
```

```
  else
    pass
  endif
end-policy
!

route-policy rpl_epe
  if destination in (161.1.1.0/28) and next-hop in (10.4.50.50) then
    set local-preference 1000
  elseif destination in (161.1.1.1/32) and next-hop in (10.4.40.40) then
    set local-preference 1000
  endif
  pass
end-policy
!

route-policy rpl_advertise_all_paths
  set path-selection all advertise
end-policy
!

router bgp 100
 bgp router-id 1.1.1.1
 ibgp policy out enforce-modifications
 address-family ipv4 unicast
  additional-paths receive
  additional-paths send
  additional-paths selection route-policy rpl_advertise_all_paths
  network 151.1.1.1/32
  allocate-label route-policy rpl_allocate_label
 !
 neighbor 1.1.1.3
  remote-as 100
  update-source Loopback0
  address-family ipv4 unicast
   advertise local-labeled-route disable
  !
  address-family ipv4 labeled-unicast
  !
 !
 neighbor 1.1.1.4
  remote-as 100
  update-source Loopback0
  address-family ipv4 unicast
   advertise local-labeled-route disable
  !
  address-family ipv4 labeled-unicast
  !
 !
!
```

The following sections depict the **show** command outputs associated with the Egress Border routers (R3, R4) and Ingress PE router (R1):

Egress Border Router R3 Output

The following commands show the BGP EPE labels allocated for eBGP neighbors 10.3.40.40 and 10.3.50.50 alongside their corresponding entries in the FIB:

```
RP/0/RP0/CPU0:R3# show bgp egress-engineering

 Egress Engineering Object: 10.3.40.40/32 (0x7fc163c62e80)
       EPE Type: Peer
```

```
       Nexthop: 10.3.40.40
       Version: 2, rn_version: 2
         Flags: 0x00000006
     Local ASN: 100
    Remote ASN: 40
     Local RID: 1.1.1.3
    Remote RID: 1.1.1.40
  Local Address: 10.3.40.3
     First Hop: 10.3.40.40
          NHID: 0
           IFH: 0x198
         Label: 24004, Refcount: 4
       rpc_set: 0x7fc14410ff18, ID: 1

 Egress Engineering Object: 10.3.50.50/32 (0x7fc163c62d88)
      EPE Type: Peer
       Nexthop: 10.3.50.50
       Version: 3, rn_version: 3
         Flags: 0x00000006
     Local ASN: 100
    Remote ASN: 50
     Local RID: 1.1.1.3
    Remote RID: 1.1.1.50
  Local Address: 10.3.50.3
     First Hop: 10.3.50.50
          NHID: 0
           IFH: 0x1a0
         Label: 24005, Refcount: 4
       rpc_set: 0x7fc144110088, ID: 2

RP/0/RP0/CPU0:R3# show mpls forwarding labels 24004
Thu Feb  3 22:11:18.459 UTC
Local  Outgoing    Prefix             Outgoing     Next Hop        Bytes
Label  Label       or ID              Interface                    Switched
------ ----------- ------------------ ------------ --------------- ------------
24004  Pop         No ID              Hu0/0/0/1    10.3.40.40      0

RP/0/RP0/CPU0:R3# show mpls forwarding labels 24005
Thu Feb  3 22:11:35.399 UTC
Local  Outgoing    Prefix             Outgoing     Next Hop        Bytes
Label  Label       or ID              Interface                    Switched
------ ----------- ------------------ ------------ --------------- ------------
24005  Pop         No ID              Hu0/0/0/2    10.3.50.50      0
```

The following output displays the BGP-LU prefixes used to advertise the EPE-enabled eBGP neighbors 10.3.40.40 and 10.3.50.50:

```
RP/0/RP0/CPU0:R3# show bgp ipv4 labeled-unicast
Thu Feb  3 22:11:57.865 UTC
BGP router identifier 1.1.1.3, local AS number 100
BGP generic scan interval 60 secs
Non-stop routing is enabled
BGP table state: Active
Table ID: 0xe0000000   RD version: 4
BGP main routing table version 4
BGP NSR Initial initsync version 2 (Reached)
BGP NSR/ISSU Sync-Group versions 0/0
BGP scan interval 60 secs

Status codes: s suppressed, d damped, h history, * valid, > best
              i - internal, r RIB-failure, S stale, N Nexthop-discard
Origin codes: i - IGP, e - EGP, ? - incomplete
```

```
   Network            Next Hop            Metric LocPrf Weight Path
*> 10.3.40.40/32      0.0.0.0                              0 i
*> 10.3.50.50/32      0.0.0.0                              0 i

Processed 2 prefixes, 2 paths
```

The details of the BGP-LU prefixes can be found below. Note that the EPE label is advertised in BGP-LU.

```
RP/0/RP0/CPU0:R3# show bgp ipv4 labeled-unicast 10.3.40.40/32
Thu Feb  3 22:12:18.210 UTC
BGP routing table entry for 10.3.40.40/32
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker                 3           3
    Local Label: 24004
Last Modified: Feb  3 19:13:07.039 for 02:59:11
Paths: (1 available, best #1)
  Advertised IPv4 Unicast paths to update-groups (with more than one peer):
    0.4
  Path #1: Received by speaker 0
  Advertised IPv4 Unicast paths to update-groups (with more than one peer):
    0.4
  Local
    0.0.0.0 from 0.0.0.0 (1.1.1.3)
      Origin IGP, localpref 100, valid, extranet, best, group-best
      Received Path ID 0, Local Path ID 1, version 3
      Origin-AS validity: not-found


RP/0/RP0/CPU0:R3# show bgp ipv4 labeled-unicast 10.3.50.50/32
Thu Feb  3 22:12:27.282 UTC
BGP routing table entry for 10.3.50.50/32
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker                 4           4
    Local Label: 24005
Last Modified: Feb  3 19:13:07.039 for 02:59:20
Paths: (1 available, best #1)
  Advertised IPv4 Unicast paths to update-groups (with more than one peer):
    0.4
  Path #1: Received by speaker 0
  Advertised IPv4 Unicast paths to update-groups (with more than one peer):
    0.4
  Local
    0.0.0.0 from 0.0.0.0 (1.1.1.3)
      Origin IGP, localpref 100, valid, extranet, best, group-best
      Received Path ID 0, Local Path ID 1, version 4
      Origin-AS validity: not-found
```

The output below depicts the BGP route and CEF details for an overlay prefix (161.1.1.0/28) learned via the EPE-enabled BGP neighbors:

```
RP/0/RP0/CPU0:R3# show bgp ipv4 unicast
Thu Feb  3 22:58:01.736 UTC
BGP router identifier 1.1.1.3, local AS number 100
BGP generic scan interval 60 secs
Non-stop routing is enabled
BGP table state: Active
Table ID: 0xe0000000   RD version: 14
BGP main routing table version 14
BGP NSR Initial initsync version 2 (Reached)
BGP NSR/ISSU Sync-Group versions 0/0
BGP scan interval 60 secs
```

```
Status codes: s suppressed, d damped, h history, * valid, > best
              i - internal, r RIB-failure, S stale, N Nexthop-discard
Origin codes: i - IGP, e - EGP, ? - incomplete
   Network            Next Hop            Metric LocPrf Weight Path
*> 10.3.40.40/32      0.0.0.0                                0 i
*> 10.3.50.50/32      0.0.0.0                                0 i
*>i151.1.1.1/32       1.1.1.1                 0    100      0 i
*> 161.1.1.0/28       10.3.40.40                             0 40 60 i
*                     10.3.50.50                             0 50 60 i

Processed 4 prefixes, 5 paths


RP/0/RP0/CPU0:R3# show bgp ipv4 unicast 161.1.1.0/28
Thu Feb  3 22:31:52.893 UTC
BGP routing table entry for 161.1.1.0/28
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker                 6           6
Last Modified: Feb  3 22:28:56.039 for 00:02:56
Paths: (2 available, best #1)
  Advertised IPv4 Unicast paths to update-groups (with more than one peer):
    0.4
  Advertised IPv4 Unicast paths to peers (in unique update groups):
    1.1.1.1
  Path #1: Received by speaker 0
  Advertised IPv4 Unicast paths to update-groups (with more than one peer):
    0.4
  Advertised IPv4 Unicast paths to peers (in unique update groups):
    1.1.1.1
  40 60
    10.3.40.40 from 10.3.40.40 (1.1.1.40)
      Origin IGP, localpref 100, valid, external, best, group-best
      Received Path ID 0, Local Path ID 1, version 5
      Origin-AS validity: (disabled)
  Path #2: Received by speaker 0
  Advertised IPv4 Unicast paths to peers (in unique update groups):
    1.1.1.1
  50 60
    10.3.50.50 from 10.3.50.50 (1.1.1.50)
      Origin IGP, localpref 100, valid, external, group-best, backup, add-path
      Received Path ID 0, Local Path ID 2, version 6
      Origin-AS validity: (disabled)


RP/0/RP0/CPU0:R3# show cef ipv4 161.1.1.0/28
Thu Feb 10 20:17:29.240 UTC
161.1.1.0/28, version 24, internal 0x5000001 0x40 (ptr 0x90684920) [1], 0x0 (0x0), 0x0 (0x0)

 Updated Feb 10 17:37:16.609
 Prefix Len 28, traffic index 0, precedence n/a, priority 4
   via 10.3.40.40/32, 5 dependencies, recursive, bgp-ext [flags 0x6020]
    path-idx 0 NHID 0x0 [0x90684c08 0x0], Internal 0x90211730
    next hop 10.3.40.40/32 via 10.3.40.40/32
   via 10.3.50.50/32, 4 dependencies, recursive, bgp-ext, backup [flags 0x6120]
    path-idx 1 NHID 0x0 [0x90685040 0x0]
    next hop 10.3.50.50/32 via 10.3.50.50/32
```

Egress Border Router R4 Output

The following outputs correspond to egress border router R4. They follow the same sequence shown for router R3.

```
RP/0/RP0/CPU0:R4# show bgp egress-engineering

Egress Engineering Object: 10.4.40.40/32 (0x7f84d2a4ae80)
       EPE Type: Peer
        Nexthop: 10.4.40.40
        Version: 2, rn_version: 2
          Flags: 0x00000006
      Local ASN: 100
     Remote ASN: 40
      Local RID: 1.1.1.4
     Remote RID: 1.1.1.40
  Local Address: 10.4.40.4
      First Hop: 10.4.40.40
           NHID: 0
            IFH: 0x198
          Label: 24004, Refcount: 4
        rpc_set: 0x7f84b010fdb8, ID: 1

 Egress Engineering Object: 10.4.50.50/32 (0x7f84d2a4ad88)
       EPE Type: Peer
        Nexthop: 10.4.50.50
        Version: 3, rn_version: 3
          Flags: 0x00000006
      Local ASN: 100
     Remote ASN: 50
      Local RID: 1.1.1.4
     Remote RID: 1.1.1.50
  Local Address: 10.4.50.4
      First Hop: 10.4.50.50
           NHID: 0
            IFH: 0x1a0
          Label: 24005, Refcount: 4
        rpc_set: 0x7f84b010ff28, ID: 2


RP/0/RP0/CPU0:R4# show mpls forwarding labels 24004
Thu Feb  3 22:34:55.059 UTC
Local  Outgoing    Prefix             Outgoing     Next Hop        Bytes
Label  Label       or ID              Interface                    Switched
------ ----------- ------------------ ------------ --------------- ------------
24004  Pop         No ID              Hu0/0/0/1    10.4.40.40      0


RP/0/RP0/CPU0:R4# show mpls forwarding labels 24005
Thu Feb  3 22:35:07.252 UTC
Local  Outgoing    Prefix             Outgoing     Next Hop        Bytes
Label  Label       or ID              Interface                    Switched
------ ----------- ------------------ ------------ --------------- ------------
24005  Pop         No ID              Hu0/0/0/2    10.4.50.50      0


RP/0/RP0/CPU0:R4# show bgp ipv4 labeled-unicast
Thu Feb  3 22:59:37.978 UTC
BGP router identifier 1.1.1.4, local AS number 100
BGP generic scan interval 60 secs
Non-stop routing is enabled
BGP table state: Active
Table ID: 0xe0000000   RD version: 14
BGP main routing table version 14
BGP NSR Initial initsync version 2 (Reached)
BGP NSR/ISSU Sync-Group versions 0/0
BGP scan interval 60 secs

Status codes: s suppressed, d damped, h history, * valid, > best
              i - internal, r RIB-failure, S stale, N Nexthop-discard
```

```
Origin codes: i - IGP, e - EGP, ? - incomplete
   Network            Next Hop            Metric LocPrf Weight Path
*> 10.4.40.40/32      0.0.0.0                                0 i
*> 10.4.50.50/32      0.0.0.0                                0 i

Processed 2 prefixes, 2 paths


RP/0/RP0/CPU0:R4# show bgp ipv4 labeled-unicast 10.4.40.40/32
Thu Feb  3 22:35:41.275 UTC
BGP routing table entry for 10.4.40.40/32
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker                 3           3
    Local Label: 24004
Last Modified: Feb  3 19:13:08.143 for 03:22:33
Paths: (1 available, best #1)
  Advertised IPv4 Unicast paths to update-groups (with more than one peer):
    0.4
  Advertised IPv4 Labeled-unicast paths to peers (in unique update groups):
    1.1.1.1
  Path #1: Received by speaker 0
  Advertised IPv4 Unicast paths to update-groups (with more than one peer):
    0.4
  Advertised IPv4 Labeled-unicast paths to peers (in unique update groups):
    1.1.1.1
  Local
    0.0.0.0 from 0.0.0.0 (1.1.1.4)
      Origin IGP, localpref 100, valid, extranet, best, group-best
      Received Path ID 0, Local Path ID 1, version 3
      Origin-AS validity: not-found


RP/0/RP0/CPU0:R4# show bgp ipv4 labeled-unicast 10.4.50.50/32
Thu Feb  3 22:35:53.259 UTC
BGP routing table entry for 10.4.50.50/32
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker                 4           4
    Local Label: 24005
Last Modified: Feb  3 19:13:08.143 for 03:22:45
Paths: (1 available, best #1)
  Advertised IPv4 Unicast paths to update-groups (with more than one peer):
    0.4
  Advertised IPv4 Labeled-unicast paths to peers (in unique update groups):
    1.1.1.1
  Path #1: Received by speaker 0
  Advertised IPv4 Unicast paths to update-groups (with more than one peer):
    0.4
  Advertised IPv4 Labeled-unicast paths to peers (in unique update groups):
    1.1.1.1
  Local
    0.0.0.0 from 0.0.0.0 (1.1.1.4)
      Origin IGP, localpref 100, valid, extranet, best, group-best
      Received Path ID 0, Local Path ID 1, version 4
      Origin-AS validity: not-found


RP/0/RP0/CPU0:R4# show bgp ipv4 unicast
Thu Feb  3 23:00:32.470 UTC
BGP router identifier 1.1.1.4, local AS number 100
BGP generic scan interval 60 secs
Non-stop routing is enabled
BGP table state: Active
Table ID: 0xe0000000   RD version: 14
```

```
BGP main routing table version 14
BGP NSR Initial initsync version 2 (Reached)
BGP NSR/ISSU Sync-Group versions 0/0
BGP scan interval 60 secs

Status codes: s suppressed, d damped, h history, * valid, > best
              i - internal, r RIB-failure, S stale, N Nexthop-discard
Origin codes: i - IGP, e - EGP, ? - incomplete
   Network            Next Hop            Metric LocPrf Weight Path
*> 10.4.40.40/32      0.0.0.0                                  0 i
*> 10.4.50.50/32      0.0.0.0                                  0 i
*>i151.1.1.1/32       1.1.1.1                  0    100      0 i
*> 161.1.1.0/28       10.4.40.40                               0 40 60 i
*                     10.4.50.50                               0 50 60 i

Processed 4 prefixes, 5 paths


RP/0/RP0/CPU0:R4# show bgp ipv4 unicast 161.1.1.0/28
Thu Feb  3 22:36:09.266 UTC
BGP routing table entry for 161.1.1.0/28
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker                 6           6
Last Modified: Feb  3 22:28:56.143 for 00:07:13
Paths: (2 available, best #1)
  Advertised IPv4 Unicast paths to update-groups (with more than one peer):
    0.4
  Advertised IPv4 Unicast paths to peers (in unique update groups):
    1.1.1.1
  Path #1: Received by speaker 0
  Advertised IPv4 Unicast paths to update-groups (with more than one peer):
    0.4
  Advertised IPv4 Unicast paths to peers (in unique update groups):
    1.1.1.1
  40 60
    10.4.40.40 from 10.4.40.40 (1.1.1.40)
      Origin IGP, localpref 100, valid, external, best, group-best
      Received Path ID 0, Local Path ID 1, version 5
      Origin-AS validity: (disabled)
  Path #2: Received by speaker 0
  Advertised IPv4 Unicast paths to peers (in unique update groups):
    1.1.1.1
  50 60
    10.4.50.50 from 10.4.50.50 (1.1.1.50)
      Origin IGP, localpref 100, valid, external, group-best, backup, add-path
      Received Path ID 0, Local Path ID 2, version 6
      Origin-AS validity: (disabled)
```

Ingress Border Router R1 Output

This section includes the outputs corresponding to ingress border router R1.

R1 learns the eBGP neighbor IP addresses via BGP-LU. In the details for each neighbor prefix, observe that the advertised BGP-LU label corresponds to the EPE label at the egress border router (R3 or R4).

```
RP/0/RP0/CPU0:R1# show bgp ipv4 labeled-unicast
Thu Feb 10 20:18:59.645 UTC
BGP router identifier 1.1.1.1, local AS number 100
BGP generic scan interval 60 secs
Non-stop routing is enabled
BGP table state: Active
Table ID: 0xe0000000   RD version: 8
BGP main routing table version 8
```

```
BGP NSR Initial initsync version 5 (Reached)
BGP NSR/ISSU Sync-Group versions 0/0
BGP scan interval 60 secs

Status codes: s suppressed, d damped, h history, * valid, > best
              i - internal, r RIB-failure, S stale, N Nexthop-discard
Origin codes: i - IGP, e - EGP, ? - incomplete
   Network            Next Hop            Metric LocPrf Weight Path
*>i10.3.40.40/32      1.1.1.3                  100      0 i
*>i10.3.50.50/32      1.1.1.3                  100      0 i
*>i10.4.40.40/32      1.1.1.4                  100      0 i
*>i10.4.50.50/32      1.1.1.4                  100      0 i

Processed 4 prefixes, 4 paths


RP/0/RP0/CPU0:R1# show bgp ipv4 labeled-unicast 10.3.40.40/32
Thu Feb  3 23:01:57.912 UTC
BGP routing table entry for 10.3.40.40/32
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker                15          15
    Local Label: 24004
Last Modified: Feb  3 22:47:43.539 for 00:14:14
Paths: (1 available, best #1)
  Not advertised to any peer
  Path #1: Received by speaker 0
  Not advertised to any peer
  Local
    1.1.1.3 (metric 30) from 1.1.1.3 (1.1.1.3)
      Received Label 24004
      Origin IGP, localpref 100, valid, internal, best, group-best, labeled-unicast
      Received Path ID 1, Local Path ID 1, version 15


RP/0/RP0/CPU0:R1# show bgp ipv4 labeled-unicast 10.3.50.50/32
Thu Feb  3 23:02:09.173 UTC
BGP routing table entry for 10.3.50.50/32
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker                16          16
    Local Label: 24005
Last Modified: Feb  3 22:47:43.539 for 00:14:25
Paths: (1 available, best #1)
  Not advertised to any peer
  Path #1: Received by speaker 0
  Not advertised to any peer
  Local
    1.1.1.3 (metric 30) from 1.1.1.3 (1.1.1.3)
      Received Label 24005
      Origin IGP, localpref 100, valid, internal, best, group-best, labeled-unicast
      Received Path ID 1, Local Path ID 1, version 16


RP/0/RP0/CPU0:R1# show bgp ipv4 labeled-unicast 10.4.40.40/32
Thu Feb  3 23:02:18.843 UTC
BGP routing table entry for 10.4.40.40/32
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker                17          17
    Local Label: 24006
Last Modified: Feb  3 22:47:43.539 for 00:14:35
Paths: (1 available, best #1)
  Not advertised to any peer
  Path #1: Received by speaker 0
```

```
  Not advertised to any peer
  Local
    1.1.1.4 (metric 30) from 1.1.1.4 (1.1.1.4)
      Received Label 24004
      Origin IGP, localpref 100, valid, internal, best, group-best, labeled-unicast
      Received Path ID 1, Local Path ID 1, version 17


RP/0/RP0/CPU0:R1# show bgp ipv4 labeled-unicast 10.4.50.50/32
Thu Feb  3 23:02:27.622 UTC
BGP routing table entry for 10.4.50.50/32
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker                 18          18
    Local Label: 24007
Last Modified: Feb  3 22:47:43.539 for 00:14:44
Paths: (1 available, best #1)
  Not advertised to any peer
  Path #1: Received by speaker 0
  Not advertised to any peer
  Local
    1.1.1.4 (metric 30) from 1.1.1.4 (1.1.1.4)
      Received Label 24005
      Origin IGP, localpref 100, valid, internal, best, group-best, labeled-unicast
      Received Path ID 1, Local Path ID 1, version 18


RP/0/RP0/CPU0:R1# show isis segment-routing label table

IS-IS 1 IS Label Table
Label         Prefix                 Interface
----------    ----------------       ---------
16001         1.1.1.1/32             Loopback0
16002         1.1.1.2/32
16003         1.1.1.3/32
16004         1.1.1.4/32
```

The following show commands depict the RIB and CEF outputs for the loopbacks of R3 and R4 learned vis ISIS-SR:

```
RP/0/RP0/CPU0:R1# show route 1.1.1.3/32

Routing entry for 1.1.1.3/32
  Known via "isis 1", distance 115, metric 30, labeled SR, type level-2
  Installed Feb 10 17:36:12.497 for 02:43:40
  Routing Descriptor Blocks
    10.1.2.2, from 1.1.1.3, via HundredGigE0/0/0/0
      Route metric is 30
  No advertising protos.


RP/0/RP0/CPU0:R1# show route 1.1.1.4/32

Routing entry for 1.1.1.4/32
  Known via "isis 1", distance 115, metric 30, labeled SR, type level-2
  Installed Feb 10 17:37:02.171 for 02:42:59
  Routing Descriptor Blocks
    10.1.2.2, from 1.1.1.4, via HundredGigE0/0/0/0
      Route metric is 30
  No advertising protos.


RP/0/RP0/CPU0:R1# show cef 1.1.1.3/32
```

```
1.1.1.3/32, version 18, labeled SR, internal 0x1000001 0x8110 (ptr 0x90cd33a0) [1], 0x0
(0x90c3eb10), 0xa28 (0x91a18378)
 Updated Feb 10 17:36:12.506
 local adjacency to HundredGigE0/0/0/0

 Prefix Len 32, traffic index 0, precedence n/a, priority 1
   via 10.1.2.2/32, HundredGigE0/0/0/0, 7 dependencies, weight 0, class 0 [flags 0x0]
    path-idx 0 NHID 0x0 [0x91de84d8 0x0]
    next hop 10.1.2.2/32
    local adjacency
     local label 16003      labels imposed {16003}


RP/0/RP0/CPU0:R1# show cef 1.1.1.4/32

1.1.1.4/32, version 20, labeled SR, internal 0x1000001 0x8110 (ptr 0x90cd32c8) [1], 0x0
(0x90c3eb58), 0xa28 (0x91a18408)
 Updated Feb 10 17:37:02.176
 local adjacency to HundredGigE0/0/0/0

 Prefix Len 32, traffic index 0, precedence n/a, priority 1
   via 10.1.2.2/32, HundredGigE0/0/0/0, 7 dependencies, weight 0, class 0 [flags 0x0]
    path-idx 0 NHID 0x0 [0x91de84d8 0x0]
    next hop 10.1.2.2/32
    local adjacency
     local label 16004      labels imposed {16004}
```

Next, we observe the BGP table for overlay prefixes at R1. In this usecase, we use prefix 161.1.1.0/28 as an overlay prefix learned from AS 40 and AS 50. Note that all BGP paths are present at R1 with a BGP next-hop unchanged. By default and without any BGP policy applied, the BGP best-path is the path from NH 10.3.40.40 (AS 40 via R3).

```
RP/0/RP0/CPU0:R1# show bgp

BGP router identifier 1.1.1.1, local AS number 100
BGP generic scan interval 60 secs
Non-stop routing is enabled
BGP table state: Active
Table ID: 0xe0000000   RD version: 21
BGP main routing table version 21
BGP NSR Initial initsync version 7 (Reached)
BGP NSR/ISSU Sync-Group versions 0/0
BGP scan interval 60 secs

Status codes: s suppressed, d damped, h history, * valid, > best
              i - internal, r RIB-failure, S stale, N Nexthop-discard
Origin codes: i - IGP, e - EGP, ? - incomplete
   Network            Next Hop            Metric LocPrf Weight Path
*>i10.3.40.40/32      1.1.1.3                  100      0 i
*>i10.3.50.50/32      1.1.1.3                  100      0 i
*>i10.4.40.40/32      1.1.1.4                  100      0 i
*>i10.4.50.50/32      1.1.1.4                  100      0 i
*> 151.1.1.1/32       0.0.0.0                  0         32768 i
*>i161.1.1.0/28       10.3.40.40               100      0 40 60 i
* i                   10.3.50.50               100      0 50 60 i
* i                   10.4.40.40               100      0 40 60 i
* i                   10.4.50.50               100      0 50 60 i

Processed 6 prefixes, 9 paths
```

```
RP/0/RP0/CPU0:R1# show bgp ipv4 unicast 161.1.1.0/28

BGP routing table entry for 161.1.1.0/28
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker                 8           8
Last Modified: Feb 10 17:38:09.280 for 02:42:57
Paths: (4 available, best #1)
  Not advertised to any peer
  Path #1: Received by speaker 0
  Not advertised to any peer
  40 60
    10.3.40.40 (metric 30) from 1.1.1.3 (1.1.1.3)
      Origin IGP, localpref 100, valid, internal, best, group-best
      Received Path ID 1, Local Path ID 1, version 8
  Path #2: Received by speaker 0
  Not advertised to any peer
  50 60
    10.3.50.50 (metric 30) from 1.1.1.3 (1.1.1.3)
      Origin IGP, localpref 100, valid, internal, group-best, add-path
      Received Path ID 2, Local Path ID 4, version 8
  Path #3: Received by speaker 0
  Not advertised to any peer
  40 60
    10.4.40.40 (metric 30) from 1.1.1.4 (1.1.1.4)
      Origin IGP, localpref 100, valid, internal, add-path
      Received Path ID 1, Local Path ID 2, version 8
  Path #4: Received by speaker 0
  Not advertised to any peer
  50 60
    10.4.50.50 (metric 30) from 1.1.1.4 (1.1.1.4)
      Origin IGP, localpref 100, valid, internal, add-path
      Received Path ID 2, Local Path ID 3, version 8
```

A ping and traceroute to the overlay prefix confirms that the traffic is directed to R3 (prefix SID 16003) and then to AS 40 (EPE label 24004 for the eBGP neighbor to AS 40 at R3).

```
RP/0/RP0/CPU0:R1# ping 161.1.1.1 source 151.1.1.1 count 10
Thu Feb  3 23:20:48.911 UTC
Type escape sequence to abort.
Sending 10, 100-byte ICMP Echos to 161.1.1.1, timeout is 2 seconds:
!!!!!!!!!!
Success rate is 100 percent (10/10), round-trip min/avg/max = 30/36/54 ms


RP/0/RP0/CPU0:R1# traceroute 161.1.1.1 source 151.1.1.1
Thu Feb  3 23:20:53.630 UTC

Type escape sequence to abort.
Tracing the route to 161.1.1.1

 1  10.1.2.2 [MPLS: Labels 16003/24004 Exp 0] 49 msec  45 msec  42 msec
 2  10.2.3.3 [MPLS: Label 24004 Exp 0] 42 msec  37 msec  37 msec
 3  10.3.40.40 44 msec  37 msec  41 msec
 4  10.40.60.60 47 msec *  55 msec
```

Now, we proceed to apply a BGP route-policy that would modify BGP best-path selection and choose instead the path from NH 10.4.50.50 (AS 50 via R4).

```
RP/0/RP0/CPU0:R1(config)# route-policy rpl_epe
RP/0/RP0/CPU0:R1(config-rpl)# if destination in (161.1.1.0/28) and next-hop in (10.4.50.50)
 then
RP/0/RP0/CPU0:R1(config-rpl-if)# set local-preference 1000
```

```
RP/0/RP0/CPU0:R1(config-rpl-if)# elseif destination in (161.1.1.1/32) and next-hop in
(10.4.40.40) then
RP/0/RP0/CPU0:R1(config-rpl-elseif)# set local-preference 1000
RP/0/RP0/CPU0:R1(config-rpl-elseif)# endif
RP/0/RP0/CPU0:R1(config-rpl)# pass
RP/0/RP0/CPU0:R1(config-rpl)# end-policy
RP/0/RP0/CPU0:R1(config)#


RP/0/RP0/CPU0:R1(config)# router bgp 100
RP/0/RP0/CPU0:R1(config-bgp)# neighbor 1.1.1.3
RP/0/RP0/CPU0:R1(config-bgp-nbr)# address-family ipv4 unicast
RP/0/RP0/CPU0:R1(config-bgp-nbr-af)# route-policy rpl_epe in
RP/0/RP0/CPU0:R1(config-bgp-nbr-af)# exit
RP/0/RP0/CPU0:R1(config-bgp-nbr)# exit
RP/0/RP0/CPU0:R1(config-bgp)# neighbor 1.1.1.4
RP/0/RP0/CPU0:R1(config-bgp-nbr)# address-family ipv4 unicast
RP/0/RP0/CPU0:R1(config-bgp-nbr-af)# route-policy rpl_epe in
RP/0/RP0/CPU0:R1(config-bgp-nbr-af)#
```

Observe the new BGP best-path selected for the overlay prefix via NH 10.4.50.50:

```
RP/0/RP0/CPU0:R1# show bgp

BGP router identifier 1.1.1.1, local AS number 100
BGP generic scan interval 60 secs
Non-stop routing is enabled
BGP table state: Active
Table ID: 0xe0000000   RD version: 20
BGP main routing table version 20
BGP NSR Initial initsync version 7 (Reached)
BGP NSR/ISSU Sync-Group versions 0/0
BGP scan interval 60 secs

Status codes: s suppressed, d damped, h history, * valid, > best
              i - internal, r RIB-failure, S stale, N Nexthop-discard
Origin codes: i - IGP, e - EGP, ? - incomplete
   Network            Next Hop            Metric LocPrf Weight Path
*>i10.3.40.40/32      1.1.1.3                  100      0 i
*>i10.3.50.50/32      1.1.1.3                  100      0 i
*>i10.4.40.40/32      1.1.1.4                  100      0 i
*>i10.4.50.50/32      1.1.1.4                  100      0 i
*> 151.1.1.1/32       0.0.0.0                  0         32768 i
* i161.1.1.0/28       10.3.40.40               100      0 40 60 i
* i                   10.3.50.50               100      0 50 60 i
* i                   10.4.40.40               100      0 40 60 i
*>i                   10.4.50.50              1000      0 50 60 i

Processed 6 prefixes, 9 paths


RP/0/RP0/CPU0:R1# show bgp ipv4 unicast 161.1.1.0/28

BGP routing table entry for 161.1.1.0/28
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker                20          20
Last Modified: Feb  3 23:13:30.539 for 00:01:33
Paths: (4 available, best #4)
  Not advertised to any peer
  Path #1: Received by speaker 0
  Not advertised to any peer
  40 60
    10.3.40.40 (metric 30) from 1.1.1.3 (1.1.1.3)
```

```
    Origin IGP, localpref 100, valid, internal, group-best, add-path
    Received Path ID 1, Local Path ID 4, version 20
Path #2: Received by speaker 0
Not advertised to any peer
50 60
  10.3.50.50 (metric 30) from 1.1.1.3 (1.1.1.3)
    Origin IGP, localpref 100, valid, internal, add-path
    Received Path ID 2, Local Path ID 3, version 8
Path #3: Received by speaker 0
Not advertised to any peer
40 60
  10.4.40.40 (metric 30) from 1.1.1.4 (1.1.1.4)
    Origin IGP, localpref 100, valid, internal, add-path
    Received Path ID 1, Local Path ID 2, version 8
Path #4: Received by speaker 0
Not advertised to any peer
50 60
  10.4.50.50 (metric 30) from 1.1.1.4 (1.1.1.4)
    Origin IGP, localpref 1000, valid, internal, best, group-best
    Received Path ID 2, Local Path ID 1, version 20
```

A ping and traceroute to the overlay prefix confirms that after the RPL policy is applied, the traffic is directed instead to R4 (prefix SID 16004) and then to AS 50 (EPE label 24005 for the eBGP neighbor to AS 50 at R4).

```
RP/0/RP0/CPU0:R1# ping 161.1.1.1 source 151.1.1.1 count 10
Thu Feb  3 23:17:43.812 UTC
Type escape sequence to abort.
Sending 10, 100-byte ICMP Echos to 161.1.1.1, timeout is 2 seconds:
!!!!!!!!!!
Success rate is 100 percent (10/10), round-trip min/avg/max = 30/35/50 ms


RP/0/RP0/CPU0:R1# traceroute 161.1.1.1 source 151.1.1.1
Thu Feb  3 23:18:01.656 UTC

Type escape sequence to abort.
Tracing the route to 161.1.1.1

 1  10.1.2.2 [MPLS: Labels 16004/24005 Exp 0] 50 msec  42 msec  45 msec
 2  10.2.4.4 [MPLS: Label 24005 Exp 0] 50 msec  42 msec  42 msec
 3  10.4.50.50 46 msec  44 msec  44 msec
 4  10.50.60.60 51 msec  *  54 msec
```

IP Lookup Fallback for BGP Peering (EPE) Segments

Table 8: Feature History Table

Feature Name	Release	Description
IP Lookup Fallback for BGP Peering (EPE) Segments	Release 7.3.3	BGP peering segments/SIDs are part of the Segment Routing Centralized BGP Egress Peer Engineering solution (BGP-EPE). A BGP-EPE-enabled border router allocates and programs BGP peering SIDs (EPE labels) to steer traffic over a specific external interface/BGP neighbor. This feature allows a BGP-EPE-enabled border router to pop the EPE label and forward traffic based on an IP-based lookup when a BGP neighbor fails. Traffic arriving with the EPE label assigned to a failed neighbor is forwarded based on a destination IP address lookup to allow traffic to be forwarded over a different directly connected external peer.

BGP peering segments/SIDs are part of the Segment Routing Centralized BGP Egress Peer Engineering solution (BGP-EPE), as described in IETF RFC 9087. A BGP-EPE-enabled border router allocates and programs BGP peering SIDs (EPE labels) to steer traffic over a specific external interface/BGP neighbor.

This feature allows a BGP-EPE-enabled border router to pop the EPE label and forward traffic based on an IP-based lookup when a BGP neighbor fails. Traffic arriving with the EPE label assigned to a failed neighbor is forwarded based on a destination IP address lookup to allow traffic to be forwarded over a different directly connected external peer.

Usage Guidelines and Limitations

The following usage guidelines and limitations apply for this feature:

- IP Lookup Fallback for BGP peering SIDs (EPE Peer-Node SIDs and Peer-Adjacencies SIDs) allocated dynamically or configured manually is supported.
- BGPv4 and BGPv6 EPE-enabled neighbors are supported
- Sub-second convergence is supported upon failure of EPE-enabled BGP neighbor with interface peering.
- Sub-second convergence is supported upon failure of EPE-enabled BGP neighbor with loopback peering over a single interface.
- Sub-second convergence is not supported upon failure of EPE-enabled BGP neighbor with loopback peering over more than one interface.
- IP Lookup Fallback for BGP Peer-Set SIDs is not supported

- MPLS egress path counters for BGP peering SIDs are not supported when IP Lookup Fallback is enabled

Enabling IP Lookup Fallback for BGP Peering (EPE) Segments

To guaranteed convergence, configure a route policy on the ingress border router to advertise all BGP paths. For example:

```
RP/0/RP0/CPU0:R1(config)# route-policy INSTALL_BACKUP
RP/0/RP0/CPU0:R1(config-rpl)# set path-selection all advertise
RP/0/RP0/CPU0:R1(config-rpl)# set path-selection backup 1 install multipath-protect
RP/0/RP0/CPU0:R1(config-rpl)# end-policy

RP/0/RP0/CPU0:R1(config)# router bgp 100
RP/0/RP0/CPU0:R1(config-bgp)# address-family ipv4 unicast
RP/0/RP0/CPU0:R1(config-bgp-af)# additional-paths selection route-policy INSTALL_BACKUP
RP/0/RP0/CPU0:R1(config-bgp-af)# exit
RP/0/RP0/CPU0:R1(config-bgp)# exit
RP/0/RP0/CPU0:R1(config)#
```

To enable IP lookup fallback for EPE segments, use the **epe backup enable** command in router BGP address family configuration mode.

To retain the local label of the primary path after reconvergence for the specified amount of time, use the **retain local-label** *minutes* command in router BGP address family configuration mode. The range of *minutes* is from 3 to 60.

The following example shows how to enable IP lookup fallback for EPE segments associated with BGPv4 EPE-enabled neighbors:

```
RP/0/RP0/CPU0:R3(config)# router bgp 100
RP/0/RP0/CPU0:R3(config-bgp)# address-family ipv4 unicast
RP/0/RP0/CPU0:R3(config-bgp-af)# epe backup enable
RP/0/RP0/CPU0:R3(config-bgp-af)# retain local-label 6
```

Running Config

```
router bgp 100
  address-family ipv4 unicast
   epe backup enable
   retain local-label 6
```

Verification

The following outputs display the forwarding entries for the EPE MPLS labels (24004 and 24005) at an egress border router **before** the IP Lookup Fallback for EPE feature is enabled. Observe that no backup is programmed.

```
RP/0/RP0/CPU0:R3# show mpls forwarding labels 24004 24005

Local  Outgoing    Prefix            Outgoing     Next Hop        Bytes
Label  Label       or ID             Interface                    Switched
------ ----------- ----------------- ------------ --------------- ------------
24004  Pop         No ID             Hu0/0/0/1    10.3.40.40      0
24005  Pop         No ID             Hu0/0/0/2    10.3.50.50      0

RP/0/RP0/CPU0:R3# show mpls forwarding labels 24004 24005 detail
```

```
Local  Outgoing    Prefix             Outgoing     Next Hop        Bytes
Label  Label       or ID              Interface                    Switched
------ ----------- ------------------ ------------ --------------- ------------
24004  Pop         No ID              Hu0/0/0/1    10.3.40.40      0
     Updated: Feb 10 18:38:16.116
     Path Flags: 0x6000 [  ]
     Version: 16, Priority: 3
     Label Stack (Top -> Bottom): { Imp-Null }
     NHID: 0x0, Encap-ID: N/A, Path idx: 0, Backup path idx: 0, Weight: 0
     MAC/Encaps: 0/0, MTU: 1500
     Outgoing Interface: HundredGigE0/0/0/1 (ifhandle 0x00000198)
     Packets Switched: 0

24005  Pop         No ID              Hu0/0/0/2    10.3.50.50      0
     Updated: Feb 10 18:38:16.116
     Path Flags: 0x6000 [  ]
     Version: 17, Priority: 3
     Label Stack (Top -> Bottom): { Imp-Null }
     NHID: 0x0, Encap-ID: N/A, Path idx: 0, Backup path idx: 0, Weight: 0
     MAC/Encaps: 0/0, MTU: 1500
     Outgoing Interface: HundredGigE0/0/0/2 (ifhandle 0x000001a0)
     Packets Switched: 0
```

The following output depicts the BGP table for an overlay prefix including its primary and backup path.

```
RP/0/RP0/CPU0:R3# show bgp ipv4 unicast 161.1.1.0/28

BGP routing table entry for 161.1.1.0/28
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker                 6           6
Last Modified: Feb  3 22:28:56.039 for 00:02:56
Paths: (2 available, best #1)
  Advertised IPv4 Unicast paths to update-groups (with more than one peer):
    0.4
  Advertised IPv4 Unicast paths to peers (in unique update groups):
    1.1.1.1
  Path #1: Received by speaker 0
  Advertised IPv4 Unicast paths to update-groups (with more than one peer):
    0.4
  Advertised IPv4 Unicast paths to peers (in unique update groups):
    1.1.1.1
  40 60
    10.3.40.40 from 10.3.40.40 (1.1.1.40)
      Origin IGP, localpref 100, valid, external, best, group-best
      Received Path ID 0, Local Path ID 1, version 5
      Origin-AS validity: (disabled)
  Path #2: Received by speaker 0
  Advertised IPv4 Unicast paths to peers (in unique update groups):
    1.1.1.1
  50 60
    10.3.50.50 from 10.3.50.50 (1.1.1.50)
      Origin IGP, localpref 100, valid, external, group-best, backup, add-path
      Received Path ID 0, Local Path ID 2, version 6
      Origin-AS validity: (disabled)


RP/0/RP0/CPU0:R3# show cef 161.1.1.0/28 detail

161.1.1.0/28, version 26, internal 0x5000001 0x40 (ptr 0x90cd2920) [1], 0x0 (0x0), 0x0 (0x0)

 Updated Feb 17 20:35:32.438
 Prefix Len 28, traffic index 0, precedence n/a, priority 4
  gateway array (0x90aa9a58) reference count 1, flags 0x102010, source rib (7), 0 backups
```

```
              [1 type 3 flags 0x48441 (0x90b5a148) ext 0x0 (0x0)]
  LW-LDI[type=0, refc=0, ptr=0x0, sh-ldi=0x0]
  gateway array update type-time 1 Feb 17 20:35:32.438
 LDI Update time Feb 17 20:35:32.438

  Level 1 - Load distribution: 0
  [0] via 10.3.40.40/32, recursive

   via 10.3.40.40/32, 3 dependencies, recursive, bgp-ext [flags 0x6020]
    path-idx 0 NHID 0x0 [0x90cd3250 0x0], Internal 0x9081e550
    next hop 10.3.40.40/32 via 10.3.40.40/32

    Load distribution: 0 (refcount 1)

    Hash  OK  Interface                 Address
    0     Y   HundredGigE0/0/0/1        10.3.40.40

   via 10.3.50.50/32, 2 dependencies, recursive, bgp-ext, backup [flags 0x6120]
    path-idx 1 NHID 0x0 [0x90cd3178 0x0]
    next hop 10.3.50.50/32 via 10.3.50.50/32
```

The following outputs display the forwarding entries for the EPE MPLS labels at an egress border router **after** the IP Lookup Fallback for EPE feature is enabled. Observe that a backup path is now programmed for an EPE local label.

```
RP/0/RP0/CPU0:R3# show mpls forwarding labels 24004 24005
Thu Feb 10 18:40:34.655 UTC
Local  Outgoing    Prefix             Outgoing     Next Hop        Bytes
Label  Label       or ID              Interface                    Switched
------ ----------- ------------------ ------------ --------------- ------------
24004  Unlabelled  No ID              Hu0/0/0/1    10.3.40.40      0
       Aggregate   No ID              default                      0            (!)
24005  Unlabelled  No ID              Hu0/0/0/2    10.3.50.50      0
       Aggregate   No ID              default                      0            (!)


RP/0/RP0/CPU0:R3# show mpls forwarding labels 24004 24005 detail

Local  Outgoing    Prefix             Outgoing     Next Hop        Bytes
Label  Label       or ID              Interface                    Switched
------ ----------- ------------------ ------------ --------------- ------------
24004  Unlabelled  No ID              Hu0/0/0/1    10.3.40.40      0
     Updated: Feb 10 18:40:25.476
     Path Flags: 0x6000 [  ]
     Version: 18, Priority: 3
     Label Stack (Top -> Bottom): { Unlabelled }
     NHID: 0x0, Encap-ID: N/A, Path idx: 0, Backup path idx: 1, Weight: 0
     MAC/Encaps: 14/14, MTU: 1500
     Outgoing Interface: HundredGigE0/0/0/1 (ifhandle 0x00000198)
     Packets Switched: 0

       Aggregate   No ID              default                      0            (!)
     Updated: Feb 10 18:40:25.476
     Path Flags: 0x100 [  BKUP, NoFwd ]
     Label Stack (Top -> Bottom): { }
     MAC/Encaps: 0/0, MTU: 0
     Packets Switched: 0
24005  Unlabelled  No ID              Hu0/0/0/2    10.3.50.50      0
     Updated: Feb 10 18:40:25.482
     Path Flags: 0x6000 [  ]
     Version: 19, Priority: 3
     Label Stack (Top -> Bottom): { Unlabelled }
```

```
     NHID: 0x0, Encap-ID: N/A, Path idx: 0, Backup path idx: 1, Weight: 0
     MAC/Encaps: 14/14, MTU: 1500
     Outgoing Interface: HundredGigE0/0/0/2 (ifhandle 0x000001a0)
     Packets Switched: 0

       Aggregate   No ID              default                       0            (!)
     Updated: Feb 10 18:40:25.482
     Path Flags: 0x100 [  BKUP, NoFwd ]
     Label Stack (Top -> Bottom): { }
     MAC/Encaps: 0/0, MTU: 0
     Packets Switched: 0


RP/0/RP0/CPU0:R3# show bgp ipv4 unicast 161.1.1.0/28
Thu Feb  3 22:31:52.893 UTC
BGP routing table entry for 161.1.1.0/28
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker                 6           6
Last Modified: Feb  3 22:28:56.039 for 00:02:56
Paths: (2 available, best #1)
  Advertised IPv4 Unicast paths to update-groups (with more than one peer):
    0.4
  Advertised IPv4 Unicast paths to peers (in unique update groups):
    1.1.1.1
  Path #1: Received by speaker 0
  Advertised IPv4 Unicast paths to update-groups (with more than one peer):
    0.4
  Advertised IPv4 Unicast paths to peers (in unique update groups):
    1.1.1.1
  40 60
    10.3.40.40 from 10.3.40.40 (1.1.1.40)
      Origin IGP, localpref 100, valid, external, best, group-best
      Received Path ID 0, Local Path ID 1, version 5
      Origin-AS validity: (disabled)
  Path #2: Received by speaker 0
  Advertised IPv4 Unicast paths to peers (in unique update groups):
    1.1.1.1
  50 60
    10.3.50.50 from 10.3.50.50 (1.1.1.50)
      Origin IGP, localpref 100, valid, external, group-best, backup, add-path
      Received Path ID 0, Local Path ID 2, version 6
      Origin-AS validity: (disabled)


RP/0/RP0/CPU0:R3# show cef 161.1.1.0/28 detail

161.1.1.0/28, version 24, internal 0x5000001 0x40 (ptr 0x90684920) [1], 0x0 (0x0), 0x0 (0x0)

 Updated Feb 10 17:37:16.610
 Prefix Len 28, traffic index 0, precedence n/a, priority 4
  gateway array (0x9045ba58) reference count 1, flags 0x102010, source rib (7), 0 backups
                [1 type 3 flags 0x48441 (0x9050c148) ext 0x0 (0x0)]
  LW-LDI[type=0, refc=0, ptr=0x0, sh-ldi=0x0]
  gateway array update type-time 1 Feb 10 17:37:16.610
 LDI Update time Feb 10 17:37:16.623

  Level 1 - Load distribution: 0
  [0] via 10.3.40.40/32, recursive

   via 10.3.40.40/32, 5 dependencies, recursive, bgp-ext [flags 0x6020]
    path-idx 0 NHID 0x0 [0x90684c08 0x0], Internal 0x90211730
    next hop 10.3.40.40/32 via 10.3.40.40/32

    Load distribution: 0 (refcount 1)
```

```
 Hash  OK  Interface                 Address
 0     Y   HundredGigE0/0/0/1        10.3.40.40

via 10.3.50.50/32, 4 dependencies, recursive, bgp-ext, backup [flags 0x6120]
 path-idx 1 NHID 0x0 [0x90685040 0x0]
 next hop 10.3.50.50/32 via 10.3.50.50/32
```

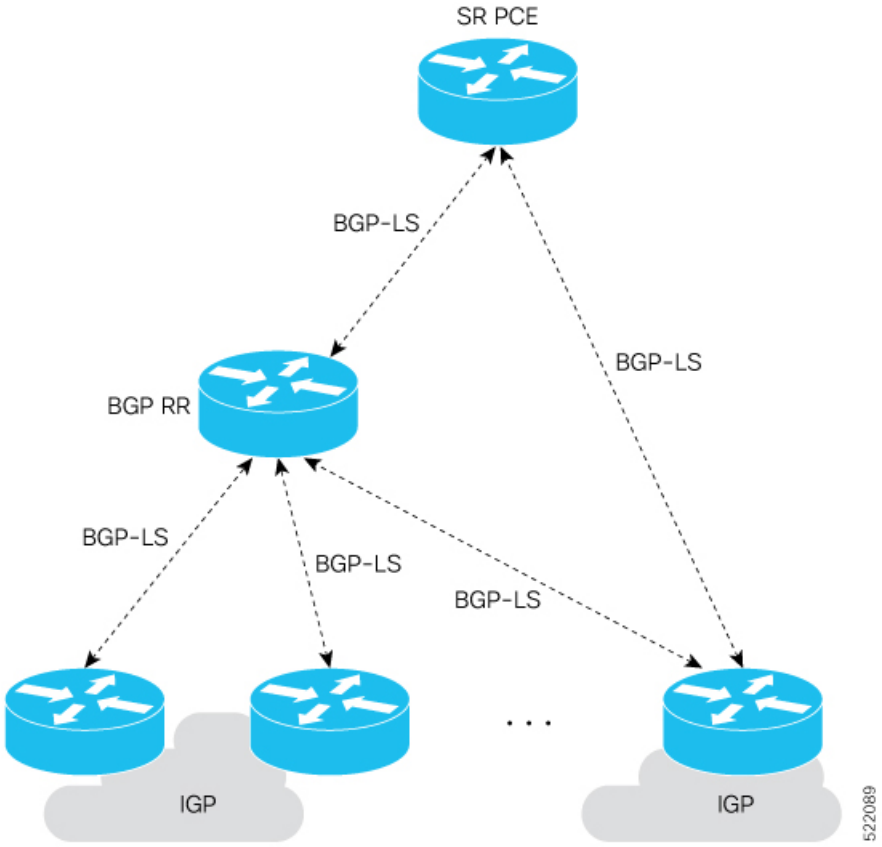
Configure BGP Link-State

BGP Link-State (LS) is an Address Family Identifier (AFI) and Sub-address Family Identifier (SAFI) originally defined to carry interior gateway protocol (IGP) link-state information through BGP. The BGP Network Layer Reachability Information (NLRI) encoding format for BGP-LS and a new BGP Path Attribute called the BGP-LS attribute are defined in RFC7752. The identifying key of each Link-State object, namely a node, link, or prefix, is encoded in the NLRI and the properties of the object are encoded in the BGP-LS attribute.

The BGP-LS Extensions for Segment Routing are documented in RFC9085.

BGP-LS applications like an SR Path Computation Engine (SR-PCE) can learn the SR capabilities of the nodes in the topology and the mapping of SR segments to those nodes. This can enable the SR-PCE to perform path computations based on SR-TE and to steer traffic on paths different from the underlying IGP-based distributed best-path computation.

The following figure shows a typical deployment scenario. In each IGP area, one or more nodes (BGP speakers) are configured with BGP-LS. These BGP speakers form an iBGP mesh by connecting to one or more route-reflectors. This way, all BGP speakers (specifically the route-reflectors) obtain Link-State information from all IGP areas (and from other ASes from eBGP peers).



Usage Guidelines and Limitations

- BGP-LS supports IS-IS and OSPFv2.
- The identifier field of BGP-LS (referred to as the Instance-ID) identifies the IGP routing domain where the NLRI belongs. The NLRIs representing link-state objects (nodes, links, or prefixes) from the same IGP routing instance must use the same Instance-ID value.
- When there is only a single protocol instance in the network where BGP-LS is operational, we recommend configuring the Instance-ID value to **0**.
- Assign consistent BGP-LS Instance-ID values on all BGP-LS Producers within a given IGP domain.
- NLRIs with different Instance-ID values are considered to be from different IGP routing instances.
- Unique Instance-ID values must be assigned to routing protocol instances operating in different IGP domains. This allows the BGP-LS Consumer (for example, SR-PCE) to build an accurate segregated multi-domain topology based on the Instance-ID values, even when the topology is advertised via BGP-LS by multiple BGP-LS Producers in the network.
- If the BGP-LS Instance-ID configuration guidelines are not followed, a BGP-LS Consumer may see duplicate link-state objects for the same node, link, or prefix when there are multiple BGP-LS Producers deployed. This may also result in the BGP-LS Consumers getting an inaccurate network-wide topology.

- The following table defines the supported extensions to the BGP-LS address family for carrying IGP topology information (including SR information) via BGP. For more information on the BGP-LS TLVs, refer to Border Gateway Protocol - Link State (BGP-LS) Parameters.

Table 9: IOS XR Supported BGP-LS Node Descriptor, Link Descriptor, Prefix Descriptor, and Attribute TLVs

TLV Code Point	Description	Produced by IS-IS	Produced by OSPFv2	Produced by BGP
256	Local Node Descriptors	X	X	—
257	Remote Node Descriptors	X	X	—
258	Link Local/Remote Identifiers	X	X	—
259	IPv4 interface address	X	X	—
260	IPv4 neighbor address	X		
261	IPv6 interface address	X	—	—
262	IPv6 neighbor address	X	—	—
263	Multi-Topology ID	X	—	—
264	OSPF Route Type	—	X	—
265	IP Reachability Information	X	X	—
266	Node MSD TLV	X	X	—
267	Link MSD TLV	X	X	—
512	Autonomous System	—	—	X
513	BGP-LS Identifier	—	—	X
514	OSPF Area-ID	—	X	—
515	IGP Router-ID	X	X	—
516	BGP Router-ID TLV	—	—	X
517	BGP Confederation Member TLV	—	—	X
1024	Node Flag Bits	X	X	—
1026	Node Name	X	X	—
1027	IS-IS Area Identifier	X	—	—
1028	IPv4 Router-ID of Local Node	X	X	—
1029	IPv6 Router-ID of Local Node	X	—	—
1030	IPv4 Router-ID of Remote Node	X	X	—
1031	IPv6 Router-ID of Remote Node	X	—	—
1034	SR Capabilities TLV	X	X	—
1035	SR Algorithm TLV	X	X	—
1036	SR Local Block TLV	X	X	—

TLV Code Point	Description	Produced by IS-IS	Produced by OSPFv2	Produced by BGP
1039	Flex Algo Definition (FAD) TLV	X	X	—
1044	Flex Algorithm Prefix Metric (FAPM) TLV	X	X	—
1088	Administrative group (color)	X	X	—
1089	Maximum link bandwidth	X	X	—
1090	Max. reservable link bandwidth	X	X	—
1091	Unreserved bandwidth	X	X	—
1092	TE Default Metric	X	X	—
1093	Link Protection Type	X	X	—
1094	MPLS Protocol Mask	X	X	—
1095	IGP Metric	X	X	—
1096	Shared Risk Link Group	X	X	—
1099	Adjacency SID TLV	X	X	—
1100	LAN Adjacency SID TLV	X	X	—
1101	PeerNode SID TLV	—	—	X
1102	PeerAdj SID TLV	—	—	X
1103	PeerSet SID TLV	—	—	X
1114	Unidirectional Link Delay TLV	X	X	—
1115	Min/Max Unidirectional Link Delay TLV	X	X	—
1116	Unidirectional Delay Variation TLV	X	X	—
1117	Unidirectional Link Loss	X	X	—
1118	Unidirectional Residual Bandwidth	X	X	—
1119	Unidirectional Available Bandwidth	X	X	—
1120	Unidirectional Utilized Bandwidth	X	X	—
1122	Application-Specific Link Attribute TLV	X	X	—
1152	IGP Flags	X	X	—
1153	IGP Route Tag	X	X	—
1154	IGP Extended Route Tag	X	—	—
1155	Prefix Metric	X	X	—
1156	OSPF Forwarding Address	—	X	—
1158	Prefix-SID	X	X	—
1159	Range	X	X	—

TLV Code Point	Description	Produced by IS-IS	Produced by OSPFv2	Produced by BGP
1161	SID/Label TLV	X	X	—
1170	Prefix Attribute Flags	X	X	—
1171	Source Router Identifier	X	—	—
1172	L2 Bundle Member Attributes TLV	X	—	—
1173	Extended Administrative Group	X	X	—

Exchange Link State Information with BGP Neighbor

The following example shows how to exchange link-state information with a BGP neighbor:

```
Router# configure
Router(config)# router bgp 1
Router(config-bgp)# neighbor 10.0.0.2
Router(config-bgp-nbr)# remote-as 1
Router(config-bgp-nbr)# address-family link-state link-state
Router(config-bgp-nbr-af)# exit
```

IGP Link-State Database Distribution

A given BGP node may have connections to multiple, independent routing domains. IGP link-state database distribution into BGP-LS is supported for both OSPF and IS-IS protocols in order to distribute this information on to controllers or applications that desire to build paths spanning or including these multiple domains.

To distribute IS-IS link-state data using BGP-LS, use the **distribute link-state** command in router configuration mode.

```
Router# configure
Router(config)# router isis isp
Router(config-isis)# distribute link-state instance-id 32
```

To distribute OSPFv2 link-state data using BGP-LS, use the **distribute link-state** command in router configuration mode.

```
Router# configure
Router(config)# router ospf 100
Router(config-ospf)# distribute link-state instance-id 32
```

Configure BGP Proxy Prefix SID

Table 10: Feature History Table

Feature Name	Release	Description
BGP Proxy Prefix SID	Release 7.3.2	This feature is a BGP extension to signal BGP prefix-SIDs. This feature allows you to attach BGP prefix SID attributes for remote prefixes learned over BGP labeled unicast (LU) sessions and propagate them as SR prefixes using BGP LU. This allows an LSP towards non-SR endpoints to use segment routing global block in the SR domain.

To support segment routing, Border Gateway Protocol (BGP) requires the ability to advertise a segment identifier (SID) for a BGP prefix. A BGP-Prefix-SID is the segment identifier of the BGP prefix segment in a segment routing network. BGP prefix SID attribute is a BGP extension to signal BGP prefix-SIDs. However, there may be routers which do not support BGP extension for segment routing. Hence, those routers also do not support BGP prefix SID attribute and an alternate approach is required.

BGP proxy prefix SID feature allows you to attach BGP prefix SID attributes for remote prefixes learnt from BGP labeled unicast (LU) neighbours which are not SR-capable and propagate them as SR prefixes. This allows an LSP towards non SR endpoints to use segment routing global block in a SR domain. Since BGP proxy prefix SID uses global label values it minimizes the use of limited resources such as ECMP-FEC and provides more scalability for the networks.

BGP proxy prefix SID feature is implemented using the segment routing mapping server (SRMS). SRMS allows the user to configure SID mapping entries to specify the prefix-SIDs for the prefixes. The mapping server advertises the local SID-mapping policy to the mapping clients. BGP acts as a client of the SRMS and uses the mapping policy to calculate the prefix-SIDs.

Configuration Example:

This example shows how to configure the BGP proxy prefix SID feature for the segment routing mapping server.

```
RP/0/RSP0/CPU0:router(config)# segment-routing
RP/0/RSP0/CPU0:router(config-sr)# mapping-server
RP/0/RSP0/CPU0:router(config-sr-ms)# prefix-sid-map
RP/0/RSP0/CPU0:router(config-sr-ms-map)# address-family ipv4
RP/0/RSP0/CPU0:router(config-sr-ms-map-af)# 1.1.1.1/32 10 range 200
RP/0/RSP0/CPU0:router(config-sr-ms-map-af)# 192.168.64.1/32 400 range 300
```

This example shows how to configure the BGP proxy prefix SID feature for the segment-routing mapping client.

```
RP/0/RSP0/CPU0:router(config)# router bgp 1
RP/0/RSP0/CPU0:router(config-bgp)# address-family ip4 unicast
RP/0/RSP0/CPU0:router(config-bgp-af)# segment-routing prefix-sid-map
```

Verification

These examples show how to verify the BGP proxy prefix SID feature.

```
RP/0/RSP0/CPU0:router# show segment-routing mapping-server prefix-sid-map ipv4
 detail
Prefix
1.1.1.1/32
    SID Index:      10
    Range:          200
    Last Prefix:    1.1.1.200/32
    Last SID Index: 209
    Flags:
Number of mapping entries: 1


RP/0/RSP0/CPU0:router# show bgp ipv4 labeled-unicast 192.168.64.1/32

BGP routing table entry for 192.168.64.1/32
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker               117         117
  Local Label: 16400
Last Modified: Oct 25 01:02:28.562 for 00:11:45Paths: (2 available, best #1)
 Advertised to peers (in unique update groups):
   201.1.1.1
 Path #1: Received by speaker 0  Advertised to peers (in unique update groups):
    201.1.1.1
  Local
   20.0.101.1 from 20.0.101.1 (20.0.101.1)      Received Label 61
   Origin IGP, localpref 100, valid, internal, best, group-best, multipath, labeled-unicast

   Received Path ID 0, Local Path ID 0, version 117
  Prefix SID Attribute Size: 7
  Label Index: 1

 RP/0/RSP0/CPU0:router# show route ipv4 unicast 192.68.64.1/32 detail

Routing entry for 192.168.64.1/32
  Known via "bgp 65000", distance 200, metric 0, [ei]-bgp, labeled SR, type internal
  Installed Oct 25 01:02:28.583 for 00:20:09
  Routing Descriptor Blocks
    20.0.101.1, from 20.0.101.1, BGP multi path
      Route metric is 0
      Label: 0x3d (61)
      Tunnel ID: None
      Binding Label: None
      Extended communities count: 0
      NHID:0x0(Ref:0)
   Route version is 0x6 (6)
  Local Label: 0x3e81 (16400)
  IP Precedence: Not Set
  QoS Group ID: Not Set
  Flow-tag: Not Set
  Fwd-class: Not Set
  Route Priority: RIB_PRIORITY_RECURSIVE (12) SVD Type RIB_SVD_TYPE_LOCAL
  Download Priority 4, Download Version 242
  No advertising protos.


RP/0/RSP0/CPU0:router# show cef ipv4 192.168.64.1/32 detail
192.168.64.1/32, version 476, labeled SR, drop adjacency, internal 0x5000001 0x80 (ptr
0x71c42b40) [1], 0x0 (0x71c11590), 0x808 (0x722b91e0)
 Updated Oct 31 23:23:48.733
```

```
 Prefix Len 32, traffic index 0, precedence n/a, priority 4
 Extensions: context-label:16400
  gateway array (0x71ae7e78) reference count 3, flags 0x7a, source rib (7), 0 backups
                [2 type 5 flags 0x88401 (0x722eb450) ext 0x0 (0x0)]
  LW-LDI[type=5, refc=3, ptr=0x71c11590, sh-ldi=0x722eb450]
  gateway array update type-time 3 Oct 31 23:49:11.720
 LDI Update time Oct 31 23:23:48.733
 LW-LDI-TS Oct 31 23:23:48.733
   via 20.0.101.1/32, 0 dependencies, recursive, bgp-ext [flags 0x6020]
    path-idx 0 NHID 0x0 [0x7129a294 0x0]
    recursion-via-/32
    unresolved
     local label 16400
     labels imposed {ExpNullv6}
```

```
RP/0/RSP0/CPU0:router# show bgp labels
BGP router identifier 2.1.1.1, local AS number 65000
BGP generic scan interval 60 secs
Non-stop routing is enabled
BGP table state: Active
Table ID: 0xe0000000   RD version: 245
BGP main routing table version 245
BGP NSR Initial initsync version 16 (Reached)
BGP NSR/ISSU Sync-Group versions 245/0
BGP scan interval 60 secs

Status codes: s suppressed, d damped, h history, * valid, > best
              i - internal, r RIB-failure, S stale, N Nexthop-discard
Origin codes: i - IGP, e - EGP, ? - incomplete
   Network            Next Hop        Rcvd Label      Local Label
*>i1.1.1.1/32         1.1.1.1         3               16010
*> 2.1.1.1/32         0.0.0.0         nolabel         3
*> 192.68.64.1/32      20.0.101.1      2               16400
*> 192.68.64.2/32      20.0.101.1      2               16401
```

BGP Best Path Computation using SR Policy Paths

Table 11: Feature History Table

Feature Name	Release Information	Feature Description
BGP Best Path Computation using SR Policy Paths	Release 7.3.4	BGP best-path selection is modified for a prefix when at least one of its paths resolves over the next hop using SR policies (SR policy in "up" state). Under this condition, paths not steered over an SR policy (those using native next-hop resolution) are considered ineligible during best-path selection. You can thus control the best path selection in order to steer traffic, preferably or exclusively, over SR policies with the desired SLA. This feature introduces the **bgp bestpath sr-policy {force | prefer}** command.

BGP selects the best path from the available pool of paths such as iBGP, eBGP, color, or noncolor paths with native next hop and SR policy next hop. BGP uses either native next hop or an SR policy next hop for best path computation. However, BGP might not consider SR policy next hop for best path computation due to other factors in best path selection. By default, BGP considers a native next hop for the best path computation during the failure.

For more information, see Best path calculation algorithm.

When multiple advertisements of the same BGP prefix are received where some have extended community color, SRTE headend with BGP multi-path enabled installs multiple routes with or without extended community color. It may be required to exclude the path resolving over native next hop SR policy paths from BGP best path selection when a prefix has multiple paths in the presence of one BGP path with the extended community color that is resolved over the SR policy.

You may want to use the egress PE to exit a domain using local preference or other attributes before the next hop metric selection. In such scenarios, when SR policy of the primary path fails, the best path is resolved over a regular IGP next hop that is the default mode of operation. Traffic doesn't select the backup path with SR policy, instead traffic moves to native LSP on the primary path.

The BGP Best Path Computation using SR Policy Paths feature allows the BGP to use the path with SR policy as the best-path, backup, and multipath.

When this feature is enabled, some paths are marked as an ineligible path for BGP best path selection. Existing BGP best path selection order is applied to the eligible paths.

Use either of the following modes for the BGP to select the SR policy path as the best path for the backup path:

- Force mode: When force mode is enabled, only SR policy paths are considered for best path calculation. Use the **bgp bestpath sr-policy force** command to enable this mode.

  In a network, when at least one path has an active SR policy, the following paths are marked as ineligible for best path selection:

  - iBGP paths with noncolor or color paths with SR policy that isn't active.
  - eBGP with color and SR policy isn't active.
  - eBGP noncolor paths

Note Local and redistributed BGP paths are always eligible for best path selection.

- Prefer mode: When prefer mode is enabled, SR policy paths and eBGP noncolor paths are eligible for best path calculation.

  Use the **bgp bestpath sr-policy prefer** command to enable this mode.

  In a network, when at least one path has an active SR policy, the following paths are marked as ineligible for best path selection:

  - iBGP paths with noncolor or color paths with SR policy that isn't active.
  - eBGP with color and SR policy isn't active.

Note Local and redistributed BGP paths are always eligible for best path selection.

Configure BGP Best Path Computation using SR Policy Paths

To enable the feature, perform the following tasks on the ingress PE router that is the head-end of SR policy:

- Configure route policy.
- Configure SR policy.
- Configure BGP with either prefer or force mode.

Configuration Example

Configure route policies on the egress PE router:

```
Router(config)#extcommunity-set opaque color9001
Router(config-ext)#9001 co-flag 01
Router(config-ext)#end-set
Router(config)#extcommunity-set opaque color9002
Router(config-ext)#9002 co-flag 01
Router(config-ext)#end-set
Router(config)#commitC

Router(config)#route-policy for9001
Router(config-rpl)#set extcommunity color color9001
```

```
Router(config-rpl)# pass
Router(config-rpl)#end-policy

Router(config)#route-policy for9002
Router(config-rpl)#set extcommunity color color9002
Router(config-rpl)#pass
Router(config-rpl)#end-policy
Router(config)#commit


Router#configure
Router(config)#route-policy add_path
Router(config-rpl)#set path-selection backup 1 install multipath-protect advertise
multipath-protect-advertise
Router(config-rpl)#end-policy

Router(config)#route-policy pass-all
Router(config-rpl)#pass
Router(config-rpl)#end-policy
Router(config)#commit
```

Configure SR policy on the egress PE router:

```
Router#configure
Router(config)#segment-routing
Router(config-sr)#traffic-eng
Router(config-sr-te)#segment-list SL201
Router(config-sr-te-sl)#index 1 mpls label 25000
Router(config-sr-te-sl)#policy POLICY_9001
Router(config-sr-te-policy)#binding-sid mpls 47700
Router(config-sr-te-policy)#color 9001 end-point ipv6 ::
Router(config-sr-te-policy)#candidate-paths
Router(config-sr-te-policy-path)#preference 10
Router(config-sr-te-policy-path-pref)#explicit segment-list SL201
Router(config-sr-te-sl)#policy POLICY_9002
Router(config-sr-te-policy)#binding-sid mpls 47701
Router(config-sr-te-policy)#color 9002 end-point ipv6 ::
Router(config-sr-te-policy)#candidate-paths
Router(config-sr-te-policy-path)#preference 10
Router(config-sr-te-policy-path-pref)#explicit segment-list SL201
Router(config-sr-te-policy-path-pref)#commit
```

Configure BGP on the Egress PE router:

```
Router(config)#router bgp 100
Router(config-bgp)#nsr
Router(config-bgp)#bgp router-id 10.1.1.2
Router(config-bgp)#bgp best-path sr-policy force
Router(config-bgp)#address-family ipv6 unicast
Router(config-bgp-af)#maximum-paths eibgp 25
Router(config-bgp-af)#additional-paths receive
Router(config-bgp-af)#additional-paths send
Router(config-bgp-af)#additional-paths selection route-policy add_path
Router(config-bgp-af)#redistribute connected
Router(config-bgp-af)#redistribute static
Router(config-bgp-af)#allocate-label all
Router(config-bgp-af)#commit
Router(config-bgp-af)#exit
Router(config-bgp)#neighbor 31::2
Router(config-bgp-nbr)#remote-as 2
Router(config-bgp-nbr)#address-family ipv6 unicast
Router(config-bgp-nbr-af)#route-policy for9001 in
Router(config-bgp-nbr-af)#route-policy pass-all out
```

```
Router(config-bgp-nbr-af)#commit
Router(config-bgp-nbr-af)#exit
Router(config-bgp)#neighbor 32::2
Router(config-bgp-nbr)#remote-as 2
Router(config-bgp-nbr)#address-family ipv6 unicast
Router(config-bgp-nbr-af)#route-policy for9002 in
Router(config-bgp-nbr-af)#route-policy pass-all out
Router(config-bgp-nbr-af)#commit
```

Verification

The following show output shows that when the **force** option is enabled, the configured SR policy path is selected as the best path instead of the default best path.

```
Router#show bgp ipv6 unicast 2001:DB8::1 brief
Status codes: s suppressed, d damped, h history, * valid, > best
              i - internal, r RIB-failure, S stale, N Nexthop-discard
Origin codes: i - IGP, e - EGP, ? - incomplete
   Network            Next Hop            Metric LocPrf Weight Path
* 2001:DB8::1         10:1:1::55                   100      0 2 i
* i                   10:1:1::55                   100      0 2 i

*                    30::2                                 0 2 I
*>                   31::2 C:9001                          0 2 I
*                    32::2 C:9002                          0 2 I
Router#
```

Use the following command to compare the best paths:

```
Router#show bgp ipv6 unicast 2001:DB8::1 bestpath-compare
BGP routing table entry for 2001:DB8::1
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker              7641        7641
    Flags: 0x240232b2+0x20050000; multipath; backup available;
Last Modified: Dec  7 03:43:57.200 for 00:34:48
Paths: (24 available, best #4)
  Advertised IPv6 Unicast paths to update-groups (with more than one peer):
    0.3 0.4
  Advertised IPv6 Unicast paths to peers (in unique update groups):
    10.1.1.55
  Path #1: Received by speaker 0
  Flags: 0x2000000000020005, import: 0x20
  Flags2: 0x00
  Not advertised to any peer
  2
    10:1:1::55 (metric 30) from 10.1.1.55 (10.1.1.55), if-handle 0x00000000
      Origin IGP, localpref 100, valid, internal
      Received Path ID 1, Local Path ID 0, version 0
      Extended community: Color[CO-Flag]:8001[01]
      Non SR-policy path is ignored due to config knob
  Path #2: Received by speaker 0
  Flags: 0x2000000000020005, import: 0x20
  Flags2: 0x00
  Not advertised to any peer
  2
    10:1:1::55 (metric 30) from 10.1.1.55 (10.1.1.55), if-handle 0x00000000
      Origin IGP, localpref 100, valid, internal
      Received Path ID 3, Local Path ID 0, version 0
      Extended community: Color[CO-Flag]:8002[01]
      Non SR-policy path is ignored due to config knob
  Path #3: Received by speaker 0
  Flags: 0x3000000000060001, import: 0x20
```

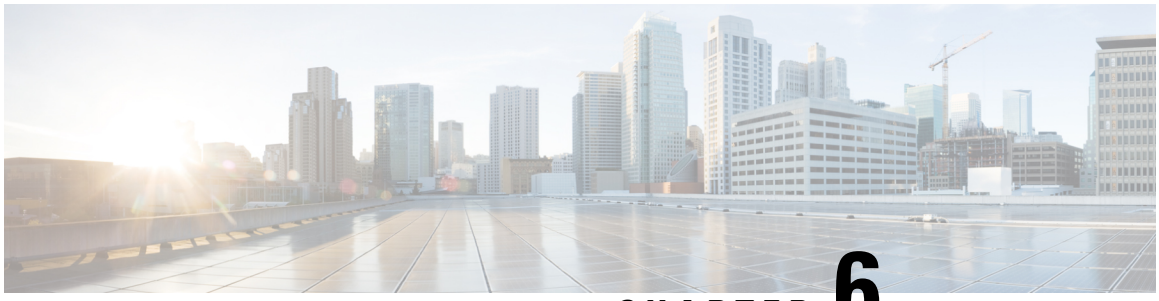
```
Flags2: 0x00
Advertised IPv6 Unicast paths to update-groups (with more than one peer):
  0.4
Advertised IPv6 Unicast paths to peers (in unique update groups):
  10.1.1.55
2
  30::2 from 30::2 (198.51.100.1), if-handle 0x00000000
    Origin IGP, localpref 100, weight 65534, valid, external, backup, add-path
    Received Path ID 0, Local Path ID 2, version 7641
    Origin-AS validity: (disabled)
    Non SR-policy path is ignored due to config knob
Path #4: Received by speaker 0
Flags: 0xb000000001070001, import: 0x20
Flags2: 0x00
Advertised IPv6 Unicast paths to update-groups (with more than one peer):
  0.3 0.4
Advertised IPv6 Unicast paths to peers (in unique update groups):
  10.1.1.55
2
  31::2 C:9001 (bsid:48900) from 31::2 (198.51.100.2), if-handle 0x00000000
    Origin IGP, localpref 100, valid, external, best, group-best, multipath
    Received Path ID 0, Local Path ID 1, version 7641
    Extended community: Color[CO-Flag]:9001[01]
    Origin-AS validity: (disabled)
    SR policy color 9001, ipv6 null endpoint, up, not-registered, bsid 48900

    best of AS 2, Overall best
Path #5: Received by speaker 0
Flags: 0xb000000000030001, import: 0x20
Flags2: 0x00
Not advertised to any peer
2
  32::2 C:9002 (bsid:48901) from 32::2 (198.51.100.3), if-handle 0x00000000
    Origin IGP, localpref 100, valid, external, multipath
    Received Path ID 0, Local Path ID 0, version 0
    Extended community: Color[CO-Flag]:9002[01]
    Origin-AS validity: (disabled)
    SR policy color 9002, up, not-registered, bsid 48901
    Higher router ID than best path (path #4)
```

Use the **show bgp process** command to verify which mode is enabled.

In the following example, you see that the **force** mode is enabled.

```
Router#show bgp process
BGP Process Information:
BGP is operating in STANDALONE mode
Autonomous System number format: ASPLAIN
Autonomous System: 100
Router ID: 10.1.1.2 (manually configured)
Default Cluster ID: 10.1.1.2
Active Cluster IDs:  10.1.1.2
Fast external fallover enabled
Platform Loadbalance paths max: 64
Platform RLIMIT max: 8589934592 bytes
Maximum limit for BMP buffer size: 1638 MB
Default value for BMP buffer size: 1228 MB
Current limit for BMP buffer size: 1228 MB
Current utilization of BMP buffer limit: 0 B
Neighbor logging is enabled
Enforce first AS enabled
Use SR-Policy admin/metric of color-extcomm Nexthop during path comparison: disabled
SR policy path force is enabled
Default local preference: 100
```

```
Default keepalive: 60
Non-stop routing is enabled
Slow peer detection enabled
ExtComm Color Nexthop validation: RIB

Update delay: 120
Generic scan interval: 60
Configured Segment-routing Local Block: [0, 0]
In use Segment-routing Local Block: [15000, 15999]
Platform support mix of sr-policy and native nexthop: Yes

Address family: IPv4 Unicast
Dampening is not enabled
Client reflection is enabled in global config
Dynamic MED is Disabled
Dynamic MED interval : 10 minutes
Dynamic MED Timer : Not Running
Dynamic MED Periodic Timer : Not Running
Scan interval: 60
Total prefixes scanned: 33
Prefixes scanned per segment: 100000
Number of scan segments: 1
Nexthop resolution minimum prefix-length: 0 (not configured)
IPv6 Nexthop resolution minimum prefix-length: 0 (not configured)
Main Table Version: 12642
Table version synced to RIB: 12642
Table version acked by RIB: 12642
IGP notification: IGPs notified
RIB has converged: version 2
RIB table prefix-limit reached ?  [No], version 0
Permanent Network Unconfigured

Node                Process     Nbrs Estb Rst Upd-Rcvd Upd-Sent Nfn-Rcv Nfn-Snt
node0_RSP1_CPU0     Speaker       53    3   2      316      823       0      53
```



CHAPTER 6

Enabling Segment Routing Flexible Algorithm

Table 12: Feature History Table

Feature Name	Release Information	Feature Description
Segment Routing Flexible Algorithm	Release 7.3.1	The Segment Routing architecture associates prefix-SIDs to an algorithm that defines how the path is computed. This feature allows for user-defined algorithms where the IGP computes paths based on a combination of metric type and constraint. An operator can assign custom SR prefix-SIDs to realize forwarding beyond link-cost-based SPF. As a result, this feature provides a traffic-engineered path computed automatically by the IGP to any destination reachable by the IGP. This release supports the following functionality: • TI-LFA (IS-IS/OSPF) • Microloop Avoidance (IS-IS) • Inter-AS Support (IS-IS) • SID Redistribution (IS-IS) • Metric minimization—avoidance, multi-plane, delay (IS-IS/OSPF) • Affinity include (IS-IS/OSPF) • Affinity exclude (IS-IS/OSPF)

Segment Routing Flexible Algorithm allows operators to customize IGP shortest path computation according to their own needs. An operator can assign custom SR prefix-SIDs to realize forwarding beyond link-cost-based SPF. As a result, Flexible Algorithm provides a traffic engineered path automatically computed by the IGP to any destination reachable by the IGP.

The SR architecture associates prefix-SIDs to an algorithm which defines how the path is computed. Flexible Algorithm allows for user-defined algorithms where the IGP computes paths based on a user-defined combination of metric type and constraint.

This document describes the IS-IS and OSPF extensions to support Segment Routing Flexible Algorithm on an MPLS data-plane.

Prerequisites for Flexible Algorithm

Segment routing must be enabled on the router before the Flexible Algorithm functionality is activated.

Building Blocks of Segment Routing Flexible Algorithm

This section describes the building blocks that are required to support the SR Flexible Algorithm functionality in IS-IS and OSPF.

Flexible Algorithm Definition

Many possible constrains may be used to compute a path over a network. Some networks are deployed with multiple planes. A simple form of constrain may be to use a particular plane. A more sophisticated form of constrain can include some extended metric, like delay, as described in [RFC7810]. Even more advanced case could be to restrict the path and avoid links with certain affinities. Combinations of these are also possible. To provide a maximum flexibility, the mapping between the algorithm value and its meaning can be defined by the user. When all the routers in the domain have the common understanding what the particular algorithm value represents, the computation for such algorithm is consistent and the traffic is not subject to looping. Here, since the meaning of the algorithm is not defined by any standard, but is defined by the user, it is called as Flexible Algorithm.

Flexible Algorithm Support Advertisement

An algorithm defines how the best path is computed by IGP. Routers advertise the support for the algorithm as a node capability. Prefix-SIDs are also advertised with an algorithm value and are tightly coupled with the algorithm itself.

An algorithm is a one octet value. Values from 128 to 255 are reserved for user defined values and are used for Flexible Algorithm representation.

Flexible Algorithm Definition Advertisement

To guarantee the loop free forwarding for paths computed for a particular Flexible Algorithm, all routers in the network must share the same definition of the Flexible Algorithm. This is achieved by dedicated router(s) advertising the definition of each Flexible Algorithm. Such advertisement is associated with the priority to make sure that all routers will agree on a single and consistent definition for each Flexible Algorithm.

Definition of Flexible Algorithm includes:

- Metric type
- Affinity constraints

To enable the router to advertise the definition for the particular Flexible Algorithm, **advertise-definition** command is used. At least one router in the area, preferably two for redundancy, must advertise the Flexible Algorithm definition. Without the valid definition being advertised, the Flexible Algorithm will not be functional.

Flexible Algorithm Prefix-SID Advertisement

To be able to forward traffic on a Flexible Algorithm specific path, all routers participating in the Flexible Algorithm will install a MPLS labeled path for the Flexible Algorithm specific SID that is advertised for the prefix. Only prefixes for which the Flexible Algorithm specific Prefix-SID is advertised is subject to Flexible Algorithm specific forwarding.

Calculation of Flexible Algorithm Path

A router may compute path for multiple Flexible Algorithms. A router must be configured to support particular Flexible Algorithm before it can compute any path for such Flexible Algorithm. A router must have a valid definition of the Flexible Algorithm before such Flexible Algorithm is used.

When computing the shortest path tree for particular Flexible Algorithm:

- All nodes that do not advertise support for such Flexible Algorithm will be pruned from the topology.
- If the Flexible Algorithm definition includes affinities that are excluded, then all links for which any of such affinities are advertised will be pruned from the topology.
- Router uses the metric that is part of the Flexible Algorithm definition. If the metric is not advertised for the particular link, such link will be pruned from the topology.

IS-IS supports Loop Free Alternate (LFA) paths, TI-LFA backup paths, and Microloop Avoidance paths for particular Flexible Algorithm. OSPF supports Loop Free Alternate (LFA) and TI-LFA backup paths for particular Flexible Algorithm. These paths are computed using the same constraints as the calculation of the primary paths for such Flexible Algorithm. These paths use Prefix-SIDs advertised specifically for such Flexible Algorithm in order to enforce a backup or microloop avoidance path.

Installation of Forwarding Entries for Flexible Algorithm Paths

Flexible Algorithm path to any prefix must be installed in the forwarding using the Prefix-SID that was advertised for such Flexible Algorithm. If the Prefix-SID for Flexible Algorithm is not known, such Flexible Algorithm path is not installed in forwarding for such prefix..

Only MPLS to MPLS entries are installed for a Flexible Algorithm path. No IP to IP or IP to MPLS entries are installed. These follow the native IGP paths computed based on the default algorithm and regular IGP metrics.

Flexible Algorithm Prefix-SID Redistribution

Previously, prefix redistribution from IS-IS to another IS-IS instance or protocol was limited to SR algorithm 0 (regular SPF) prefix SIDs; SR algorithm 1 (Strict SPF) and SR algorithms 128-255 (Flexible Algorithm) prefix SIDs were not redistributed along with the prefix. The Segment Routing IS-IS Flexible Algorithm Prefix SID Redistribution feature allows redistribution of strict and flexible algorithms prefix SIDs from IS-IS to another IS-IS instance or protocols. This feature is enabled automatically when you configure redistribution of IS-IS Routes with strict or Flexible Algorithm SIDs.

Configuring Flexible Algorithm

Note For information about the commands usage, see the Segment Routing Command Reference for Cisco 8000 Series Routers.

The following ISIS and OSPF configuration sub-mode is used to configure Flexible Algorithm:

```
flex-algo algorithm number
```

algorithm number —value from 128 to 255

Commands under Flexible Algorithm Configuration Mode

The following commands are used to configure Flexible Algorithm definition under the flex-algo sub-mode:

```
metric-type delay
```

Note By default the regular IGP metric is used. If delay metric is enabled, the advertised delay on the link is used as a metric for Flexible Algorithm computation.

```
affinity {include-any | include-all | exclude-any}  name1, name2, …
```

name—name of the affinity map

```
priority priority value
```

priority value—priority used during the Flexible Algorithm definition election.

The following command is used to enable advertisement of the Flexible Algorithm definition in IS-IS:

```
advertise-definition
```

Commands for Affinity Configuration

The following command is used for defining the affinity-map. Affinity-map associates the name with the particular bit positions in the Extended Admin Group bitmask.

affinity-map *name* **bit-position** *bit number*

- *name*—name of the affinity-map.
- *bit number*—bit position in the Extended Admin Group bitmask.

The following command is used to associate the affinity with an interface:

affinity flex-algo *name 1, name 2, …*

name—name of the affinity-map

Command for Prefix-SID Configuration

The following command is used to advertise prefix-SID for default and strict-SPF algorithm:

prefix-sid [**strict-spf** | **algorithm** *algorithm-number*] [**index** | **absolute**] *sid value*

- *algorithm-number*—Flexible Algorithm number
- *sid value*—SID value

IS-IS Enhancements: max-metric and data plane updates

With the IOS XR Release 7.8.1, the new optional keyword **anomaly** is introduced to the **interface** submode of **affinity flex-algo**. This keyword option helps to advertise flex-algo affinity on PM anomaly. The following command is used to associate the affinity with an interface:

router isis *instance* **interface** *type interface-path-id* **affinity flex-algo anomaly** *name 1, name 2, …*

router ospf *process* **area** *area* **interface** *type interface-path-id* **affinity flex-algo anomaly** *name 1, name 2, …*

name—name of the affinity-map

You can configure both normal and anomaly values. For the following example, the **blue** affinity is advertised. However, if a metric is received with the anomaly flag set, it will change to **red**:

```
Router# configure
Router(config)# router isis 1
Router(config-isis)#flex-algo 128
Router(config-isis-flex-algo)# interface GigabitEthernet0/0/0/2
Router(config-isis-flex-algo)# affinity flex-algo blue
Router(config-isis-flex-algo)# affinity flex-algo anomaly red
```

Example: Configuring IS-IS Flexible Algorithm

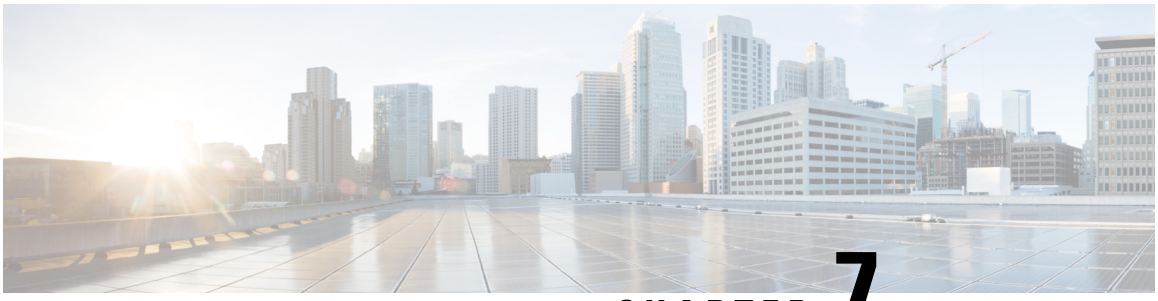
```
router isis 1
 affinity-map red bit-position 65
 affinity-map blue bit-position 8
 affinity-map green bit-position 201

 flex-algo 128
  advertise-definition
```

```
  affinity exclude-any red
  affinity include-any blue
 !
 flex-algo 129
  affinity exclude-any green
 !
!
address family ipv4 unicast
 segment-routing mpls
!
interface Loopback0
 address-family ipv4 unicast
  prefix-sid algorithm 128 index 100
  prefix-sid algorithm 129 index 101
!
!
interface GigabitEthernet0/0/0/0
 affinity flex-algo red
!
interface GigabitEthernet0/0/0/1
 affinity flex-algo blue red
!
interface GigabitEthernet0/0/0/2
 affinity flex-algo blue
!
```

Example: Configuring OSPF Flexible Algorithm

```
router ospf 1
 flex-algo 130
  priority 200
  affinity exclude-any
   red
   blue
  !
  metric-type delay
 !
 flex-algo 140
  affinity include-all
   green
  !
  affinity include-any
   red
  !
 !

  interface Loopback0
   prefix-sid index 10
   prefix-sid strict-spf index 40
   prefix-sid algorithm 128 absolute 16128
   prefix-sid algorithm 129 index 129
   prefix-sid algorithm 200 index 20
   prefix-sid algorithm 210 index 30
  !
 !

  interface GigabitEthernet0/0/0/0
   flex-algo affinity
    color red
    color blue
   !
  !
```

```
affinity-map
 color red bit-position 10
 color blue bit-position 11
!
```



CHAPTER 7

Configure Performance Measurement

Network performance metrics is a critical measure for traffic engineering (TE) in service provider networks. Network performance metrics include the following:

- Packet loss
- Delay
- Delay variation
- Bandwidth utilization

These network performance metrics provide network operators information about the performance characteristics of their networks for performance evaluation and help to ensure compliance with service level agreements. The service-level agreements (SLAs) of service providers depend on the ability to measure and monitor these network performance metrics. Network operators can use Segment Routing Performance Measurement (SR-PM) feature to monitor the network metrics for links and end-to-end TE label switched paths (LSPs).

The following table explains the functionalities supported by performance measurement feature for measuring delay for links or SR policies.

Table 13: Performance Measurement Functionalities

Functionality	Details
Profiles	You can configure different profiles for different types of delay measurements. Delay profile type interfaces is used for link-delay measurement. Delay profile type sr-policy is used for SR policy delay measurements. Delay profile allows you to schedule probe and configure metric advertisement parameters for delay measurement.
Protocols	Two-Way Active Measurement Protocol (TWAMP) Light (using RFC 5357 with IP/UDP encap).
Probe and burst scheduling	Schedule probes and configure metric advertisement parameters for delay measurement.
Metric advertisements	Advertise measured metrics periodically using configured thresholds. Also supports accelerated advertisements using configured thresholds.
Measurement history and counters	Maintain packet delay and loss measurement history and also session counters and packet advertisement counters.

Liveness Monitoring

Liveness refers to the ability of the network to confirm that a specific path, segment, or a node is operational and capable of forwarding packets. Liveness checks are essential for maintaining network availability and reliability.

Benefits

- **Fault Detection**: You can quickly identify if a device is down, which allows for immediate response and troubleshooting.
- **Load Balancing**: You can identify if the devices in a network are live, so work can be distributed more evenly across the network, preventing overloading of specific components and improving overall performance.
- **System Health**: You can provide an ongoing snapshot of a system's health, helping to identify potential issues before they become significant problems.
- **Maintenance Planning**: Liveness information can also help with maintenance planning, as system administrators can understand which components are live or down and plan maintenance and downtime accordingly without significant disruption to services.
- **Security**: Regular liveness checks can also play a role in maintaining network security. Administrators can take proactive steps to mitigate the damage and prevent future incidents by identifying unusual activity that might indicate a security breach or attack.

You can determine liveness for SR Policy and IP Endpoint.

Delay Measurement

Delay measuremnt is a mechanism used to measure the latency or delay experienced by data packets when they traverse a network.

The PM for delay measuremnt uses the IP/UDP packet format defined in RFC 5357 (TWAMP-Light) for probes. Two-Way Active Measurement Protocol (TWAMP) adds two-way or round-trip measurement capabilities. TWAMP employs time stamps applied at the echo destination (reflector) to enable greater accuracy. In the case of TWAMP Light, the Session-Reflector doesn’t necessarily know about the session state. The Session-Reflector simply copies the Sequence Number of the received packet to the Sequence Number field of the reflected packet. The controller receives the reflected test packets and collects two-way metrics. This architecture allows for collection of two-way metrics.

Benefits

- Network Troubleshooting: You can quickly and easily identify areas in your network with high delay and resolve network problems using delay measurement.
- Network Planning and Optimization: You can easily understand the performance of your network under various conditions and design a network that can handle expected traffic loads.

- Quality of Service (QoS): You can ensure quality of service standards are being met by continuously monitoring the delay in your network.

Supported Delay Measurement Methods

You can measure delay using the following methods:

- Use to monitor delay experienced by data packets in a single link or path between two nodes in a network.
- : Use to monitor the amount of time it takes for a data packet to travel from a source device to a specific IP endpoint within a network.
- : Use to to monitor the end-to-end delay experienced by the traffic sent over an SR policy.

Link Delay Measurement

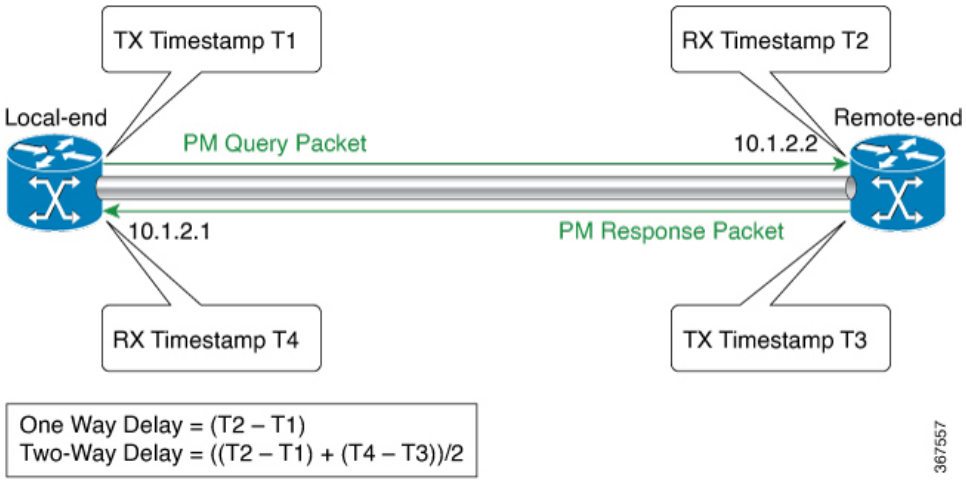
Table 14: Feature History Table

Feature Name	Release Information	Feature Description
Link Delay Measurement using TWAMP Light Encoding	Release 7.3.1	The PM for link delay uses the IP/UDP packet format defined in RFC 5357 (TWAMP-Light) for probes. Two-Way Active Measurement Protocol (TWAMP) adds two-way or round-trip measurement capabilities. TWAMP employs time stamps applied at the echo destination (reflector) to enable greater accuracy.

The PM for link delay uses the IP/UDP packet format defined in RFC 5357 (TWAMP-Light) for probes. Two-Way Active Measurement Protocol (TWAMP) adds two-way or round-trip measurement capabilities. TWAMP employs time stamps applied at the echo destination (reflector) to enable greater accuracy. In the case of TWAMP Light, the Session-Reflector doesn't necessarily know about the session state. The Session-Reflector simply copies the Sequence Number of the received packet to the Sequence Number field of the reflected packet. The controller receives the reflected test packets and collects two-way metrics. This architecture allows for collection of two-way metrics.

The following figure explains the PM query and response for link delay.

Figure 6: Performance Measurement for Link Delay



The PM query and response for link delay can be described in the following steps:

1. The local-end router sends PM query packets periodically to the remote side once the egress line card on the router applies timestamps on packets.
2. Ingress line card on the remote-end router applies time-stamps on packets as soon as they are received.
3. The remote-end router sends the PM packets containing time-stamps back to the local-end router. The remote-end router time-stamps the packet just before sending it for two-way measurement.
4. The local-end router time-stamps the packet as soon as the packet is received for two-way measurement.
5. One-way delay and optionally two-way delay is measured using the time-stamp values in the PM packet.

Restrictions and Usage Guidelines for PM for Link Delay

The following restrictions and guidelines apply for the PM for link delay feature for different links.

- For LSPs, remote-end line card needs to be MPLS and multicast MAC address capable.
- For broadcast links, only point-to-point (P2P) links are supported. P2P configuration on IGP is required for flooding the value.
- For link bundles, the hashing function may select a member link for forwarding but the reply may come from the remote line card on a different member link of the bundle.
- For one-way delay measurement, clocks should be synchronized on two end-point nodes of the link using PTP.
- Link delay measurement is supported on IPv4 unnumbered interfaces. An IPv4 unnumbered interface is identified by a node ID (a loopback address) and the local SNMP index assigned to the interface. Note that the reply messages could be received on any interface, since the packets are routed at the responder based on the loopback address used to identify the link.

Configuration Example: PM for Link Delay

This example shows how to configure performance-measurement functionalities for link delay as a global default profile. The default values for the different parameters in the PM for link delay is given as follows:

- **probe measurement mode**: The default measurement mode for probe is two-way delay measurement. If you are configuring one-way delay measurement, hardware clocks must be synchronized between the local-end and remote-end routers using precision time protocol (PTP).
- **protocol**: Interface delay measurement uses TWAMP-Light (RFC5357) with IP/UDP encap.
- **burst-interval**: Interval for sending probe packet. The default value is 3000 milliseconds and the range is from 30 to 15000 milliseconds.
- **computation interval**: Interval for metric computation. Default is 30 seconds; range is 1 to 3600 seconds.
- **periodic advertisement**: Periodic advertisement is enabled by default.
- **periodic-advertisement interval**: The default value is 120 seconds and the interval range is from 30 to 3600 seconds.
- **periodic-advertisement threshold**: The default value of periodic advertisement threshold is 10 percent.
- **periodic-advertisement minimum change**: The default value is 1000 microseconds (usec) and the range is from 0 to 10000 microseconds.
- **accelerated advertisement**: Accelerated advertisement is disabled by default.
- **accelerated-advertisement threshold**: The default value is 20 percent and the range is from 0 to 100 percent.
- **accelerated-advertisement minimum change**: The default value is 1000 microseconds and the range is from 1 to 100000 microseconds.

```
RP/0/0/CPU0:router(config)# performance-measurement delay-profile interfaces
RP/0/0/CPU0:router(config-pm-dm-intf)# probe
RP/0/0/CPU0:router(config-pm-dm-intf-probe)# measurement-mode one-way
RP/0/0/CPU0:router(config-pm-dm-intf-probe)# burst-interval 60
RP/0/0/CPU0:router(config-pm-dm-intf-probe)# exit

RP/0/0/CPU0:router(config-pm-dm-intf)# advertisement periodic
RP/0/0/CPU0:router(config-pm-dm-intf-adv-per)# interval 120
RP/0/0/CPU0:router(config-pm-dm-intf-adv-per)# threshold 20
RP/0/0/CPU0:router(config-pm-dm-intf-adv-per)# minimum-change 1000
RP/0/0/CPU0:router(config-pm-dm-intf-adv-per)# exit

RP/0/0/CPU0:router(config-pm-dm-intf)# advertisement accelerated
RP/0/0/CPU0:router(config-pm-dm-intf-adv-acc)# threshold 30
RP/0/0/CPU0:router(config-pm-dm-intf-adv-acc)# minimum-change 1000
RP/0/0/CPU0:router(config-pm-dm-intf-adv-per)# exit
```

Configure the UDP Destination Port

Configuring the UDP port for TWAMP-Light protocol is optional. By default, PM uses port 862 as the TWAMP-reserved UDP destination port.

The UDP port is configured for each PM measurement probe type (delay, loss, protocol, authentication mode, etc.) on querier and responder nodes. If you configure a different UDP port, the UDP port for each PM measurement probe type must match on the querier and the responder nodes.

Note The same UDP destination port is used for delay measurement for links and SR Policy.

This example shows how to configure the UDP destination port.

```
Router(config)# performance-measurement
Router(config-perf-meas)# protocol twamp-light
Router(config-pm-protocol)# measurement delay unauthenticated
Router(config-pm-proto-mode)# querier-dst-port 12000
```

Enable PM for Link Delay Over an Interface

This example shows how to enable PM for link delay over an interface.

```
RP/0/0/CPU0:router(config)# performance-measurement
RP/0/0/CPU0:router(config-perf-meas)# interface TenGigE0/0/0/0
RP/0/0/CPU0:router(config-pm-intf)# delay-measurement
RP/0/0/CPU0:router(config-pm-dm-intf)# exit
```

Verification

```
RP/0/0/CPU0:router# show performance-measurement profile interface
Thu Dec 12 14:13:16.029 PST

------------------------------------------------------------------------------
0/0/CPU0
------------------------------------------------------------------------------
Interface Delay-Measurement:
  Profile configuration:
    Measurement Type                          : Two-Way
    Probe computation interval                : 30 (effective: 30) seconds
    Type of services                          : Traffic Class: 6, DSCP: 48
    Burst interval                            : 3000 (effective: 3000) mSec
    Burst count                               : 10 packets
    Encap mode                                : UDP
    Payload Type                              : TWAMP-light
    Destination sweeping mode                 : Disabled
    Periodic advertisement                    : Enabled
      Interval                                : 120 (effective: 120) sec
      Threshold                               : 10%
      Minimum-Change                          : 500 uSec
    Advertisement accelerated                 : Disabled
    Threshold crossing check                  : Minimum-delay

RP/0/0/CPU0:router# show performance-measurement summary detail location 0/2/CPU0

Thu Dec 12 14:09:59.162 PST

------------------------------------------------------------------------------
0/2/CPU0
------------------------------------------------------------------------------
Total interfaces                              : 1
Total SR Policies                             : 0
```

```
Total RSVP-TE tunnels                          : 0
Total Maximum PPS                              : 2000 pkts/sec
Total Interfaces PPS                           : 0 pkts/sec
Maximum Allowed Multi-hop PPS                  : 2000 pkts/sec
Multi Hop Requested PPS                        : 0 pkts/sec (0% of max allowed)
Dampened Multi Hop Requested PPS               : 0% of max allowed
Inuse Burst Interval Adjustment Factor         : 100% of configuration


Interface Delay-Measurement:
  Total active sessions                        : 1
  Counters:
    Packets:
      Total sent                               : 26
      Total received                           : 26
    Errors:
        TX:
          Reason interface down                : 0
          Reason no MPLS caps                  : 0
          Reason no IP address                 : 0
          Reason other                         : 0
        RX:
          Reason negative delay                : 0
          Reason delay threshold exceeded      : 0
          Reason missing TX timestamp          : 0
          Reason missing RX timestamp          : 0
          Reason probe full                    : 0
          Reason probe not started             : 0
          Reason control code error            : 0
          Reason control code notif            : 0
    Probes:
      Total started                            : 3
      Total completed                          : 2
      Total incomplete                         : 0
      Total advertisements                     : 0

SR Policy Delay-Measurement:
  Total active sessions                        : 0
  Counters:
    Packets:
      Total sent                               : 0
      Total received                           : 0
    Errors:
        TX:
          Reason interface down                : 0
          Reason no MPLS caps                  : 0
          Reason no IP address                 : 0
          Reason other                         : 0
        RX:
          Reason negative delay                : 0
          Reason delay threshold exceeded      : 0
          Reason missing TX timestamp          : 0
          Reason missing RX timestamp          : 0
          Reason probe full                    : 0
          Reason probe not started             : 0
          Reason control code error            : 0
          Reason control code notif            : 0
    Probes:
      Total started                            : 0
      Total completed                          : 0
      Total incomplete                         : 0
      Total advertisements                     : 0

RSVP-TE Delay-Measurement:
```

```
  Total active sessions                       : 0
  Counters:
    Packets:
      Total sent                              : 0
      Total received                          : 0
    Errors:
        TX:
          Reason interface down               : 0
          Reason no MPLS caps                 : 0
          Reason no IP address                : 0
          Reason other                        : 0
        RX:
          Reason negative delay               : 0
          Reason delay threshold exceeded     : 0
          Reason missing TX timestamp         : 0
          Reason missing RX timestamp         : 0
          Reason probe full                   : 0
          Reason probe not started            : 0
          Reason control code error           : 0
          Reason control code notif           : 0
    Probes:
      Total started                           : 0
      Total completed                         : 0
      Total incomplete                        : 0
      Total advertisements                    : 0

Global Delay Counters:
  Total packets sent                          : 26
  Total query packets received                : 26
  Total invalid session id                    : 0
  Total missing session                       : 0
RP/0/0/CPU0:router# show performance-measurement interfaces detail
Thu Dec 12 14:16:09.692 PST

------------------------------------------------------------------------------
0/0/CPU0
------------------------------------------------------------------------------

------------------------------------------------------------------------------
0/2/CPU0
------------------------------------------------------------------------------

Interface Name: GigabitEthernet0/2/0/0 (ifh: 0x1004060)
  Delay-Measurement                 : Enabled
  Loss-Measurement                  : Disabled
  Configured IPv4 Address           : 10.10.10.2
  Configured IPv6 Address           : 10:10:10::2
  Link Local IPv6 Address           : fe80::3a:6fff:fec9:cd6b
  Configured Next-hop Address       : Unknown
  Local MAC Address                 : 023a.6fc9.cd6b
  Next-hop MAC Address              : 0291.e460.6707
  Primary VLAN Tag                  : None
  Secondary VLAN Tag                : None
  State                             : Up

  Delay Measurement session:
    Session ID          : 1

    Last advertisement:
      Advertised at: Dec 12 2019 14:10:43.138 (326.782 seconds ago)
      Advertised reason: First advertisement
      Advertised delays (uSec): avg: 839, min: 587, max: 8209, variance: 297

    Next advertisement:
```

```
  Threshold check scheduled in 1 more probe (roughly every 120 seconds)
  Aggregated delays (uSec): avg: 751, min: 589, max: 905, variance: 112
  Rolling average (uSec): 756

Current Probe:
  Started at Dec 12 2019 14:15:43.154 (26.766 seconds ago)
  Packets Sent: 9, received:9
  Measured delays (uSec): avg: 795, min: 631, max: 1199, variance: 164
  Next probe scheduled at Dec 12 2019 14:16:13.132 (in 3.212 seconds)
  Next burst packet will be sent in 0.212 seconds
  Burst packet sent every 3.0 seconds
  Probe samples:
    Packet Rx Timestamp       Measured Delay (nsec)
    Dec 12 2019 14:15:43.156          689223
    Dec 12 2019 14:15:46.156          876561
    Dec 12 2019 14:15:49.156          913548
    Dec 12 2019 14:15:52.157         1199620
    Dec 12 2019 14:15:55.156          794008
    Dec 12 2019 14:15:58.156          631437
    Dec 12 2019 14:16:01.157          656440
    Dec 12 2019 14:16:04.157          658267
    Dec 12 2019 14:16:07.157          736880
```

You can also use the following commands for verifying the PM for link delay on the local-end router.

Command	Description
show performance-measurement history probe interfaces [*interface*]	Displays the PM link-delay probe history for interfaces.
show performance-measurement history aggregated interfaces [*interface*]	Displays the PM link-delay aggregated history for interfaces.
show performance-measurement history advertisement interfaces [*interface*]	Displays the PM link-delay advertisement history for interfaces.
show performance-measurement counters [interface *interface*] [location *location-name*]	Displays the PM link-delay session counters.

You can also use the following commands for verifying the PM for link-delay configuration on the remote-end router.

Command	Description
show performance-measurement responder summary [**location** *location-name*]	Displays the PM for link-delay summary on the remote-end router (responder).
show performance-measurement responder interfaces [*interface*]	Displays PM for link-delay for interfaces on the remote-end router.
show performance-measurement responder counters [**interface** *interface*] [**location** *location-name*]	Displays the PM link-delay session counters on the remote-end router.

Delay Normalization

Performance measurement (PM) measures various link characteristics like packet loss and delay. Such characteristics can be used by IS-IS as a metric for Flexible Algorithm computation. Low latency routing using dynamic delay measurement is one of the primary use cases for Flexible Algorithm technology.

Delay is measured in microseconds. If delay values are taken as measured and used as link metrics during the IS-IS topology computation, some valid ECMP paths might be unused because of the negligible difference in the link delay.

The Delay Normalization feature computes a normalized delay value and uses the normalized value instead. This value is advertised and used as a metric during the Flexible Algorithm computation.

The normalization is performed when the delay is received from the delay measurement component. When the next value is received, it is normalized and compared to the previous saved normalized value. If the values are different, then the LSP generation is triggered.

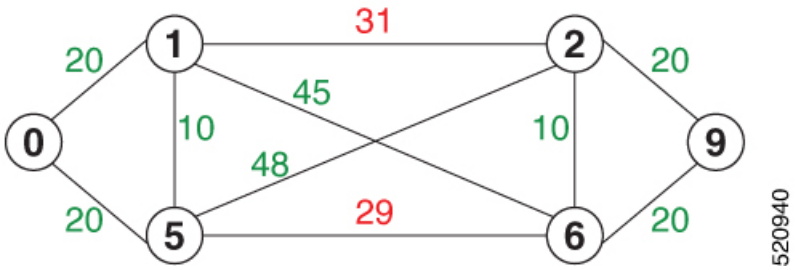
The following formula is used to calculate the normalized value:

- **Dm** – measured Delay
- **Int** – configured normalized Interval
- **Off** – configured normalized Offset (must be less than the normalized interval Int)
- **Dn** – normalized Delay
- **a** = Dm / Int (rounded down)
- **b** = a * Int + Off

If the measured delay (Dm) is less than or equal to **b**, then the normalized delay (Dn) is equal to **b**. Otherwise, Dn is **b** + **Int**.

Example

The following example shows a low-latency service. The intent is to avoid high-latency links (1-6, 5-2). Links 1-2 and 5-6 are both low-latency links. The measured latency is not equal, but the difference is insignificant.



We can normalize the measured latency before it is advertised and used by IS-IS. Consider a scenario with the following:

- Interval = 10
- Offset = 3

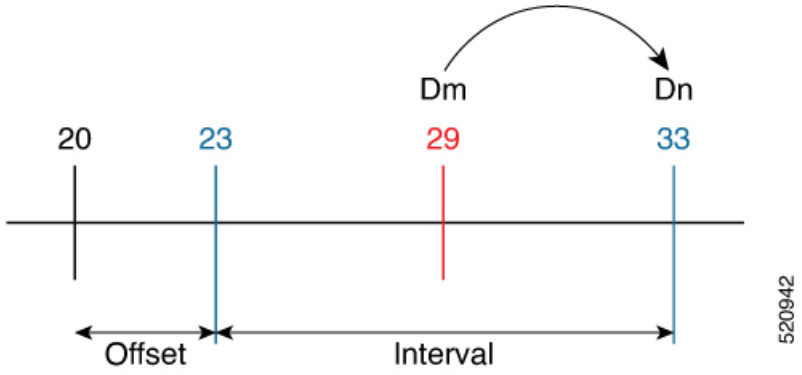
The measured delays will be normalized as follows:

- **Dm** = 29

a = 29 / 10 = 2 (2.9, rounded down to 2)

b = 2 * 10 + 3 = 23

In this case, **Dm** (29) is greater than **b** (23); so **Dn** is equal to **b**+**I** (23 + 10) = **33**

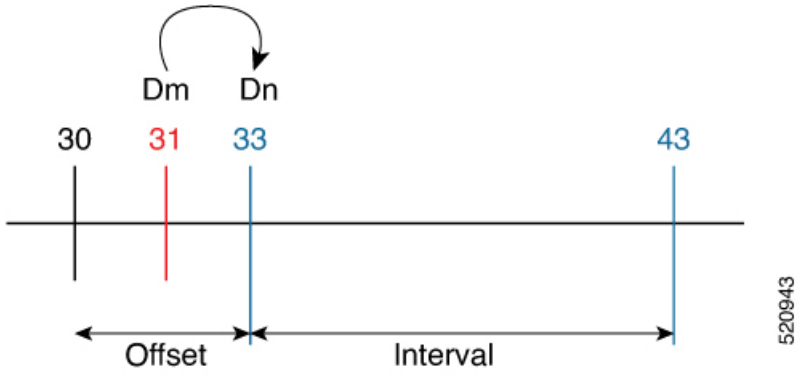


- **Dm** = 31

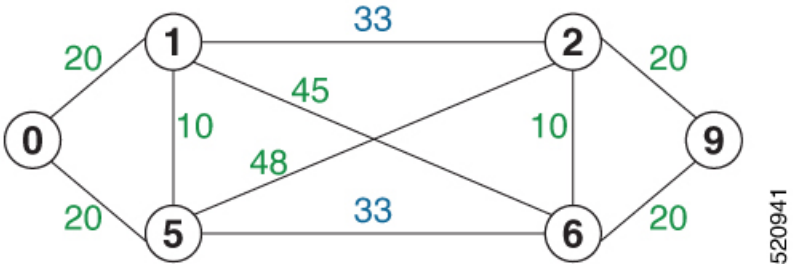
  a = 31 / 10 = 3 (3.1, rounded down to 3)

  b = 3 * 10 + 3 = 33

  In this case, **Dm** (31) is less than **b** (33); so **Dn** is **b** = **33**



The link delay between 1-2 and 5-6 is normalized to 33.



Configuration

Delay normalization is disabled by default. To enable and configure delay normalization, use the **delay normalize interval** *interval* [**offset** *offset*] command.

- *interval* – The value of the normalize interval in microseconds.
- *offset* – The value of the normalized offset in microseconds. This valude must be smaller than the value of normalized interval.

IS-IS Configuration

```
router isis 1
 interface GigEth 0/0/0/0
  delay normalize interval 10 offset 3
  address-family ipv4 unicast
   metric 77
```

OSPF Configuration

```
router ospf 1
 area 0
  interface GigabitEthernet0/0/0/0
   delay normalize interval 10 offset 3
  !
 !
!
```

Delay Measurement for IP Endpoint

Table 15: Feature History Table

Feature Name	Release Information	Feature Description
IP Endpoint Delay Measurement Monitoring	Release 7.4.1	This feature measures the end-to-end delay and monitors liveness of a specified IP endpoint node, including VRF-aware (awareness of multiple customers belonging to different VRFs). This feature is supported on IPv4, IPv6, and MPLS data planes.

Delay for an IP endpoint is the amount of time it takes for a data packet to travel from a source device to a specific IP endpoint within a network.

To measure a delay for a packet, also called a probe, is sent from a source device to the target IP endpoint.

The time from when the packet leaves the source to when it arrives at the endpoint is measured and recorded as the delay.

You can measure one-way delay, Two-way delay, and Roundtrip delay or delay in loop-back mode. For more information on Delay measurement, see Link Delay Measurement and Measurement Modes.

Collecting IP Endpoint Probe Statistics

- Statistics associated with the probe for delay metrics are available via Histogram and Streaming Telemetry.
- Model Driven Telemetry (MDT) is supported for the following data:

    - Summary, endpoint, session, and counter show command bags.
    - History buffers data

- Model Driven Telemetry (MDT) and Event Driven Telemetry (EDT) are supported for the following data:
    - Delay metrics computed in the last probe computation-interval (event: probe-completed)
    - Delay metrics computed in the last aggregation-interval; that is, end of the periodic advertisement-interval (event: advertisement-interval expired)
    - Delay metrics last notified (event: notification-triggered)

- The following xpaths for MDT/EDT is supported:
    - `Cisco-IOS-XR-perf-meas-oper:performance-measurement/nodes/node/endpoints/endpoint-delay/endpoint-last-probes`
    - `Cisco-IOS-XR-perf-meas-oper:performance-measurement/nodes/node/endpoints/endpoint-delay/endpoint-last-aggregations`
    - `Cisco-IOS-XR-perf-meas-oper:performance-measurement/nodes/node/endpoints/endpoint-delay/endpoint-last-advertisements`

Supported Features

- IPv6 Endpoint Delay in Default VRF (over SRv6)
- SRv6 Endpoint Delay in Default VRF (Endpoint can be Node SID, Flex-Algo SID, Packed uSID carrier)
- IPv6 Endpoint Delay in VRF (static uDT6)
- IPv6 Endpoint Delay in VRF (dynamic uDT6 encap)
- IPv4 Endpoint Delay in VRF or GRT (static uDT4)
- IPv4 Endpoint Delay in VRF or GRT (dynamic uDT4 encap)

Guidelines and Limitations

You can specify a custom labeled path through one or more user-configured segment-lists. User-configured segment-list represents the forwarding path from sender to reflector when the probe is configured in delay-measurement mode.

- SR PM is supported on hardware that supports Precision Time Protocol (PTP). This requirement applies to both one-way and two-way delay measurement.

  See the "**Configuring Precision Time Protocol**" chapter in the *System Management Configuration Guide for Cisco 8000 Series Routers* for Restrictions for PTP and the Timing Hardware Support Matrix.

- Examples of the custom segment-list include:
    - Probe in delay-measurement mode with a segment-list that includes Flex-Algo prefix SID of the endpoint

- Probe in delay-measurement mode with a segment-list that includes a SID-list with labels to reach the endpoint or the sender (forward direction)
- Probe in delay-measurement mode with a segment-list that includes BSID associated with SR policy to reach the end point.

- Endpoint segment list configuration is not supported under nondefault VRF.

IP Endpoint Liveness Detection in an SR MPLS Network

IP endpoint liveness detection leverages the loopback measurement-mode. The following workflow describes the sequence of events.

1. The sender creates and transmits the PM probe packets.

   The IP destination address (DA) on the probe packets is set to the loopback value of the sender itself.

   The transmit timestamp (T1) is added to the payload.

   The probe packet is encapsulated with the label corresponding to the endpoint.
2. The network delivers the PM probe packets following the LSP toward the endpoint.
3. The end-point receives the PM probe packets.

   Packets are forwarded back to the sender based on the forwarding entry associated with the IP DA of the PM probe packet. If an LSP exists, the probe packet is encapsulated with the label of the sender.
4. The sender node receives the PM probe packets.

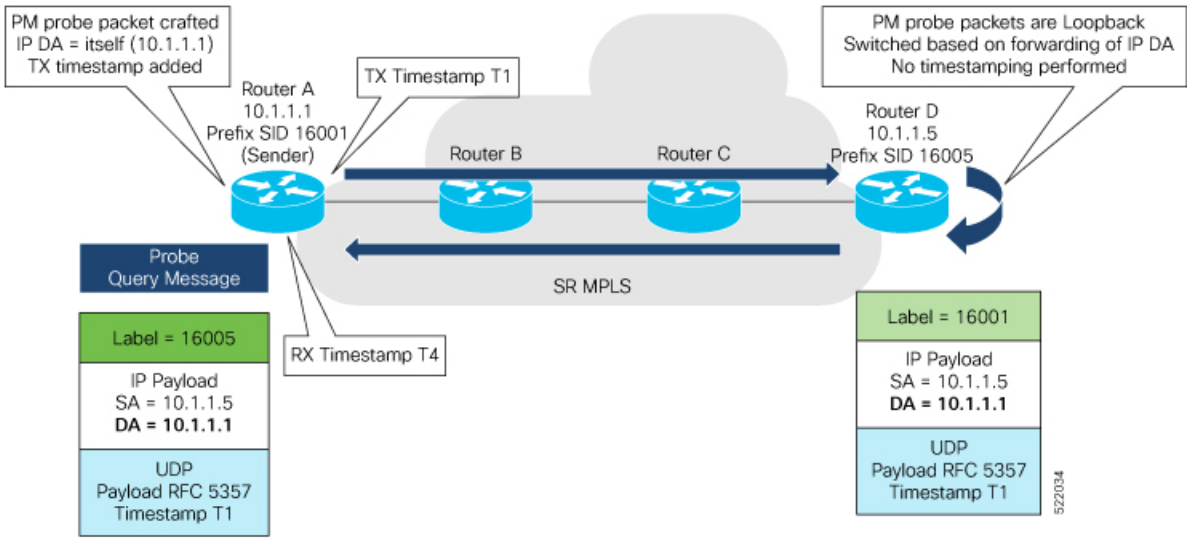
   The received timestamp (T4) stored.

   If the sender node doesn't receive the specified number of probe packets (based on the configured multiplier), the sender node declares the PM session as down.

The following figure illustrates a liveness detection probe toward an IP endpoint learned by the IGP. The network interconnecting the sender and reflector provides MPLS connectivity with Segment Routing.

The liveness detection multiplier is set to 5 to specify the number of consecutive missed probe packets before the PM session is declared as down.

Figure 7: IP Endpoint Liveness Detection



Configuration Example

```
RouterA(config)# performance-measurement
RouterA(config-perf-meas)# endpoint ipv4 1.1.1.5
RouterA(config-pm-ep)# source-address ipv4 1.1.1.1
RouterA(config-pm-ep)# liveness-detection
RouterA(config-pm-ep-ld)# exit
RouterA(config-pm-ep)# exit
RouterA(config-perf-meas)# liveness-profile endpoint default
RouterA(config-pm-ld-ep)# liveness-detection
RouterA(config-pm-ld-ep-ld)# multiplier 5
RouterA(config-pm-ld-ep-ld)# exit
RouterA(config-pm-ld-ep)# probe
RouterA(config-pm-ld-ep-probe)# measurement-mode loopback
```

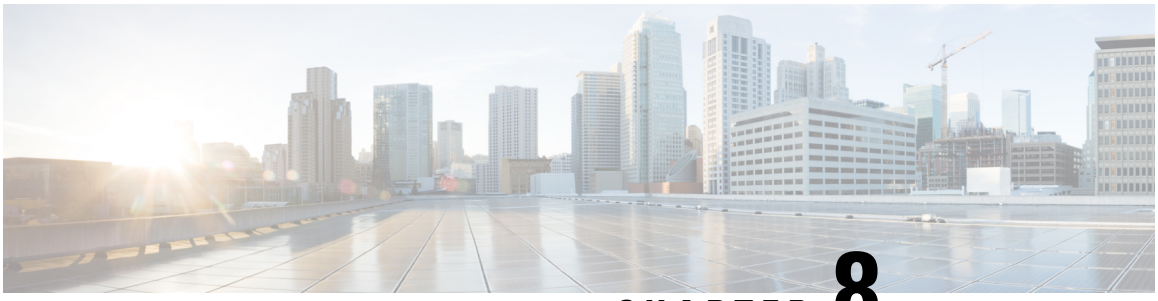
Running Configuration

```
performance-measurement
 endpoint ipv4 1.1.1.5
  source-address ipv4 1.1.1.1
  liveness-detection
  !
 !
 liveness-profile endpoint default
  liveness-detection
   multiplier 5
  !
  probe
   measurement-mode loopback
  !
 !
!
end
```

Verification

```
RouterA# show performance-measurement endpoint ipv4 1.1.1.5
```

```
--------------------------------------------------------------------------------
0/RSP0/CPU0
--------------------------------------------------------------------------------

Endpoint name: IPv4-1.1.1.5-vrf-default
  Source address              : 1.1.1.1
  VRF name                    : default
  Liveness Detection          : Enabled
  Profile Keys:
    Profile name              : default
    Profile type              : Endpoint Liveness Detection

  Segment-list                : None
  Session State: Down
  Missed count: 0
```



CHAPTER 8

Configure Topology-Independent Loop-Free Alternate (TI-LFA)

Topology-Independent Loop-Free Alternate (TI-LFA) uses segment routing to provide link, node, and Shared Risk Link Groups (SRLG) protection in topologies where other fast reroute techniques cannot provide protection.

- Classic Loop-Free Alternate (LFA) is topology dependent, and therefore cannot protect all destinations in all networks. A limitation of LFA is that, even if one or more LFAs exist, the optimal LFA may not always be provided.
- Remote LFA (RLFA) extends the coverage to 90-95% of the destinations, but it also does not always provide the most desired repair path. RLFA also adds more operational complexity by requiring a targeted LDP session to the RLFAs to protect LDP traffic.

TI-LFA provides a solution to these limitations while maintaining the simplicity of the IPFRR solution.

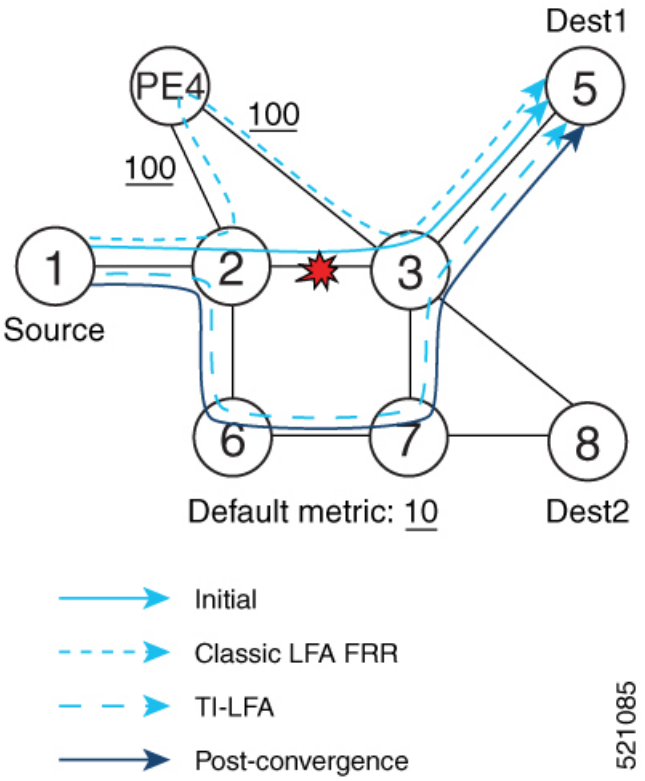
The goal of TI-LFA is to reduce the packet loss that results while routers converge after a topology change due to a link or node failure. Rapid failure repair (< 50 msec) is achieved through the use of pre-calculated backup paths that are loop-free and safe to use until the distributed network convergence process is completed.

The optimal repair path is the path that the traffic will eventually follow after the IGP has converged. This is called the post-convergence path. This path is preferred for the following reasons:

- Optimal for capacity planning — During the capacity-planning phase of the network, the capacity of a link is provisioned while taking into consideration that such link with be used when other links fail.
- Simple to operate — There is no need to perform a case-by-case adjustments to select the best LFA among multiple candidate LFAs.
- Fewer traffic transitions — Since the repair path is equal to the post-convergence path, the traffic switches paths only once.

The following topology illustrates the optimal and automatic selection of the TI-LFA repair path.

Figure 8: TI-LFA Repair Path



Node 2 protects traffic to destination Node 5.

With classic LFA, traffic would be steered to Node 4 after a failure of the protected link. This path is not optimal, since traffic is routed over edge node Node 4 that is connected to lower capacity links.

TI-LFA calculates a post-convergence path and derives the segment list required to steer packets along the post-convergence path without looping back.

In this example, if the protected link fails, the shortest path from Node2 to Node5 would be:

Node2 → Node6 → Node7 → Node3 → Node5

Node7 is the PQ-node for destination Node5. TI-LFA encodes a single segment (prefix SID of Node7) in the header of the packets on the repair path.

TI-LFA Protection Types

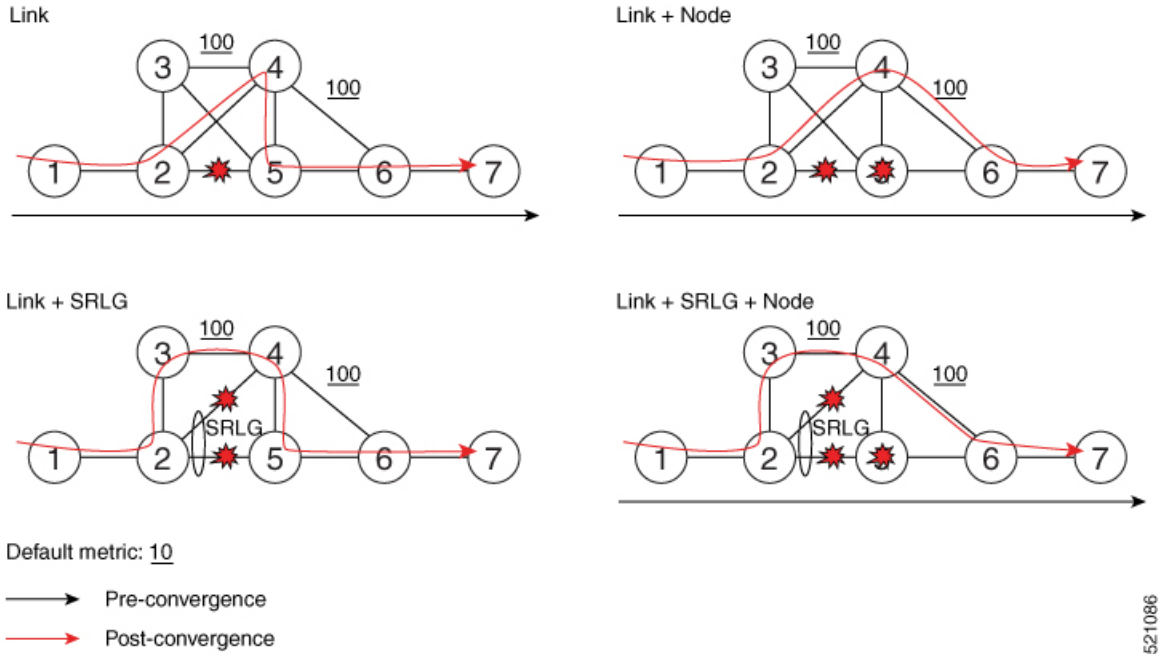
TI-LFA supports the following protection:

- Link protection — The link is excluded during the post-convergence backup path calculation.
- Node protection — The neighbor node is excluded during the post convergence backup path calculation.
- Shared Risk Link Groups (SRLG) protection — SRLG refer to situations in which links in a network share a common fiber (or a common physical attribute). These links have a shared risk: when one link fails, other links in the group might also fail. TI-LFA SRLG protection attempts to find the post-convergence backup path that excludes the SRLG of the protected link. All local links that share any SRLG with the protecting link are excluded.

When you enable link protection, you can also enable node protection, SRLG protection, or both, and specify a tiebreaker priority in case there are multiple LFAs.

The following example illustrates the link, node, and SRLG protection types. In this topology, Node2 applies different protection models to protect traffic to Node7.

Figure 9: TI-LFA Protection Types



- Behaviors and Limitations of TI-LFA, on page 125
- Configuring TI-LFA for IS-IS, on page 127
- Configuring TI-LFA for OSPF, on page 128
- TI-LFA Node and SRLG Protection: Examples, on page 130
- Configuring Global Weighted SRLG Protection, on page 131

Behaviors and Limitations of TI-LFA

Table 16: Feature History Table

Feature Name	Release Information	Feature Description
BFDv6-triggered TI-LFA	Release 7.3.2	Topology-Independent Loop-Free Alternate (TI-LFA) uses segment routing to provide link, node, and Shared Risk Link Groups (SRLG) protection in topologies where other fast reroute techniques cannot provide protection. BFDv6-triggered TI-LFA allows you to obtain link, node, and SRLG protection using the Bidirectional Forwarding Detection (BFD) over IPv6 protocol.

Feature Name	Release Information	Feature Description
BFDv4-triggered TI-LFA	Release 7.3.1	Topology-Independent Loop-Free Alternate (TI-LFA) uses segment routing to provide link, node, and Shared Risk Link Groups (SRLG) protection in topologies where other fast reroute techniques cannot provide protection. BFD-triggered TI-LFA allows you to obtain link, node, and SRLG protection by using the Bidirectional Forwarding Detection (BFD) protocol.

The TI-LFA behaviors and limitations are listed below:

- IGP directly programs a TI-LFA backup path requiring three or fewer labels, including the label of the protected destination prefix.
- Programming of TI-LFA backup paths requiring more than three labels is not supported.

TI-LFA Functionality	IS-IS[1]	OSPFv2
Protected Traffic Types		
Protection for SR labeled traffic	Supported	Supported
Protection of IPv4 unlabeled traffic	Supported (IS-ISv4)	Supported
Protection of IPv6 unlabeled traffic	Supported (IS-ISv6)	N/A
Protection Types		
Link Protection	Supported	Supported
Node Protection	Supported	Supported
Local SRLG Protection	Supported	Supported
Weighted Remote SRLG Protection	Supported	Supported
Line Card Disjoint Protection	Unsupported	Unsupported
Interface Types		
Ethernet Interfaces	Supported	Supported
Ethernet Bundle Interfaces	Supported	Supported
TI-LFA over GRE Tunnel as Protecting Interface	Unsupported	Unsupported
Additional Functionality		
Maximum number of labels that can be pushed on the backup path (including the label of the protected prefix)	3	3
BFDv4-triggered	Supported	Supported
BFDv6-triggered	Supported	N/A

TI-LFA Functionality	IS-IS[1]	OSPFv2
Prefer backup path with lowest total metric	Unsupported	Unsupported
Prefer backup path from ECMP set	Unsupported	Unsupported
Prefer backup path from non-ECMP set	Unsupported	Unsupported
Load share prefixes across multiple backups paths	Supported	Supported
Limit backup computation up to the prefix priority	Supported	Supported

[1] Unless specified, IS-IS support is IS-ISv4 and IS-ISv6

Configuring TI-LFA for IS-IS

This task describes how to enable per-prefix Topology Independent Loop-Free Alternate (TI-LFA) computation to converge traffic flows around link, node, and SRLG failures.

Before you begin

Ensure that the following topology requirements are met:

- Routers are configured with IS-IS.
- Segment routing for IS-IS is configured. See Enabling Segment Routing for IS-IS Protocol, on page 15.

SUMMARY STEPS

1. **configure**
2. **router isis** *instance-id*
3. **interface** *type interface-path-id*
4. **address-family** {**ipv4** | **ipv6**} [**unicast**]
5. **fast-reroute per-prefix**
6. **fast-reroute per-prefix ti-lfa**
7. **fast-reroute per-prefix tiebreaker** {**node-protecting** | **srlg-disjoint**} **index** *priority*

DETAILED STEPS

	Command or Action	Purpose
Step 1	**configure** **Example:** `RP/0/RP0/CPU0:router# configure`	Enters mode.
Step 2	**router isis** *instance-id* **Example:** `RP/0/RP0/CPU0:router(config)# router isis 1`	Enables IS-IS routing for the specified routing instance, and places the router in router configuration mode. **Note** You can change the level of routing to be performed by a particular routing instance by using the **is-type** router configuration command.

	Command or Action	Purpose
Step 3	**interface** *type interface-path-id* **Example:** `RP/0/RP0/CPU0:router(config-isis)# interface GigabitEthernet0/0/0/1`	Enters interface configuration mode.
Step 4	**address-family** {**ipv4** | **ipv6**} [**unicast**] **Example:** `RP/0/RP0/CPU0:router(config-isis-if)# address-family ipv4 unicast`	Specifies the IPv4 or IPv6 address family, and enters router address family configuration mode.
Step 5	**fast-reroute per-prefix** **Example:** `RP/0/RP0/CPU0:router(config-isis-if-af)# fast-reroute per-prefix`	Enables per-prefix fast reroute .
Step 6	**fast-reroute per-prefix ti-lfa** **Example:** `RP/0/RP0/CPU0:router(config-isis-if-af)# fast-reroute per-prefix ti-lfa`	Enables per-prefix TI-LFA fast reroute link protection.
Step 7	**fast-reroute per-prefix tiebreaker** {**node-protecting** | **srlg-disjoint**} **index** *priority* **Example:** `RP/0/RP0/CPU0:router(config-isis-if-af)# fast-reroute per-prefix srlg-disjoint index 100`	Enables TI-LFA node or SRLG protection and specifies the tiebreaker priority. Valid *priority* values are from 1 to 255. The lower the *priority* value, the higher the priority of the rule. Link protection always has a lower priority than node or SRLG protection. **Note** The same attribute cannot be configured more than once on an interface.

TI-LFA has been successfully configured for segment routing.

Configuring TI-LFA for OSPF

This task describes how to enable per-prefix Topology Independent Loop-Free Alternate (TI-LFA) computation to converge traffic flows around link, node, and SRLG failures.

Note TI-LFA can be configured on the instance, area, or interface. When configured on the instance or area, all interfaces in the instance or area inherit the configuration.

Before you begin

Ensure that the following topology requirements are met:

- Routers are configured with OSPF.
- Segment routing for OSPF is configured. See Enabling Segment Routing for OSPF Protocol, on page 41.

SUMMARY STEPS

1. **configure**
2. **router ospf** *process-name*
3. **area** *area-id*
4. **interface** *type interface-path-id*
5. **fast-reroute per-prefix**
6. **fast-reroute per-prefix ti-lfa**
7. **fast-reroute per-prefix tiebreaker** {**node-protecting** | **srlg-disjoint**} **index** *priority*

DETAILED STEPS

	Command or Action	Purpose
Step 1	**configure** **Example:** `RP/0/RP0/CPU0:router# configure`	Enters mode.
Step 2	**router ospf** *process-name* **Example:** `RP/0/RP0/CPU0:router(config)# router ospf 1`	Enables OSPF routing for the specified routing process, and places the router in router configuration mode.
Step 3	**area** *area-id* **Example:** `RP/0/RP0/CPU0:router(config-ospf)# area 1`	Enters area configuration mode.
Step 4	**interface** *type interface-path-id* **Example:** `RP/0/RP0/CPU0:router(config-ospf-ar)# interface GigabitEthernet0/0/0/1`	Enters interface configuration mode.
Step 5	**fast-reroute per-prefix** **Example:** `RP/0/RP0/CPU0:router(config-ospf-ar-if)# fast-reroute per-prefix`	Enables per-prefix fast reroute.

	Command or Action	Purpose

Step 6	**fast-reroute per-prefix ti-lfa**
	Example:
	`RP/0/RP0/CPU0:router(config-ospf-ar-if)#`
	`fast-reroute per-prefix ti-lfa`
		Enables per-prefix TI-LFA fast reroute link protection.

Step 7	**fast-reroute per-prefix tiebreaker** {**node-protecting** | **srlg-disjoint**} **index** *priority*
	Example:
	`RP/0/RP0/CPU0:router(config-isis-ar-if)#`
	`fast-reroute per-prefix srlg-disjoint index 100`
		Enables TI-LFA node or SRLG protection and specifies the tiebreaker priority. Valid *priority* values are from 1 to 255. The higher the *priority* value, the higher the priority of the rule. Link protection always has a lower priority than node or SRLG protection.
		Note The same attribute cannot be configured more than once on an interface.

TI-LFA has been successfully configured for segment routing.

TI-LFA Node and SRLG Protection: Examples

The following examples show the configuration of the tiebreaker priority for TI-LFA node and SRLG protection, and the behavior of post-convergence backup-path. These examples use OSPF, but the same configuration and behavior applies to IS-IS.

Example: Enable link-protecting and node-protecting TI-LFA

```
router ospf 1
 area 1
  interface GigabitEthernet0/0/2/1
    fast-reroute per-prefix
    fast-reroute per-prefix ti-lfa
    fast-reroute per-prefix tiebreaker node-protecting index 100
```

Both link-protecting and node-protecting TI-LFA backup paths will be computed. If the priority associated with the node-protecting tiebreaker is higher than any other tiebreakers, then node-protecting post-convergence backup paths will be selected, if it is available.

Example: Enable link-protecting and SRLG-protecting TI-LFA

```
router ospf 1
 area 1
  interface GigabitEthernet0/0/2/1
    fast-reroute per-prefix
    fast-reroute per-prefix ti-lfa
    fast-reroute per-prefix tiebreaker srlg-disjoint index 100
```

Both link-protecting and SRLG-protecting TI-LFA backup paths will be computed. If the priority associated with the SRLG-protecting tiebreaker is higher than any other tiebreakers, then SRLG-protecting post-convergence backup paths will be selected, if it is available.

Example: Enable link-protecting, node-protecting and SRLG-protecting TI-LFA

```
router ospf 1
 area 1
  interface GigabitEthernet0/0/2/1
    fast-reroute per-prefix
    fast-reroute per-prefix ti-lfa
    fast-reroute per-prefix tiebreaker node-protecting index 200
    fast-reroute per-prefix tiebreaker srlg-disjoint index 100
```

Link-protecting, node-protecting, and SRLG-protecting TI-LFA backup paths will be computed. If the priority associated with the node-protecting tiebreaker is highest from all tiebreakers, then node-protecting post-convergence backup paths will be selected, if it is available. If the node-protecting backup path is not available, SRLG-protecting post-convergence backup path will be used, if it is available.

Configuring Global Weighted SRLG Protection

A shared risk link group (SRLG) is a set of links sharing a common resource and thus shares the same risk of failure. The existing loop-free alternate (LFA) implementations in interior gateway protocols (IGPs) support SRLG protection. However, the existing implementation considers only the directly connected links while computing the backup path. Hence, SRLG protection may fail if a link that is not directly connected but shares the same SRLG is included while computing the backup path. Global weighted SRLG protection feature provides better path selection for the SRLG by associating a weight with the SRLG value and using the weights of the SRLG values while computing the backup path.

To support global weighted SRLG protection, you need information about SRLGs on all links in the area topology. You can flood SRLGs for remote links using ISIS or manually configuring SRLGS on remote links.

Configuration Examples: Global Weighted SRLG Protection

There are three types of configurations that are supported for the global weighted SRLG protection feature.

- local SRLG with global weighted SRLG protection
- remote SRLG flooding
- remote SRLG static provisioning

This example shows how to configure the local SRLG with global weighted SRLG protection feature.

```
RP/0/RP0/CPU0:router(config)# srlg
RP/0/RP0/CPU0:router(config-srlg)# interface TenGigE0/0/0/0
RP/0/RP0/CPU0:router(config-srlg-if)# name group1
RP/0/RP0/CPU0:router(config-srlg-if)# exit
RP/0/RP0/CPU0:router(config-srlg)# interface TenGigE0/0/0/1
RP/0/RP0/CPU0:router(config-srlg-if)# name group1
RP/0/RP0/CPU0:router(config-srlg)# name group value 100
RP/0/RP0/CPU0:router(config)# router isis 1
RP/0/RP0/CPU0:router(config-isis)# address-family ipv4 unicast
RP/0/RP0/CPU0:router(config-isis-if-af)# fast-reroute per-prefix srlg-protection
weighted-global
RP/0/RP0/CPU0:router(config-isis-if-af)# fast-reroute per-prefix tiebreaker srlg-disjoint
index 1
RP/0/RP0/CPU0:router(config-isis)# interface TenGigE0/0/0/0
RP/0/RP0/CPU0:router(config-isis-if)# point-to-point
RP/0/RP0/CPU0:router(config-isis-if)# address-family ipv4 unicast
```

```
RP/0/RP0/CPU0:router(config-isis-if-af)# fast-reroute per-prefix
RP/0/RP0/CPU0:router(config-isis-if-af)# fast-reroute per-prefix ti-lfa
RP/0/RP0/CPU0:router(config-isis)# srlg
RP/0/RP0/CPU0:router(config-isis-srlg)# name group1
RP/0/RP0/CPU0:router(config-isis-srlg-name)# admin-weight 5000
```

This example shows how to configure the global weighted SRLG protection feature with remote SRLG flooding.The configuration includes local and remote router configuration. On the local router, the global weighted SRLG protection is enabled by using the **fast-reroute per-prefix srlg-protection weighted-global** command. In the remote router configuration, you can control the SRLG value flooding by using the **advertise application lfa link-attributes srlg** command. You should also globally configure SRLG on the remote router.

The local router configuration for global weighted SRLG protection with remote SRLG flooding is as follows:

```
RP/0/RP0/CPU0:router(config)# router isis 1
RP/0/RP0/CPU0:router(config-isis)# address-family ipv4 unicast
RP/0/RP0/CPU0:router(config-isis-if-af)# fast-reroute per-prefix srlg-protection
weighted-global
RP/0/RP0/CPU0:router(config-isis-if-af)# fast-reroute per-prefix tiebreaker srlg-disjoint
index 1
RP/0/RP0/CPU0:router(config-isis-if-af)# exit
RP/0/RP0/CPU0:router(config-isis)# interface TenGigE0/0/0/0
RP/0/RP0/CPU0:router(config-isis-if)# point-to-point
RP/0/RP0/CPU0:router(config-isis-if)# address-family ipv4 unicast
RP/0/RP0/CPU0:router(config-isis-if-af)# fast-reroute per-prefix
RP/0/RP0/CPU0:router(config-isis-if-af)# fast-reroute per-prefix ti-lfa
RP/0/RP0/CPU0:router(config-isis-if-af)# exit
RP/0/RP0/CPU0:router(config-isis)# srlg
RP/0/RP0/CPU0:router(config-isis-srlg)# name group1
RP/0/RP0/CPU0:router(config-isis-srlg-name)# admin-weight 5000
```

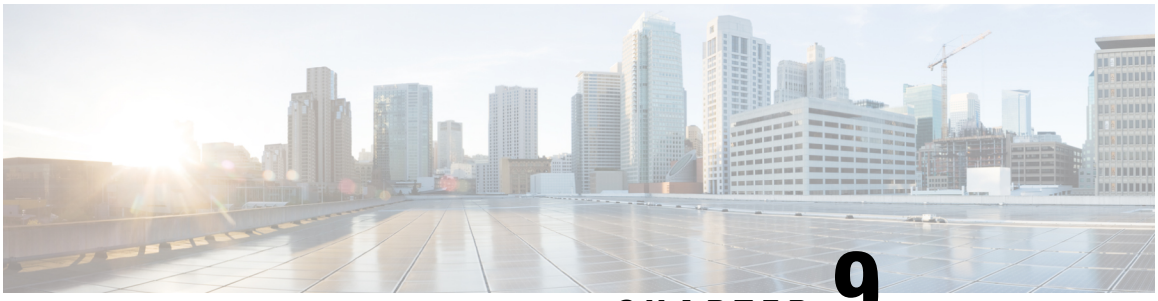
The remote router configuration for global weighted SRLG protection with remote SRLG flooding is as follows:

```
RP/0/RP0/CPU0:router(config)# srlg
RP/0/RP0/CPU0:router(config-srlg)# interface TenGigE0/0/0/0
RP/0/RP0/CPU0:router(config-srlg-if)# name group1
RP/0/RP0/CPU0:router(config-srlg-if)# exit
RP/0/RP0/CPU0:router(config-srlg)# interface TenGigE0/0/0/1
RP/0/RP0/CPU0:router(config-srlg-if)# name group1
RP/0/RP0/CPU0:router(config-srlg)# name group value 100
RP/0/RP0/CPU0:router(config-srlg)# exit
RP/0/RP0/CPU0:router(config)# router isis 1
RP/0/RP0/CPU0:(config-isis)# address-family ipv4 unicast
RP/0/RP0/CPU0:router(config-isis-af)# advertise application lfa link-attributes srlg
```

This example shows configuring the global weighted SRLG protection feature with static provisioning of SRLG values for remote links. You should perform these configurations on the local router.

```
RP/0/RP0/CPU0:router(config)# srlg
RP/0/RP0/CPU0:router(config-srlg)# interface TenGigE0/0/0/0
RP/0/RP0/CPU0:router(config-srlg-if)# name group1
RP/0/RP0/CPU0:router(config-srlg-if)# exit
RP/0/RP0/CPU0:router(config-srlg)# interface TenGigE0/0/0/1
RP/0/RP0/CPU0:router(config-srlg-if)# name group1
RP/0/RP0/CPU0:router(config-srlg)# name group value 100
RP/0/RP0/CPU0:router(config-srlg)# exit
RP/0/RP0/CPU0:router(config)# router isis 1
RP/0/RP0/CPU0:router(config-isis)# address-family ipv4 unicast
RP/0/RP0/CPU0:router(config-isis-if-af)# fast-reroute per-prefix srlg-protection
```

```
weighted-global
RP/0/RP0/CPU0:router(config-isis-if-af)# fast-reroute per-prefix tiebreaker srlg-disjoint
index 1
RP/0/RP0/CPU0:router(config-isis)# interface TenGigE0/0/0/0
RP/0/RP0/CPU0:router(config-isis-if)# point-to-point
RP/0/RP0/CPU0:router(config-isis-if)# address-family ipv4 unicast
RP/0/RP0/CPU0:router(config-isis-if-af)# fast-reroute per-prefix
RP/0/RP0/CPU0:router(config-isis-if-af)# fast-reroute per-prefix ti-lfa
RP/0/RP0/CPU0:router(config-isis)# srlg
RP/0/RP0/CPU0:router(config-isis-srlg)# name group1
RP/0/RP0/CPU0:router(config-isis-srlg-name)# admin-weight 5000
RP/0/RP0/CPU0:router(config-isis-srlg-name)# static ipv4 address 10.0.4.1 next-hop ipv4
address 10.0.4.2
RP/0/RP0/CPU0:router(config-isis-srlg-name)# static ipv4 address 10.0.4.2 next-hop ipv4
address 10.0.4.1
```



CHAPTER 9

Configure Segment Routing Microloop Avoidance

The Segment Routing Microloop Avoidance feature enables link-state routing protocols, such as IS-IS and OSPF, to prevent or avoid microloops during network convergence after a topology change.

About Segment Routing Microloop Avoidance

IP hop-by-hop routing may induce microloops (uLoops) at any topology transition. Microloops are a day-one IP challenge. Microloops are brief packet loops that occur in the network following a topology change (link down, link up, or metric change events). Microloops are caused by the non-simultaneous convergence of different nodes in the network. If a node converges and sends traffic to a neighbor node that has not converged yet, traffic may be looped between these two nodes, resulting in packet loss, jitter, and out-of-order packets.

Segment Routing resolves the microloop problem. A router with the Segment Routing Microloop Avoidance feature detects if microloops are possible for a destination on the post-convergence path following a topology change associated with a remote link event.

If a node determines that a microloop could occur on the new topology, the IGP computes a microloop-avoidant path to steer the traffic to that destination loop-free over the post-convergence path.

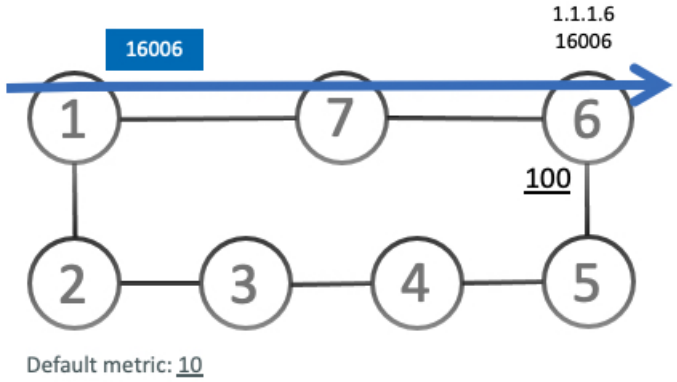
The IGP updates the forwarding table and temporarily (based on a RIB update delay timer) installs the SID-list imposition entries associated with the microloop-avoidant path for the destination with possible microloops.

After the RIB update delay timer expires, IGP updates the forwarding table, removing the microloop-avoidant SID list and traffic now natively follows the post-convergence path.

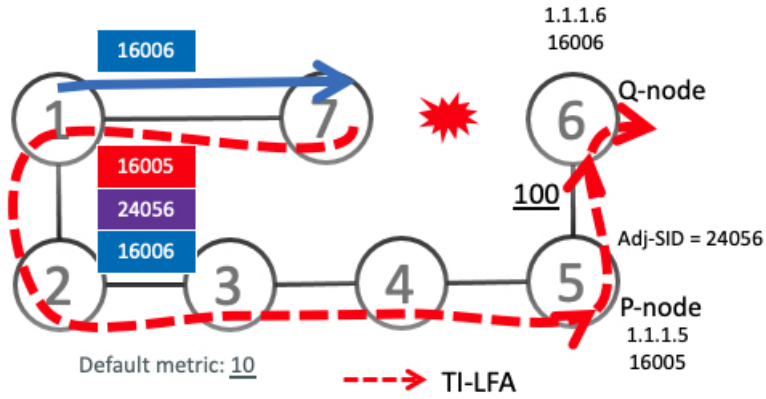
SR microloop avoidance is a local behavior and therefore not all nodes need to implement it to get the benefits.

In the topology below, microloops can occur after the failure of the link between Node6 and Node7.

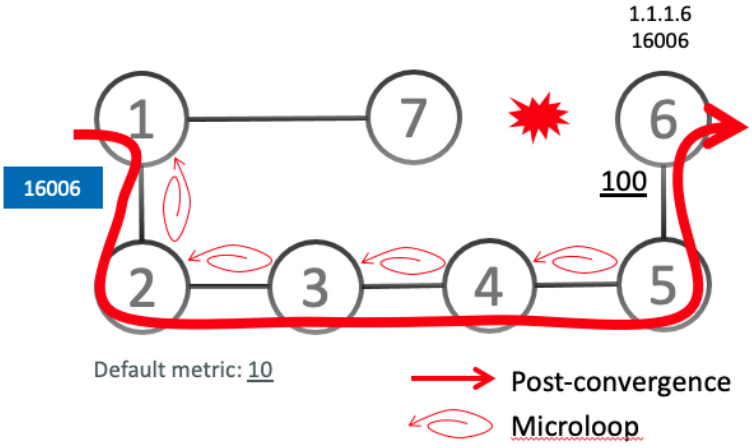
At steady state, Node1 sends traffic to node 6 (16006) via Node7. Node 7 is configured with TI-LFA to protect traffic to Node6.



TI-LFA on Node7 pre-computes a backup path for traffic to Node6 (prefix SID 16006) that will be activated if the link between Node7 and Node6 goes down. In this network, the backup path would steer traffic toward Node5 (prefix SID 16005) and then via link between Node5 and Node6 (adj-SID 24056). All nodes are notified of the topology change due to the link failure.



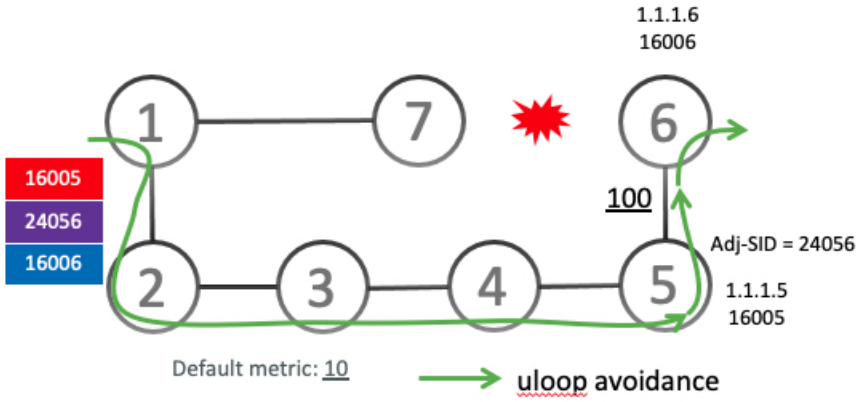
However, if nodes along the path do not converge at the same time, microloops can be introduced. For example, if Node2 converged before Node3, Node3 would send traffic back to Node2 as the shortest IGP path to Node6. The traffic between Node2 and Node3 creates a microloop.



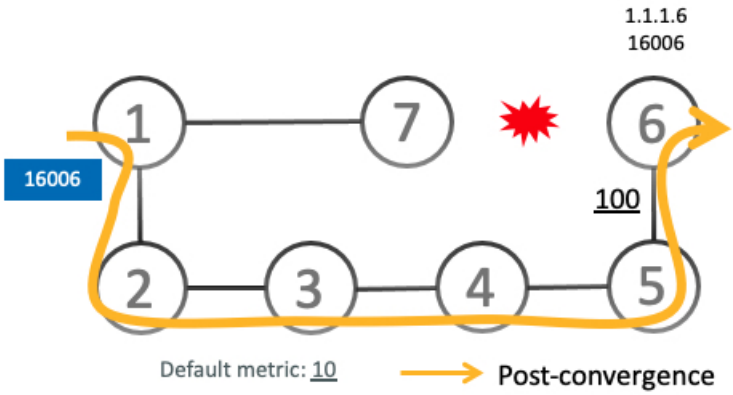
With microloop avoidance configured on Node1, a post-convergence path is computed and possible microloops on the post-convergence path for any destination are detected.

If microloops are possible on the post-convergence path to Node6, a microloop-avoidant path is constructed to steer the traffic to Node6 loop-free over the microloop-avoidant path {16005, 24056, 16006}.

Node1 updates the forwarding table and installs the SID-list imposition entries for those destinations with possible microloops, such as Node6. All nodes converge and update their forwarding tables, using SID lists where needed.



After the RIB update delay timer expires, the microloop-avoidant path is replaced with regular forwarding paths; traffic now natively follows the post-convergence path.



Configure Segment Routing Microloop Avoidance for IS-IS

This task describes how to enable Segment Routing Microloop Avoidance and set the Routing Information Base (RIB) update delay value for IS-IS.

Before you begin

Ensure that the following topology requirements are met:

- Router interfaces are configured as per the topology.
- Routers are configured with IS-IS.

- Segment routing for IS-IS is configured. See Enabling Segment Routing for IS-IS Protocol, on page 15.

SUMMARY STEPS

1. **configure**
2. **router isis** *instance-id*
3. **address-family** {**ipv4** | **ipv6**} [**unicast**]
4. **microloop avoidance segment-routing**
5. **microloop avoidance rib-update-delay** *delay-time*

DETAILED STEPS

	Command or Action	Purpose
Step 1	**configure** **Example:** `RP/0/RP0/CPU0:router# configure`	Enters mode.
Step 2	**router isis** *instance-id* **Example:** `RP/0/RP0/CPU0:router(config)# router isis 1`	Enables IS-IS routing for the specified routing instance, and places the router in router configuration mode. You can change the level of routing to be performed by a particular routing instance by using the **is-type** router configuration command.
Step 3	**address-family** {**ipv4** | **ipv6**} [**unicast**] **Example:** `RP/0/RP0/CPU0:router(config-isis)# address-family ipv4 unicast`	Specifies the IPv4 or IPv6 address family and enters router address family configuration mode.
Step 4	**microloop avoidance segment-routing** **Example:** `RP/0/RP0/CPU0:router(config-isis-af)# microloop avoidance segment-routing`	Enables Segment Routing Microloop Avoidance.
Step 5	**microloop avoidance rib-update-delay** *delay-time* **Example:** `RP/0/RP0/CPU0:router(config-isis-af)# microloop avoidance rib-update-delay 3000`	Specifies the amount of time the node uses the microloop avoidance policy before updating its forwarding table. The *delay-time* is in milliseconds. The range is from 1-60000. The default value is 5000.

Configure Segment Routing Microloop Avoidance for OSPF

This task describes how to enable Segment Routing Microloop Avoidance and set the Routing Information Base (RIB) update delay value for OSPF.

Before you begin

Ensure that the following topology requirements are met:

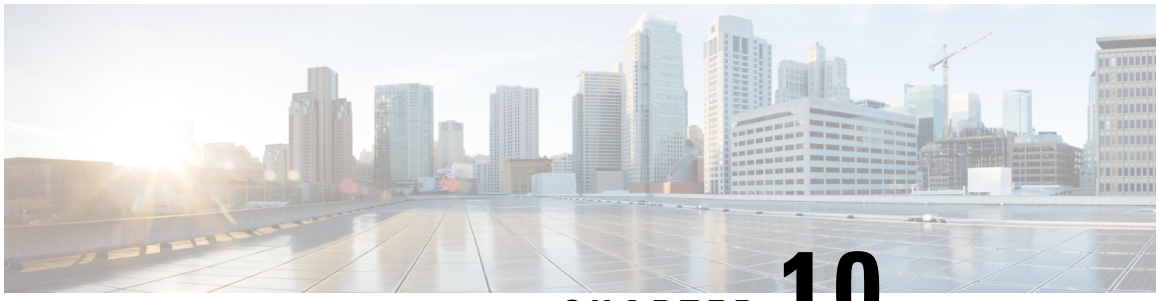
- Router interfaces are configured as per the topology.
- Routers are configured with OSPF.
- Segment routing for OSPF is configured. See Enabling Segment Routing for OSPF Protocol, on page 41.

SUMMARY STEPS

1. **configure**
2. **router ospf** *process-name*
3. **microloop avoidance segment-routing**
4. **microloop avoidance rib-update-delay** *delay-time*

DETAILED STEPS

	Command or Action	Purpose
Step 1	**configure** **Example:** `RP/0/RP0/CPU0:router# configure`	Enters mode.
Step 2	**router ospf** *process-name* **Example:** `RP/0/RP0/CPU0:router(config)# router ospf 1`	Enables OSPF routing for the specified routing process, and places the router in router configuration mode.
Step 3	**microloop avoidance segment-routing** **Example:** `RP/0/RP0/CPU0:router(config-ospf)# microloop avoidance segment-routing`	Enables Segment Routing Microloop Avoidance.
Step 4	**microloop avoidance rib-update-delay** *delay-time* **Example:** `RP/0/RP0/CPU0:router(config-ospf)# microloop avoidance rib-update-delay 3000`	Specifies the amount of time the node uses the microloop avoidance policy before updating its forwarding table. The *delay-time* is in milliseconds. The range is from 1-60000. The default value is 5000.



CHAPTER 10

Configure Segment Routing Mapping Server

The mapping server is a key component of the interworking between LDP and segment routing. It enables SR-capable nodes to interwork with LDP nodes. The mapping server advertises Prefix-to-SID mappings in IGP on behalf of other non-SR-capable nodes.

Segment Routing Mapping Server

The mapping server functionality in Cisco IOS XR segment routing centrally assigns prefix-SIDs for some or all of the known prefixes. A router must be able to act as a mapping server, a mapping client, or both.

- A router that acts as a mapping server allows the user to configure SID mapping entries to specify the prefix-SIDs for some or all prefixes. This creates the local SID-mapping policy. The local SID-mapping policy contains non-overlapping SID-mapping entries. The mapping server advertises the local SID-mapping policy to the mapping clients.
- A router that acts as a mapping client receives and parses remotely received SIDs from the mapping server to create remote SID-mapping entries.
- A router that acts as a mapping server and mapping client uses the remotely learnt and locally configured mapping entries to construct the non-overlapping consistent active mapping policy. IGP instance uses the active mapping policy to calculate the prefix-SIDs of some or all prefixes.

The mapping server automatically manages the insertions and deletions of mapping entries to always yield an active mapping policy that contains non-overlapping consistent SID-mapping entries.

- Locally configured mapping entries must not overlap each other.
- The mapping server takes the locally configured mapping policy, as well as remotely learned mapping entries from a particular IGP instance, as input, and selects a single mapping entry among overlapping mapping entries according to the preference rules for that IGP instance. The result is an active mapping policy that consists of non-overlapping consistent mapping entries.
- At steady state, all routers, at least in the same area or level, must have identical active mapping policies.

Segment Routing Mapping Server Restrictions

- The position of the mapping server in the network is not important. However, since the mapping advertisements are distributed in IGP using the regular IGP advertisement mechanism, the mapping server needs an IGP adjacency to the network.
- The role of the mapping server is crucial. For redundancy purposes, you should configure multiple mapping servers in the networks.
- The mapping server functionality does not support a scenario where SID-mapping entries learned through one IS-IS instance are used by another IS-IS instance to determine the prefix-SID of a prefix. For example, mapping entries learnt from remote routers by 'router isis 1' cannot be used to calculate prefix-SIDs for prefixes learnt, advertised, or downloaded to FIB by 'router isis 2'. A mapping server is required for each IS-IS area.
- Segment Routing Mapping Server does not support Virtual Routing and Forwarding (VRF) currently.

Segment Routing and LDP Interoperability

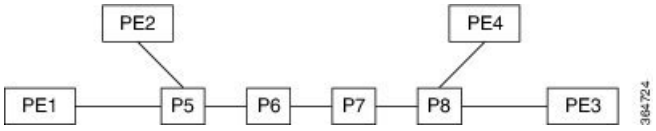
IGP provides mechanisms through which segment routing (SR) interoperate with label distribution protocol (LDP). The control plane of segment routing co-exists with LDP.

The Segment Routing Mapping Server (SRMS) functionality in SR is used to advertise SIDs for destinations, in the LDP part of the network, that do not support SR. SRMS maintains and advertises segment identifier (SID) mapping entries for such destinations. IGP propagates the SRMS mapping entries and interacts with SRMS to determine the SID value when programming the forwarding plane. IGP installs prefixes and corresponding labels, into routing information base (RIB), that are used to program the forwarding information base (FIB).

Example: Segment Routing LDP Interoperability

Consider a network with a mix of segment routing (SR) and label distribution protocol (LDP). A continuous multiprotocol label switching (MPLS) LSP (Labeled Switched Path) can be established by facilitating interoperability. One or more nodes in the SR domain act as segment routing mapping server (SRMS). SRMS advertises SID mappings on behalf of non-SR capable nodes. Each SR-capable node learns about SID assigned to non-SR capable nodes without explicitly configuring individual nodes.

Consider a network as shown in the following image. This network is a mix of both LDP and SR-capable nodes.



In this mixed network:

- Nodes P6, P7, P8, PE4 and PE3 are LDP-capable
- Nodes PE1, PE2, P5 and P6 are SR-capable
- Nodes PE1, PE2, P5 and P6 are configured with segment routing global block (SRGB) of (100, 200)
- Nodes PE1, PE2, P5 and P6 are configured with node segments of 101, 102, 105 and 106 respectively

A service flow must be established from PE1 to PE3 over a continuous MPLS tunnel. This requires SR and LDP to interoperate.

LDP to SR

The traffic flow from LDP to SR (right to left) involves:

1. PE3 learns a service route whose nhop is PE1. PE3 has an LDP label binding from the nhop P8 for the FEC PE1. PE3 forwards the packet P8.
2. P8 has an LDP label binding from its nhop P7 for the FEC PE1. P8 forwards the packet to P7.
3. P7 has an LDP label binding from its nhop P6 for the FEC PE1. P7 forwards the packet to P6.
4. P6 does not have an LDP binding from its nhop P5 for the FEC PE1. But P6 has an SR node segment to the IGP route PE1. P6 forwards the packet to P5 and swaps its local LDP label for FEC PE1 by the equivalent node segment 101. This process is called label merging.
5. P5 pops 101, assuming PE1 has advertised its node segment 101 with the penultimate-pop flag set and forwards to PE1.
6. PE1 receives the tunneled packet and processes the service label.

The end-to-end MPLS tunnel is established from an LDP LSP from PE3 to P6 and the related node segment from P6 to PE1.

SR to LDP

Suppose that the operator configures P5 as a Segment Routing Mapping Server (SRMS) and advertises the mappings (P7, 107), (P8, 108), (PE3, 103) and (PE4, 104). If PE3 was SR-capable, the operator may have configured PE3 with node segment 103. Because PE3 is non-SR capable, the operator configures that policy at the SRMS; the SRMS advertises the mapping on behalf of the non-SR capable nodes. Multiple SRMS servers can be provisioned in a network for redundancy. The mapping server advertisements are only understood by the SR-capable nodes. The SR capable routers install the related node segments in the MPLS data plane in exactly the same manner if node segments were advertised by the nodes themselves.

The traffic flow from SR to LDP (left to right) involves:

1. PE1 installs the node segment 103 with nhop P5 in exactly the same manner if PE3 had advertised node segment 103.
2. P5 swaps 103 for 103 and forwards to P6.
3. The nhop for P6 for the IGP route PE3 is non-SR capable. (P7 does not advertise the SR capability.) However, P6 has an LDP label binding from that nhop for the same FEC. (For example, LDP label 1037.) P6 swaps 103 for 1037 and forwards to P7. We refer to this process as label merging.
4. P7 swaps this label with the LDP label received from P8 and forwards to P8.
5. P8 pops the LDP label and forwards to PE3.
6. PE3 receives the packet and processes as required.

The end-to-end MPLS LSP is established from an SR node segment from PE1 to P6 and an LDP LSP from P6 to PE3.

Configuring Mapping Server

Perform these tasks to configure the mapping server and to add prefix-SID mapping entries in the active local mapping policy.

SUMMARY STEPS

1. **configure**
2. **segment-routing**
3. **mapping-server**
4. **prefix-sid-map**
5. **address-family ipv4** | **ipv6**
6. *ip-address/prefix-length first-SID-value* **range** *range*
7. Use the **commit** or **end** command.
8. **show segment-routing mapping-server prefix-sid-map** [**ipv4** | **ipv6**] [**detail**]

DETAILED STEPS

	Command or Action	Purpose
Step 1	**configure** **Example:** `RP/0/RP0/CPU0:router# configure`	Enters mode.
Step 2	**segment-routing** **Example:** `RP/0/RP0/CPU0:router(config)# segment-routing`	Enables segment routing.
Step 3	**mapping-server** **Example:** `RP/0/RP0/CPU0:router(config-sr)# mapping-server`	Enables mapping server configuration mode.
Step 4	**prefix-sid-map** **Example:** `RP/0/RP0/CPU0:router(config-sr-ms)# prefix-sid-map`	Enables prefix-SID mapping configuration mode. **Note** Two-way prefix SID can be enabled directly under IS-IS or through a mapping server.
Step 5	**address-family ipv4**	**ipv6** **Example:** This example shows the address-family for ipv4: `RP/0/RP0/CPU0:router(config-sr-ms-map)# address-family ipv4`	Configures address-family for IS-IS.

	Command or Action	Purpose
	This example shows the address-family for ipv6: `RP/0/RP0/CPU0:router(config-sr-ms-map)#` **`address-family ipv6`**	
Step 6	*ip-address/prefix-length first-SID-value* **range** *range* **Example:** `RP/0/RP0/CPU0:router(config-sr-ms-map-af)#` **`10.1.1.1/32 10 range 200`** `RP/0/RP0/CPU0:router(config-sr-ms-map-af)#` **`20.1.0.0/16 400 range 300`**	Adds SID-mapping entries in the active local mapping policy. In the configured example: • Prefix 10.1.1.1/32 is assigned prefix-SID 10, prefix 10.1.1.2/32 is assigned prefix-SID 11,…, prefix 10.1.1.199/32 is assigned prefix-SID 200 • Prefix 20.1.0.0/16 is assigned prefix-SID 400, prefix 20.2.0.0/16 is assigned prefix-SID 401,…, and so on.
Step 7	Use the **commit** or **end** command.	**commit** —Saves the configuration changes and remains within the configuration session. **end** —Prompts user to take one of these actions: • **Yes** — Saves configuration changes and exits the configuration session. • **No** —Exits the configuration session without committing the configuration changes. • **Cancel** —Remains in the configuration session, without committing the configuration changes.

Step 8	**show segment-routing mapping-server prefix-sid-map** [**ipv4** | **ipv6**] [**detail**]
	Example:
	
	`RP/0/RP0/CPU0:router#` **`show segment-routing mapping-server prefix-sid-map ipv4`**
	`Prefix SID Index Range Flags`
	`20.1.1.0/24 400 300`
	`10.1.1.1/32 10 200`
	
	`Number of mapping entries: 2`
	
	`RP/0/RP0/CPU0:router#` **`show segment-routing mapping-server prefix-sid-map ipv4 detail`**
	`Prefix`
	`20.1.1.0/24`
	`SID Index:` **`400`**
	`Range:` **`300`**
	`Last Prefix: 20.2.44.0/24`
	`Last SID Index: 699`
	`Flags:`
	`10.1.1.1/32`
	`SID Index:` **`10`**
	`Range:` **`200`**
	`Last Prefix: 10.1.1.200/32`
	`Last SID Index: 209`
		Displays information about the locally configured prefix-to-SID mappings.
		Note Specify the address family for IS-IS.

	Command or Action	Purpose
	Flags: Number of mapping entries: 2	

What to do next

Enable the advertisement of the local SID-mapping policy in the IGP.

Enable Mapping Advertisement

In addition to configuring the static mapping policy, you must enable the advertisement of the mappings in the IGP.

Perform these steps to enable the IGP to advertise the locally configured prefix-SID mapping.

Configure Mapping Advertisement for IS-IS

SUMMARY STEPS

1. **router isis** *instance-id*
2. **address-family** { **ipv4** | **ipv6** } [**unicast**]
3. **segment-routing prefix-sid-map advertise-local**
4. Use the **commit** or **end** command.
5. **show isis database verbose**

DETAILED STEPS

	Command or Action	Purpose
Step 1	**router isis** *instance-id* **Example:** `RP/0/RP0/CPU0:router(config)# router isis 1`	Enables IS-IS routing for the specified routing instance, and places the router in router configuration mode. • You can change the level of routing to be performed by a particular routing instance by using the **is-type** router configuration command.
Step 2	**address-family** { **ipv4** | **ipv6** } [**unicast**] **Example:** The following is an example for ipv4 address family: `RP/0/RP0/CPU0:router(config-isis)# address-family ipv4 unicast`	Specifies the IPv4 or IPv6 address family, and enters router address family configuration mode.
Step 3	**segment-routing prefix-sid-map advertise-local** **Example:**	Configures IS-IS to advertise locally configured prefix-SID mappings.

	Command or Action	Purpose
	`RP/0/RP0/CPU0:router(config-isis-af)# segment-routing prefix-sid-map advertise-local`	
Step 4	Use the **commit** or **end** command.	**commit** —Saves the configuration changes and remains within the configuration session. **end** —Prompts user to take one of these actions: • **Yes** — Saves configuration changes and exits the configuration session. • **No** —Exits the configuration session without committing the configuration changes. • **Cancel** —Remains in the configuration session, without committing the configuration changes.

Step 5	**show isis database verbose**
	Example:
	`RP/0/RP0/CPU0:router# show isis database verbose`
	`<...removed...>`
	`SID Binding: 10.1.1.1/32 F:0 M:0 S:0 D:0 A:0 Weight:0 Range:200`
	`SID: Start:10, Algorithm:0, R:0 N:0 P:0 E:0 V:0 L:0`
	`SID Binding: 20.1.1.0/24 F:0 M:0 S:0 D:0 A:0 Weight:0 Range:300`
	`SID: Start:400, Algorithm:0, R:0 N:0 P:0 E:0 V:0 L:0`
		Displays IS-IS prefix-SID mapping advertisement and TLV.

Configure Mapping Advertisement for OSPF

SUMMARY STEPS

1. **router ospf** *process-name*
2. **segment-routing prefix-sid-map advertise-local**
3. Use the **commit** or **end** command.
4. **show ospf database opaque-area**

DETAILED STEPS

	Command or Action	Purpose
Step 1	**router ospf** *process-name* **Example:** `RP/0/RP0/CPU0:router(config)# router ospf 1`	Enables OSPF routing for the specified routing instance, and places the router in router configuration mode.

	Command or Action	Purpose
Step 2	**segment-routing prefix-sid-map advertise-local** **Example:** `RP/0/RP0/CPU0:router(config-ospf)# segment-routing prefix-sid-map advertise-local`	Configures OSPF to advertise locally configured prefix-SID mappings.
Step 3	Use the **commit** or **end** command.	**commit** —Saves the configuration changes and remains within the configuration session. **end** —Prompts user to take one of these actions: • **Yes** — Saves configuration changes and exits the configuration session. • **No** —Exits the configuration session without committing the configuration changes. • **Cancel** —Remains in the configuration session, without committing the configuration changes.

Step 4	**show ospf database opaque-area**
	Example:
	`RP/0/RP0/CPU0:router# show ospf database opaque-area`
	`<...removed...>`
	`Extended Prefix Range TLV: Length: 24`
	`AF : 0`
	`Prefix : 10.1.1.1/32`
	`Range Size: 200`
	`Flags : 0x0`
	`SID sub-TLV: Length: 8`
	`Flags : 0x60`
	`MTID : 0`
	`Algo : 0`
	`SID Index : 10`
		Displays OSP prefix-SID mapping advertisement and TLV.

Enable Mapping Client

By default, mapping client functionality is enabled.

You can disable the mapping client functionality by using the **segment-routing prefix-sid-map receive disable** command.

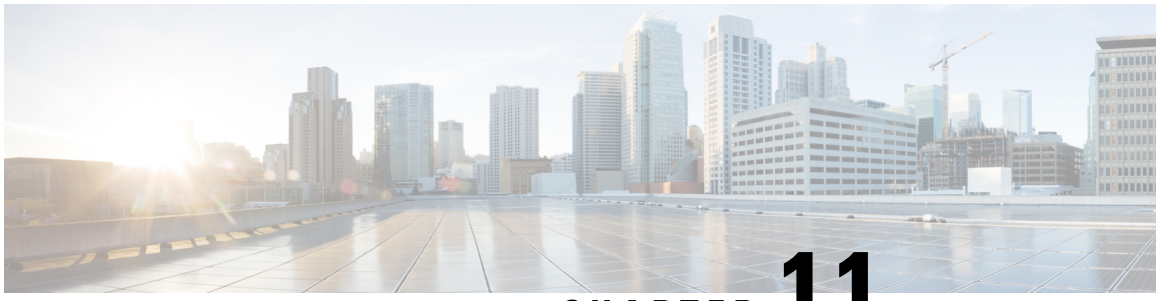
You can re-enable the mapping client functionality by using the **segment-routing prefix-sid-map receive** command.

The following example shows how to enable the mapping client for IS-IS:

```
RP/0/RP0/CPU0:router(config)# router isis 1
RP/0/RP0/CPU0:router(config-isis)# address-family ipv4 unicast
RP/0/RP0/CPU0:router(config-isis-af)# segment-routing prefix-sid-map receive
```

The following example shows how to enable the mapping client for OSPF:

```
RP/0/RP0/CPU0:router(config)# router ospf 1
RP/0/RP0/CPU0:router(config-ospf)# segment-routing prefix-sid-map receive
```



CHAPTER 11

Using Segment Routing OAM

Segment Routing Operations, Administration, and Maintenance (OAM) helps service providers to monitor label-switched paths (LSPs) and quickly isolate forwarding problems to assist with fault detection and troubleshooting in the network. The Segment Routing OAM feature provides support for BGP prefix SID, Nil-FEC (forwarding equivalence classes) LSP Ping and Traceroute functionality.

MPLS Ping and Traceroute for BGP and IGP Prefix-SID

MPLS Ping and Traceroute operations for Prefix SID are supported for various BGP and IGP scenarios, for example:

- Within an IS-IS level or OSPF area
- Across IS-IS levels or OSPF areas
- Route redistribution from IS-IS to OSPF and from OSPF to IS-IS
- Anycast Prefix SID
- Combinations of BGP and LDP signaled LSPs

The MPLS LSP Ping feature is used to check the connectivity between ingress Label Switch Routers (LSRs) and egress LSRs along an LSP. MPLS LSP ping uses MPLS echo request and reply messages, similar to Internet Control Message Protocol (ICMP) echo request and reply messages, to validate an LSP. The destination IP address of the MPLS echo request packet is different from the address used to select the label stack. The destination IP address is defined as a 127.x.y.z/8 address and it prevents the IP packet from being IP switched to its destination, if the LSP is broken.

The MPLS LSP Traceroute feature is used to isolate the failure point of an LSP. It is used for hop-by-hop fault localization and path tracing. The MPLS LSP Traceroute feature relies on the expiration of the Time to Live (TTL) value of the packet that carries the echo request. When the MPLS echo request message hits a transit node, it checks the TTL value and if it is expired, the packet is passed to the control plane, else the

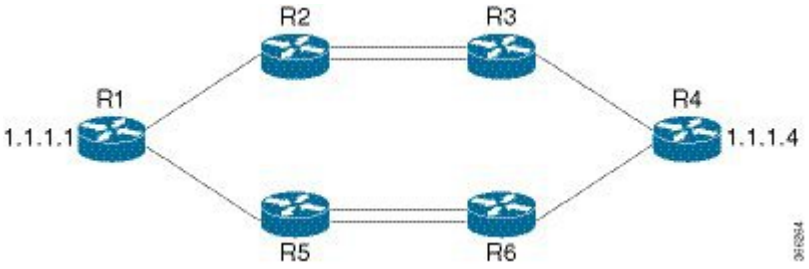
message is forwarded. If the echo message is passed to the control plane, a reply message is generated based on the contents of the request message.

The MPLS LSP Tree Trace (traceroute multipath) operation is also supported for BGP and IGP Prefix SID. MPLS LSP Tree Trace provides the means to discover all possible equal-cost multipath (ECMP) routing paths of an LSP to reach a destination Prefix SID. It uses multipath data encoded in echo request packets to query for the load-balancing information that may allow the originator to exercise each ECMP. When the packet TTL expires at the responding node, the node returns the list of downstream paths, as well as the multipath information that can lead the operator to exercise each path in the MPLS echo reply. This operation is performed repeatedly for each hop of each path with increasing TTL values until all ECMP are discovered and validated.

MPLS echo request packets carry Target FEC Stack sub-TLVs. The Target FEC sub-TLVs are used by the responder for FEC validation. The BGP and IGP IPv4 prefix sub-TLV has been added to the Target FEC Stack sub-TLV. The IGP IPv4 prefix sub-TLV contains the prefix SID, the prefix length, and the protocol (IS-IS or OSPF). The BGP IPv4 prefix sub-TLV contains the prefix SID and the prefix length.

Examples: MPLS Ping, Traceroute, and Tree Trace for Prefix-SID

These examples use the following topology:



MPLS Ping for Prefix-SID

```
RP/0/RP0/CPU0:router-arizona# ping mpls ipv4 1.1.1.4/32
Thu Dec 17 01:01:42.301 PST

Sending 5, 100-byte MPLS Echos to 1.1.1.4,
      timeout is 2 seconds, send interval is 0 msec:

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
  'L' - labeled output interface, 'B' - unlabeled output interface,
  'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
  'M' - malformed request, 'm' - unsupported tlvs, 'N' - no rx label,
  'P' - no rx intf label prot, 'p' - premature termination of LSP,
  'R' - transit router, 'I' - unknown upstream index,
  'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.

!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 2/2/3 ms
```

MPLS Traceroute for Prefix-SID

```
RP/0/RP0/CPU0:router-arizona# traceroute mpls ipv4 1.1.1.4/32
Thu Dec 17 14:45:05.563 PST
```

```
Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
  'L' - labeled output interface, 'B' - unlabeled output interface,
  'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
  'M' - malformed request, 'm' - unsupported tlvs, 'N' - no rx label,
  'P' - no rx intf label prot, 'p' - premature termination of LSP,
  'R' - transit router, 'I' - unknown upstream index,
  'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.

  0 12.12.12.1 MRU 4470 [Labels: 16004 Exp: 0]
L 1 12.12.12.2 MRU 4470 [Labels: 16004 Exp: 0] 3 ms
L 2 23.23.23.3 MRU 4470 [Labels: implicit-null Exp: 0] 3 ms
! 3 34.34.34.4 11 ms
```

MPLS Tree Trace for Prefix-SID

```
RP/0/RP0/CPU0:router-arizona# traceroute mpls multipath ipv4 1.1.1.4/32
Thu Dec 17 14:55:46.549 PST

Starting LSP Path Discovery for 1.1.1.4/32

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
  'L' - labeled output interface, 'B' - unlabeled output interface,
  'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
  'M' - malformed request, 'm' - unsupported tlvs, 'N' - no rx label,
  'P' - no rx intf label prot, 'p' - premature termination of LSP,
  'R' - transit router, 'I' - unknown upstream index,
  'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.

LL!
Path 0 found,
 output interface TenGigE0/0/0/0 nexthop 12.12.12.2  source 12.12.12.1 destination 127.0.0.0
 L!
Path 1 found,
 output interface TenGigE0/0/0/0 nexthop 12.12.12.2  source 12.12.12.1 destination 127.0.0.2
 LL!
Path 2 found,
 output interface TenGigE0/0/0/1 nexthop 15.15.15.5  source 15.15.15.1 destination 127.0.0.1
 L!
Path 3 found,
 output interface TenGigE0/0/0/1 nexthop 15.15.15.5  source 15.15.15.1 destination 127.0.0.0

Paths (found/broken/unexplored) (4/0/0)
 Echo Request (sent/fail) (10/0)
 Echo Reply (received/timeout) (10/0)
 Total Time Elapsed 53 ms
```

MPLS LSP Ping and Traceroute Nil FEC Target

The Nil-FEC LSP ping and traceroute operations are extensions of regular MPLS ping and traceroute.

Nil-FEC LSP Ping/Traceroute functionality supports segment routing and MPLS Static. It also acts as an additional diagnostic tool for all other LSP types. This feature allows operators to provide the ability to freely test any label stack by allowing them to specify the following:

- label stack
- outgoing interface
- nexthop address

In the case of segment routing, each segment nodal label and adjacency label along the routing path is put into the label stack of an echo request message from the initiator Label Switch Router (LSR); MPLS data plane forwards this packet to the label stack target, and the label stack target sends the echo message back.

The following table shows the syntax for the ping and traceroute commands.

Table 17: LSP Ping and Traceroute Nil FEC Commands

Command Syntax
ping mpls nil-fec labels {*label*[**,***label*]} [**output** {**interface** *tx-interface*} [**nexthop** *nexthop-ip-addr*]]
traceroute mpls nil-fec labels {*label*[**,***label*]} [**output** {**interface** *tx-interface*} [**nexthop** *nexthop-ip-addr*]]

Examples: LSP Ping and Traceroute for Nil_FEC Target

These examples use the following topology:

```
Node loopback IP address: 172.18.1.3   172.18.1.4   172.18.1.5   172.18.1.7
Node label:                            16004        16005        16007
Nodes:                    Arizona ---- Utah ------- Wyoming ---- Texas

Interface:            GigabitEthernet0/0/0/1   GigabitEthernet0/0/0/1
Interface IP address:         10.1.1.3              10.1.1.4
```

```
RP/0/RP0/CPU0:router-utah# show mpls forwarding

Tue Jul  5 13:44:31.999 EDT
Local  Outgoing    Prefix             Outgoing     Next Hop        Bytes
Label  Label       or ID              Interface                    Switched
------ ----------- ------------------ ------------ --------------- ------------
16004  Pop         No ID              Gi0/0/0/1    10.1.1.4        1392
       Pop         No ID              Gi0/0/0/2    10.1.2.2        0
16005  16005       No ID              Gi0/0/0/0    10.1.1.4        0
       16005       No ID              Gi0/0/0/1    10.1.2.2        0
16007  16007       No ID              Gi0/0/0/0    10.1.1.4        4752
       16007       No ID              Gi0/0/0/1    10.1.2.2        0
24000  Pop         SR Adj (idx 0)     Gi0/0/0/0    10.1.1.4        0
24001  Pop         SR Adj (idx 2)     Gi0/0/0/1    10.1.1.4        0
24002  Pop         SR Adj (idx 0)     Gi0/0/0/1    10.1.2.2        0
24003  Pop         SR Adj (idx 2)     Gi0/0/0/1    10.1.2.2        0
24004  Pop         No ID              tt10         point2point     0
24005  Pop         No ID              tt11         point2point     0
24006  Pop         No ID              tt12         point2point     0
24007  Pop         No ID              tt13         point2point     0
24008  Pop         No ID              tt30         point2point     0
```

Ping Nil FEC Target

```
RP/0/RP0/CPU0:router-arizona# ping mpls nil-fec labels 16005,16007 output interface
GigabitEthernet 0/0/0/1 nexthop 10.1.1.4 repeat 1
Sending 1, 72-byte MPLS Echos with Nil FEC labels 16005,16007,
     timeout is 2 seconds, send interval is 0 msec:

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
  'L' - labeled output interface, 'B' - unlabeled output interface,
  'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
  'M' - malformed request, 'm' - unsupported tlvs, 'N' - no label entry,
  'P' - no rx intf label prot, 'p' - premature termination of LSP,
  'R' - transit router, 'I' - unknown upstream index,
  'd' - see DDMAP for return code,
  'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.
!
Success rate is 100 percent (1/1), round-trip min/avg/max = 1/1/1 ms
 Total Time Elapsed 0 ms
```

Traceroute Nil FEC Target

```
RP/0/RP0/CPU0:router-arizona# traceroute mpls nil-fec labels 16005,16007 output interface
GigabitEthernet 0/0/0/1 nexthop 10.1.1.4
Tracing MPLS Label Switched Path with Nil FEC labels 16005,16007, timeout is 2 seconds

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
  'L' - labeled output interface, 'B' - unlabeled output interface,
  'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
  'M' - malformed request, 'm' - unsupported tlvs, 'N' - no label entry,
  'P' - no rx intf label prot, 'p' - premature termination of LSP,
  'R' - transit router, 'I' - unknown upstream index,
  'd' - see DDMAP for return code,
  'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.
  0 10.1.1.3 MRU 1500 [Labels: 16005/16007/explicit-null Exp: 0/0/0]
L 1 10.1.1.4 MRU 1500 [Labels: implicit-null/16007/explicit-null Exp: 0/0/0] 1 ms
L 2 10.1.1.5 MRU 1500 [Labels: implicit-null/explicit-null Exp: 0/0] 1 ms
! 3 10.1.1.7 1 ms
```

Segment Routing Ping and Traceroute

Table 18: Feature History Table

Feature Name	Release Information	Feature Description
SR OAM for SR Policy (Policy Name / Binding SID / Custom label stack)	Release 7.3.1	This feature extends SR OAM ping and traceroute function for an SR policy (or binding SID)-LSP end-point combination. This addresses the limitations of the Nil-FEC LSP Ping and Traceroute function which cannot perform a ping operation to a segment list that is not associated with an installed SR policy. Also, it cannot validate egress device-specific SR policies.

Segment Routing Ping

The MPLS LSP ping feature is used to check the connectivity between ingress and egress of LSP. MPLS LSP ping uses MPLS echo request and reply messages, similar to Internet Control Message Protocol (ICMP) echo request and reply messages, to validate an LSP. Segment routing ping is an extension of the MPLS LSP ping to perform the connectivity verification on the segment routing control plane.

Note Segment routing ping can only be used when the originating device is running segment routing.

You can initiate the segment routing ping operation only when Segment Routing control plane is available at the originator, even if it is not preferred. This allows you to validate the SR path before directing traffic over the path. Segment Routing ping can use either generic FEC type or SR control-plane FEC type (SR-OSPF, SR-ISIS). In mixed networks, where some devices are running MPLS control plane (for example, LDP) or do not understand SR FEC, generic FEC type allows the device to successfully process and respond to the echo request. By default, generic FEC type is used in the target FEC stack of segment routing ping echo request. Generic FEC is not coupled to a particular control plane; it allows path verification when the advertising protocol is unknown or might change during the path of the echo request. If you need to specify the target FEC, you can select the FEC type as OSPF, IS-IS, or BGP. This ensures that only devices that are running segment routing control plane, and can therefore understand the segment routing IGP FEC, respond to the echo request.

Configuration Examples

These examples show how to use segment routing ping to test the connectivity of a segment routing control plane. In the first example, FEC type is not specified. You can also specify the FEC type as shown in the other examples.

```
RP/0/RP0/CPU0:router# ping sr-mpls 10.1.1.2/32

Sending 5, 100-byte MPLS Echos to 10.1.1.2/32,
      timeout is 2 seconds, send interval is 0 msec:

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
  'L' - labeled output interface, 'B' - unlabeled output interface,
```

```
  'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
  'M' - malformed request, 'm' - unsupported tlvs, 'N' - no rx label,
  'P' - no rx intf label prot, 'p' - premature termination of LSP,
  'R' - transit router, 'I' - unknown upstream index,
  'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.

!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/2/5 ms

RP/0/RP0/CPU0:router# ping sr-mpls 10.1.1.2/32 fec-type generic

Sending 5, 100-byte MPLS Echos to 10.1.1.2/32,
      timeout is 2 seconds, send interval is 0 msec:

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
  'L' - labeled output interface, 'B' - unlabeled output interface,
  'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
  'M' - malformed request, 'm' - unsupported tlvs, 'N' - no rx label,
  'P' - no rx intf label prot, 'p' - premature termination of LSP,
  'R' - transit router, 'I' - unknown upstream index,
  'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.

!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/2 ms


RP/0/RP0/CPU0:router# ping sr-mpls 10.1.1.2/32 fec-type igp ospf

Sending 5, 100-byte MPLS Echos to 10.1.1.2/32,
      timeout is 2 seconds, send interval is 0 msec:

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
  'L' - labeled output interface, 'B' - unlabeled output interface,
  'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
  'M' - malformed request, 'm' - unsupported tlvs, 'N' - no rx label,
  'P' - no rx intf label prot, 'p' - premature termination of LSP,
  'R' - transit router, 'I' - unknown upstream index,
  'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.

!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/2 ms


RP/0/RP0/CPU0:router# ping sr-mpls 10.1.1.2/32 fec-type igp isis

Sending 5, 100-byte MPLS Echos to 10.1.1.2/32,
      timeout is 2 seconds, send interval is 0 msec:

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
  'L' - labeled output interface, 'B' - unlabeled output interface,
  'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
  'M' - malformed request, 'm' - unsupported tlvs, 'N' - no rx label,
  'P' - no rx intf label prot, 'p' - premature termination of LSP,
  'R' - transit router, 'I' - unknown upstream index,
  'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.

!!!!!
```

```
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/2 ms


RP/0/RP0/CPU0:router# ping sr-mpls 10.1.1.2/32 fec-type bgp

Sending 5, 100-byte MPLS Echos to 10.1.1.2/32,
      timeout is 2 seconds, send interval is 0 msec:

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
  'L' - labeled output interface, 'B' - unlabeled output interface,
  'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
  'M' - malformed request, 'm' - unsupported tlvs, 'N' - no rx label,
  'P' - no rx intf label prot, 'p' - premature termination of LSP,
  'R' - transit router, 'I' - unknown upstream index,
  'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.

!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/2 ms
```

Ping for SR Policy

You can perform the ping operation for an SR policy (or binding SID), and LSP end-point combination. Use the **ping** command's **policy name lsp-end-point** and **policy binding-sid lsp-end-point** options to perform this task. You can instantiate the policy through the CLI, Netconf, PCEP or BGP-TE process.

IPv6 policies are not supported for SR OAM function.

Note As a prerequisite, you must enable the MPLS OAM function.

```
Router(config)# mpls oam
Router(config)# commit

Router# ping sr-mpls policy name srte_c_4_ep_10.0.0.1 lsp-end-point 209.165.201.1
Router# ping sr-mpls policy binding-sid 1000 lsp-end-point 209.165.201.1
```

Segment Routing Traceroute

The MPLS LSP traceroute is used to isolate the failure point of an LSP. It is used for hop-by-hop fault localization and path tracing. The MPLS LSP traceroute feature relies on the expiration of the Time to Live (TTL) value of the packet that carries the echo request. When the MPLS echo request message hits a transit node, it checks the TTL value and if it is expired, the packet is passed to the control plane, else the message is forwarded. If the echo message is passed to the control plane, a reply message is generated based on the contents of the request message. Segment routing traceroute feature extends the MPLS LSP traceroute functionality to segment routing networks.

Similar to segment routing ping, you can initiate the segment routing traceroute operation only when Segment Routing control plane is available at the originator, even if it is not preferred. Segment Routing traceroute can use either generic FEC type or SR control-plane FEC type (SR-OSPF, SR-ISIS). By default, generic FEC type is used in the target FEC stack of segment routing traceroute echo request. If you need to specify the target FEC, you can select the FEC type as OSPF, IS-IS, or BGP. This ensures that only devices that are running segment routing control plane, and can therefore understand the segment routing IGP FEC, respond to the echo request.

The existence of load balancing at routers in an MPLS network provides alternate paths for carrying MPLS traffic to a target router. The multipath segment routing traceroute feature provides a means to discover all possible paths of an LSP between the ingress and egress routers.

Configuration Examples

These examples show how to use segment routing traceroute to trace the LSP for a specified IPv4 prefix SID address. In the first example, FEC type is not specified. You can also specify the FEC type as shown in the other examples.

```
RP/0/RP0/CPU0:router# traceroute sr-mpls 10.1.1.2/32

Tracing MPLS Label Switched Path to 10.1.1.2/32, timeout is 2 seconds

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
  'L' - labeled output interface, 'B' - unlabeled output interface,
  'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
  'M' - malformed request, 'm' - unsupported tlvs, 'N' - no rx label,
  'P' - no rx intf label prot, 'p' - premature termination of LSP,
  'R' - transit router, 'I' - unknown upstream index,
  'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.

  0 10.12.12.1 MRU 1500 [Labels: implicit-null Exp: 0]
! 1 10.12.12.2 3 ms


RP/0/RP0/CPU0:router# traceroute sr-mpls 10.1.1.2/32  fec-type generic

Tracing MPLS Label Switched Path to 10.1.1.2/32, timeout is 2 seconds

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
  'L' - labeled output interface, 'B' - unlabeled output interface,
  'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
  'M' - malformed request, 'm' - unsupported tlvs, 'N' - no rx label,
  'P' - no rx intf label prot, 'p' - premature termination of LSP,
  'R' - transit router, 'I' - unknown upstream index,
  'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.

  0 10.12.12.1 MRU 1500 [Labels: implicit-null Exp: 0]
! 1 10.12.12.2 2 ms


RP/0/RP0/CPU0:router# traceroute sr-mpls 10.1.1.2/32  fec-type  igp ospf

Tracing MPLS Label Switched Path to 10.1.1.2/32, timeout is 2 seconds

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
  'L' - labeled output interface, 'B' - unlabeled output interface,
  'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
  'M' - malformed request, 'm' - unsupported tlvs, 'N' - no rx label,
  'P' - no rx intf label prot, 'p' - premature termination of LSP,
  'R' - transit router, 'I' - unknown upstream index,
  'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.

  0 10.12.12.1 MRU 1500 [Labels: implicit-null Exp: 0]
! 1 10.12.12.2 2 ms
```

```
RP/0/RP0/CPU0:router# traceroute sr-mpls 10.1.1.2/32  fec-type  igp isis

Tracing MPLS Label Switched Path to 10.1.1.2/32, timeout is 2 seconds

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
  'L' - labeled output interface, 'B' - unlabeled output interface,
  'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
  'M' - malformed request, 'm' - unsupported tlvs, 'N' - no rx label,
  'P' - no rx intf label prot, 'p' - premature termination of LSP,
  'R' - transit router, 'I' - unknown upstream index,
  'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.

  0 10.12.12.1 MRU 1500 [Labels: implicit-null Exp: 0]
! 1 10.12.12.2 2 ms


RP/0/RP0/CPU0:router#traceroute sr-mpls 10.1.1.2/32 fec-type bgp


Tracing MPLS Label Switched Path to 10.1.1.2/32, timeout is 2 seconds

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
  'L' - labeled output interface, 'B' - unlabeled output interface,
  'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
  'M' - malformed request, 'm' - unsupported tlvs, 'N' - no rx label,
  'P' - no rx intf label prot, 'p' - premature termination of LSP,
  'R' - transit router, 'I' - unknown upstream index,
  'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.


  0 10.12.12.1 MRU 1500 [Labels: implicit-null/implicit-null Exp: 0/0]
! 1 10.12.12.2 2 ms
```

This example shows how to use multipath traceroute to discover all the possible paths for a IPv4 prefix SID.

```
RP/0/RP0/CPU0:router# traceroute sr-mpls multipath 10.1.1.2/32

Starting LSP Path Discovery for 10.1.1.2/32

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
  'L' - labeled output interface, 'B' - unlabeled output interface,
  'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
  'M' - malformed request, 'm' - unsupported tlvs, 'N' - no rx label,
  'P' - no rx intf label prot, 'p' - premature termination of LSP,
  'R' - transit router, 'I' - unknown upstream index,
  'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.

!
Path 0 found,
 output interface GigabitEthernet0/0/0/2 nexthop 10.13.13.2
source 10.13.13.1 destination 127.0.0.0
!
Path 1 found,
 output interface Bundle-Ether1 nexthop 10.12.12.2
source 10.12.12.1 destination 127.0.0.0

Paths (found/broken/unexplored) (2/0/0)
Echo Request (sent/fail) (2/0)
```

```
Echo Reply (received/timeout) (2/0)
Total Time Elapsed 14 ms
```

Traceroute for SR Policy

You can perform the traceroute operation for an SR policy (or binding SID), and LSP end-point combination. Use the **traceroute** command's **policy name lsp-end-point** and **policy binding-sid lsp-end-point** options to perform this task. You can instantiate the policy through the CLI, Netconf, PCEP or BGP-TE process.

IPv6 policies are not supported for SR OAM function.

Note As a prerequisite, you must enable the MPLS OAM function.

```
Router(config)# mpls oam
Router(config)# commit

Router# traceroute sr-mpls policy name srte_c_4_ep_10.0.0.1 lsp-end-point 209.165.201.1
Router# traceroute sr-mpls policy binding-sid 1000 lsp-end-point 209.165.201.1
```

Segment Routing Policy Nil-FEC Ping and Traceroute

Segment routing OAM supports Nil-FEC LSP ping and traceroute operations to verify the connectivity for segment routing MPLS data plane. For the existing Nil-FEC ping and traceroute commands, you need to specify the entire outgoing label stack, outgoing interface, as well as the next hop. SR policy Nil-FEC ping and SR policy Nil-FEC traceroute enhancements extend the data plane validation functionality of installed SR policies through Nil-FEC ping and traceroute commands while simplifying the operational process. Instead of specifying the entire outgoing label-stack, interface, and next-hop, you can use the policy name or the policy binding-SID label value to initiate Nil-FEC ping and traceroute operations for the SR policies. Specification of outgoing interface and next-hop is also not required for policy Nil-FEC OAM operations.

Restrictions and Usage Guidelines

The following restrictions and guidelines apply for this feature:

- You cannot select a specific candidate path for SR policy Nil-FEC ping and traceroute.
- You cannot use SR policy Nil-FEC ping or traceroute for non-selected candidate paths.

Examples: SR Policy Nil-FEC Ping

These examples show how to use SR policy Nil-FEC ping for a SR policy. The first example refers the SR policy-name while the second example refers the BSID.

```
RP/0/0/CPU0:router# ping sr-mpls nil-fec policy name POLICY1
Thu Feb 22 06:56:50.006 PST
Sending 5, 100-byte MPLS Echos with Nil FEC for SR-TE Policy POLICY1,
      timeout is 2 seconds, send interval is 0 msec:
Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
  'L' - labeled output interface, 'B' - unlabeled output interface,
  'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
  'M' - malformed request, 'm' - unsupported tlvs, 'N' - no rx label,
  'P' - no rx intf label prot, 'p' - premature termination of LSP,
  'R' - transit router, 'I' - unknown upstream index,
  'X' - unknown return code, 'x' - return code 0
```

```
Type escape sequence to abort.
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/5/22 ms


RP/0/0/CPU0:router# ping sr-mpls nil-fec policy binding-sid  100001
Thu Dec 17 12:41:02.381 EST
Sending 5, 100-byte MPLS Echos with Nil FEC with labels [16002,16003],
      timeout is 2 seconds, send interval is 0 msec:
Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
  'L' - labeled output interface, 'B' - unlabeled output interface,
  'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
  'M' - malformed request, 'm' - unsupported tlvs, 'N' - no rx label,
  'P' - no rx intf label prot, 'p' - premature termination of LSP,
  'R' - transit router, 'I' - unknown upstream index,
  'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 2/3/3 ms
```

Examples: SR Policy Nil-FEC Traceroute

These examples show how to use SR policy Nil-FEC traceroute for a SR policy. The first example refers the SR policy-name while the second example refers the binding SID (BSID).

```
RP/0/0/CPU0:router#  traceroute sr-mpls nil-fec policy name POLICY1
Thu Feb 22 06:57:03.637 PST
Tracing MPLS Label Switched Path with Nil FEC for SR-TE Policy POLICY1, timeout is 2 seconds
Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
  'L' - labeled output interface, 'B' - unlabeled output interface,
  'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
  'M' - malformed request, 'm' - unsupported tlvs, 'N' - no rx label,
  'P' - no rx intf label prot, 'p' - premature termination of LSP,
  'R' - transit router, 'I' - unknown upstream index,
  'X' - unknown return code, 'x' - return code 0
Type escape sequence to abort.
  0 11.11.11.1 MRU 1500 [Labels: 16003/explicit-null Exp: 0/0]
L 1 11.11.11.2 MRU 1500 [Labels: implicit-null/explicit-null Exp: 0/0] 4 ms
! 2 14.14.14.3 2 ms


RP/0/0/CPU0:router# traceroute sr-mpls nil-fec binding-sid 100001
Tracing MPLS Label Switched Path with Nil FEC with labels [16002/16004], timeout is 2 seconds

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
  'L' - labeled output interface, 'B' - unlabeled output interface,
  'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
  'M' - malformed request, 'm' - unsupported tlvs, 'N' - no rx label,
  'P' - no rx intf label prot, 'p' - premature termination of LSP,
  'R' - transit router, 'I' - unknown upstream index,
  'X' - unknown return code, 'x' - return code 0
Type escape sequence to abort.
  0 99.1.2.1 MRU 4470 [Labels: 16002/16004/explicit-null Exp: 0/0/0]
L 1 99.1.2.2 MRU 4470 [Labels: 16004/explicit-null Exp: 0/0] 3 ms
L 2 99.2.6.6 MRU 4470 [Labels: implicit-null Exp: 0] 3 ms
! 3 99.4.6.4 11 ms
```

Segment Routing Data Plane Monitoring

Table 19: Feature History Table

Feature Name	Release Information	Feature Description
Segment Routing Data Plane Monitoring	Release 7.3.1	Unreported traffic drops in MPLS networks could be difficult to detect and isolate. They can be caused by user configuration, out-of-sync neighbors, or incorrect data-plane programming. Segment Routing Data Plane Monitoring (SR DPM) provides a scalable solution to address data-plane consistency verification and unreported traffic drops. SR DPM validates the actual data plane status of all FIB entries associated with SR IGP prefix SIDs.

The primary benefits of SR DPM include:

- **Automation** – A node automatically verifies the integrity of the actual forwarding entries exercised by transit traffic.
- **Comprehensive Coverage** – Tests validate forwarding consistency for each set of destination prefixes across each combination of upstream and downstream neighbors and across all ECMP possibilities.
- **Scalability** – SR DPM is a highly scalable solution due to its localized detection process.
- **Proactive and Reactive modes of operation** – Solution caters to both continuous and on-demand verification.
- **Standards-based** – SR DPM uses existing MPLS OAM tools and leverages SR to enforce test traffic path.

DPM performs data plane validation in two phases:

- Adjacency Validation—Using special MPLS echo request packets, adjacency validation ensures that all local links are able to forward and receive MPLS traffic correctly from their neighbors. It also ensures that DPM is able to verify all local adjacency SID labels and to flag any inconsistencies, including traffic drops, forwarding by the local or neighboring device to an incorrect neighbor that is not associated with the specified adjacency, or forwarding by the local or neighboring device to the correct neighbor but over an incorrect link not associated with the specified adjacency. DPM validates the following adjacencies for each link when available:
  - Unprotected adjacency
  - Protected adjacency
  - Static adjacency
  - Dynamic adjacency
  - Shared adjacency

Note Observe the following limitations for adjacency validation:

- The adjacency validation phase only validates links that are participating in IGP (OSPF and IS-IS) instances. If one or more link is not part of the IGP, it will not be validated since there are no Adjacency SID labels.
- Adjacency validation only validates physical and bundle links, including broadcast links.

- Prefix Validation—Prefix validation identifies any forwarding inconsistency of any IGP Prefix SID reachable from the device. The validation is done for all upstream and downstream neighbor combinations of each prefix SID, and identifies inconsistencies in the downstream neighbor. The prefix validation phase simulates customer traffic path by validating both ingress and egress forwarding chain at the DPM processing node.

  Since prefix validation is localized to a device running DPM as well as its immediate neighbors, it does not suffer from scale limitations of end-to-end monitoring.

  Prefix validation builds on top of adjacency validation by using special MPLS echo requests that travel to the upstream node, return to the DPM-processing node, and time-to-live (TTL) expire at the immediate downstream node, thus exercising entire forwarding path towards the downstream.

Note Observe the following limitations for prefix validation:

- Because prefix validation builds on top of adjacency validation, if a link is not part of adjacency validation, it is not used in prefix validation.
- If all adjacencies are marked as “Faulty” during adjacency validation, prefix validation is not performed.
- If a node only has downstream links at a specific node, but no upstream node (possible in certain PE node scenarios), Prefix Validation is not performed.
- Prefix validation does not support TI-LFA.

DPM maintains a database of all prefixes and adjacencies being monitored.

The prefix database is populated by registering as a redistribution client to RIB, which enables DPM to keep the database up-to-date whenever IGP pushes a new prefix SID to RIB, deletes an existing prefix SID, or when the path of an existing prefix SID is modified.

DPM maintains the following prefix data:

- IPv4 Prefix
- Prefix Length
- Prefix SID label
- Error stats

DPM also maintains a list of all local adjacencies. DPM maintains a database that contains local links, their respective local and remote adjacency labels and IP addresses, and error stats.

SR-DPM Operation: Example

The following SR-DPM operation example use the following scenarios:

Figure 10: Test Iteration A Path

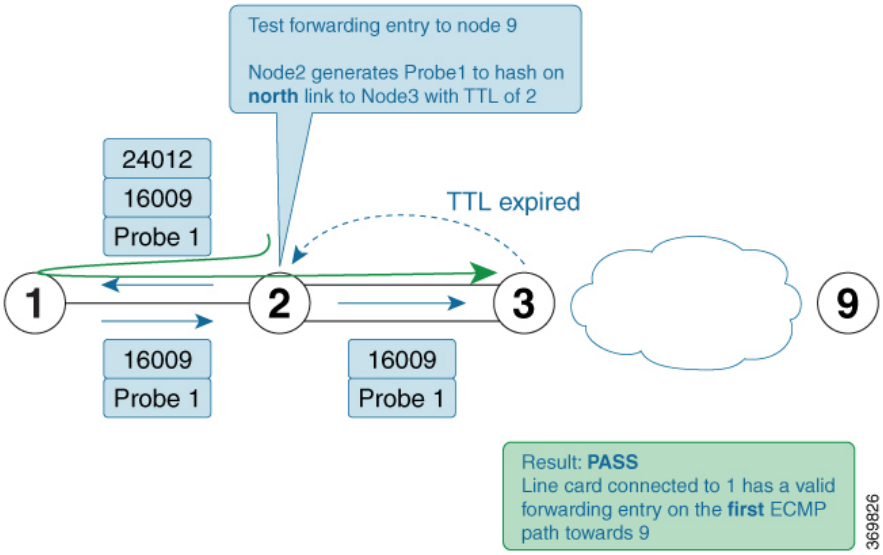
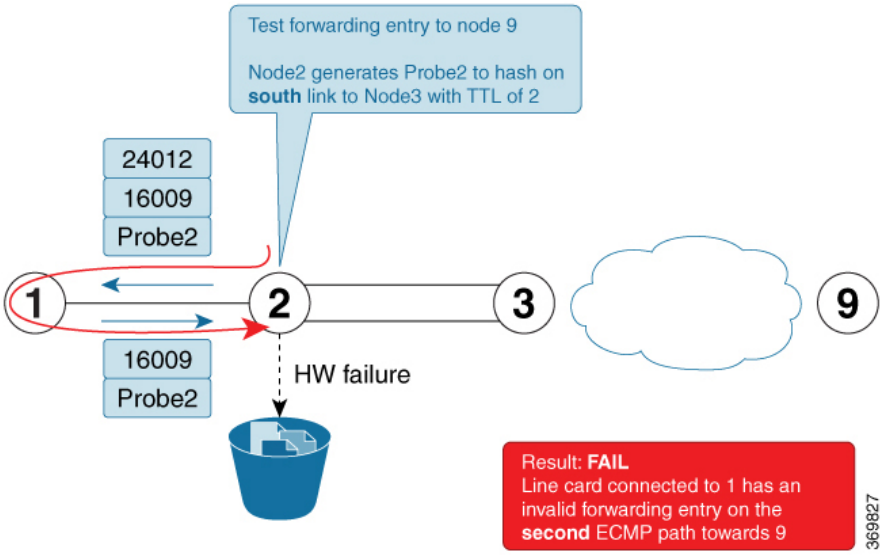


Figure 11: Test Iteration B Path



Node 2 is a DPM-capable device. DPM is enabled ***in proactive mode*** to perform forwarding consistency tests for all prefix-SIDs in the network. For each destination prefix, the router identifies the directly connected upstream and downstream neighbors used to reach a given destination.

Using node 9 as the prefix under test (prefix-SID = 16009), node 1 is designated as the upstream node and node 3 as the downstream nodes with 2 ECMPs.

1. Node 2 generates test traffic (MPLS OAM ping with source_ip of node 2) to test its forwarding for every upstream/downstream combination. In this case, two combinations exist:
   - Prefix-SID node 9 - test iteration A path = Node 2 to Node 1 to Node 2 to Node 3 (via **top** ECMP)
   - Prefix-SID node 9 - test iteration B path = Node 2 to Node 1 to Node 2 to Node 3 (via **bottom** ECMP)
2. Node 2 adds a label stack in order to enforce the desired path for the test traffic. For example, two labels are added to the packet for test iterations A and B:
   - The top label is equal to the adjacency-SID on node 1 for the interface facing node 2 (adjacency SID = 24012). The bottom label is the prefix-SID under test (16009). The test traffic is sent on the interface facing node 1.
   - The top label (after being POPed at node 1) causes the test traffic to come back to node 2. This returning traffic is completely hardware-switched based on the forwarding entry for the prefix-SID under test (16009). Note that the labeled test traffic has a time-to-live (TTL) of 2 and it will never be forwarded beyond the downstream router(s).
   - When test traffic reaches node 3, a TTL expired response is sent back to node 2. If the response packet arrives over the expected interface (**top** ECMP link) then the forwarding verification on node 2 for the first iteration towards node 9 is considered to be a success.
   - The difference between the test traffic for test iteration A and B in this example is the destination_ip of the MPLS OAM ping. Node 2 calculates them in this order to exercise a given ECMP path (if present). Thus, test traffic for iteration A is hashed onto the **top** ECMP and test traffic for iteration B is hashed onto the **bottom** ECMP link.
3. The DPM tests are then repeated for the remaining prefix-SIDs in the network

Configure SR DPM

To configure SR-DPM, complete the following configurations:

- Enable SR DPM
- Configure SR DPM interval timer
- Configure SR DPM rate limit

Enable SR DPM

Use the **mpls oam dpm** command to enable SR DPM and enter MPLS OAM DPM command mode.

```
Router(config)# mpls oam dpm
Router(config-oam-dpm)#
```

Configure SR DPM Interval Timer

Use the **interval** *minutes* command in MPLS OAM DPM command mode to specify how often to run DPM scan. The range is from 1 to 3600 minutes. The default is 30 minutes.

```
Router(config-oam-dpm)# interval 240
Router(config-oam-dpm#
```

Configure SR DPM Rate Limit

Use the **pps** *pps* command in MPLS OAM DPM command mode to rate limit the number of echo request packets per second (PPS) generated by DPM. The range is from 1 to 250 PPS. The default is 50 PPS.

Note If the specified rate limit is more than the rate limit for overall MPLS OAM requests, DPM generates an error message.

```
Router(config-oam-dpm)# pps 45
Router(config-oam-dpm#
```

Verification

```
Router# show mpls oam dpm summary
  Displays the overall status of SR-DPM from the last run.
Router# show mpls oam dpm adjacency summary
  Displays the result of DPM adjacency SID verification for all local interfaces from the
last run.
Router# show mpls oam dpm adjacency interface
  Displays the result of DPM adjacency SID verification for all adjacencies for the specified
 local interface.
Router# show mpls oam dpm counters
  Outputs various counters for DPM from last run as well as since the start of DPM process.
Router# show mpls oam dpm prefix summary
  Displays the result of DPM prefix SID verification for all reachable IGP prefix SIDs from
 the last run.
Router# show mpls oam dpm prefix prefix
  Displays the result of DPM prefix SID verification for the specified prefix including all
 upstream and downstream combinations.
Router# show mpls oam dpm trace
  Returns logged traces for DPM.
```

In addition, the existing **show mpls oam** command is extended to specify DPM counters.

```
Router# show mpls oam counters packet dpm
```

Data Plane Validation Support for SR-MPLS IPv6-based LSPs

Table 20: Feature History Table

Feature Name	Release Information	Feature Description
Data Plane Validation for SR-MPLS IPv6-based Controller Instantiated LSPs	Release 7.3.6 Release 24.2.1	You can now verify the network configuration and paths and policies set up, without interrupting or potentially disrupting live network traffic, for SR-MPLS (Segment Routing over Multiprotocol Label Switching) IPv6-based Label Switched Paths (LSPs). With this feature, you can validate controller instantiated LSPs programmed directly into the forwarding hardware. Previously, SR data plane validation was possible over IPv4-based LSPs. The feature introduces these changes: **CLI:** • The **dataplane-only** keyword is introduced in the **traceroute sr-mpls** and **ping sr-mpls** commands. **YANG Data Models:** • `Cisco-IOS-XR-mpls-traceroute-act.yang` • `Cisco-IOS-XR-mpls-ping-act.yang` See (GitHub, Yang Data Models Navigator)

With this configuration, you can validate the SR-MPLS (Segment Routing over Multiprotocol Label Switching) IPv6-based LSPs and policies without disrupting the live network traffic. You can also validate Service-layer Application Programming Interface (SL-API) initiated LSPs such as controller instantiated LSPs. For more information about SL-API, refer *Use Service Layer API to Bring your Controller on Cisco IOS XR Router* chapter in *Programmability Configuration Guide*.

In the earlier releases, you could perform SR-MPLS data plane validation over IPv4-based LSPs. For more information, refer Segment Routing Ping, on page 156 and Segment Routing Traceroute, on page 158 sections.

Examples: SR-MPLS Data Plane Validation over IPv6-based LSPs

The following example shows how to use segment routing ping to validate SR-MPLS over IPv6-based LSPs:

```
Router#ping sr-mpls dataplane-only 2001:DB8::1/32
Tue Jan 16 15:05:19.120 EST

Sending 5, 100-byte MPLS Echos with Nil FEC to 2001:DB8::1/32,
      timeout is 2 seconds, send interval is 0 msec:

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
  'L' - labeled output interface, 'B' - unlabeled output interface,
  'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
  'M' - malformed request, 'm' - unsupported tlvs, 'N' - no rx label,
  'P' - no rx intf label prot, 'p' - premature termination of LSP,
  'R' - transit router, 'I' - unknown upstream index,
  'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.

!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/2/8 ms
```

The following example shows how to use segment routing traceroute to validate SR-MPLS over IPv6-based LSPs:

```
Router#traceroute sr-mpls dataplane-only 2001:DB8::1/32
Tue Jan 16 15:08:54.681 EST

Tracing MPLS Label Switched Path with Nil FEC to 2001:DB8::1/32, timeout is 2 seconds

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
  'L' - labeled output interface, 'B' - unlabeled output interface,
  'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
  'M' - malformed request, 'm' - unsupported tlvs, 'N' - no rx label,
  'P' - no rx intf label prot, 'p' - premature termination of LSP,
  'R' - transit router, 'I' - unknown upstream index,
  'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.

  0 11:11:11::1 MRU 1500 [Labels: 18004/explicit-null Exp: 0/0]
L 1 11:11:11::2 MRU 1500 [Labels: implicit-null/explicit-null Exp: 0/0] 3 ms
! 2 15:15:15::4 3 ms
```

The following example shows how to trace the SR-MPLS LSPs with Nil-FEC that includes labels:

```
Router#traceroute sr-mpls nil-fec labels 18004 output interface GigabitEthernet 0/0/0/0
nexthop 10:10:10::2
Tue Jan 16 15:28:03.162 EST

Tracing MPLS Label Switched Path with Nil FEC with labels [18004], timeout is 2 seconds

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
  'L' - labeled output interface, 'B' - unlabeled output interface,
  'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
  'M' - malformed request, 'm' - unsupported tlvs, 'N' - no rx label,
  'P' - no rx intf label prot, 'p' - premature termination of LSP,
  'R' - transit router, 'I' - unknown upstream index,
  'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.

  0 10:10:10::1 MRU 1500 [Labels: 18004/explicit-null Exp: 0/0]
```

```
L 1 10:10:10::2 MRU 1500 [Labels: implicit-null/explicit-null Exp: 0/0] 2 ms
! 2 15:15:15::4 2 ms
```