Configuring MPLS Segment Routing OAM

This chapter describes the Multiprotocol Label Switching (MPLS) segment routing OAM functionality.

- About MPLS Segment Routing OAM, on page 1
- Guidelines and Limitations for MPLS SR OAM, on page 2
- MPLS Ping and Traceroute for Nil FEC, on page 3
- MPLS Ping and Traceroute for BGP and IGP Prefix SID, on page 4
- Verifying Segment Routing OAM, on page 4
- Examples for using Ping and Traceroute CLI commands, on page 6

About MPLS Segment Routing OAM

MPLS segment routing (SR) has been deployed on the Cisco Nexus 9000 Series switches. As MPLS segment routing (SR) is deployed, a few diagnostic tools are required to help resolve the misconfigurations or failures in the segment routing network. Segment Routing Operations, Administration, and Maintenance (OAM) helps service providers monitor label-switched paths (LSPs) and quickly isolate forwarding problems to assist with fault detection and troubleshooting in the network.

MPLS SR OAM provides two main functions for diagnostics purposes:

1. MPLS ping
2. MPLS traceroute

The segment routing OAM feature provides support for the following FEC types:

- Ping and traceroute to SR-IGP IS-IS IPv4 prefixes. This allows validation of prefix SIDs distributed in an IS-IS SR underlay.
- Ping and traceroute to BGP IPv4 prefixes. This allows validation of prefix SIDs distributed in a BGP SR underlay.
- Ping and traceroute to Generic IPv4 prefixes. This allows validation of prefix SIDs distributed in an SR underlay agnostic to the protocol that performed the distribution. The validation is performed by checking the Unicast Routing Information Base (URIB) and Unicast Label Information Base (ULIB).
- Ping and traceroute to Nil FEC prefixes. This allows a less comprehensive data-plane-only validation for any MPLS SR prefix, with finer-grained control over the path the ping or traceroute takes. The path may be specified using an SR-TE policy name or SR-TE policy color and endpoint.

To enable MPLS OAM on Cisco Nexus 9000 Series switches, use the **feature mpls oam** CLI command. Use the **no feature mpls oam** CLI command to disable MPLS OAM on Cisco Nexus 9000 Series switches.

Segment Routing Ping

Similar to how an IP ping validates connectivity to an IP host, MPLS ping is used to validate unidirectional continuity along an MPLS Label-Switched Path (LSP). By providing a FEC representing the LSP to be validated, MPLS ping performs the following:

- Confirms that the echo requests for the FEC reach an endpoint for the LSP. Except for the Nil FEC, for all other FEC types it confirms that the endpoint is the correct egress for that FEC.
- Measures coarse round trip time.
- Measures coarse round trip delay.

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.

Segment Routing Traceroute

MPLS traceroute verifies forwarding and control plane at each hop of the LSP to isolate faults. Traceroute sends MPLS echo requests with monotonically increasing time-to-live (TTL), starting with TTL of 1. Upon TTL expiry, transit node processes the request in software and verifies if it has an LSP to the target FEC and intended transit node. The transit node sends echo reply containing return code specifying the result of above verification and label stack to reach the next-hop, as well as ID of the next-hop towards destination, if verification is successful. Originator processes echo reply to build the next echo request containing TTL+1. This process is repeated until the destination replies that it is the egress for the FEC.

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

Guidelines and Limitations for MPLS SR OAM

MPLS OAM Nil FEC has the following guidelines and limitations:

- MPLS OAM Nil FEC is supported on the Cisco Nexus 9300-FX platform switches.
- MPLS OAM Nil FEC is not supported on Cisco Nexus 9500 platform switches with -R line cards.
- For all new FEC types supported in Cisco NX-OS Release 9.3(1), only a one-label stack is supported. FEC-Stack change TLV support and the associated validations are not supported. This limitation is not applicable to Nil FEC.

- In Cisco NX-OS Release 9.3(1), the SR-IGP "any" prefix type and the adjacency SIDs described in RFC 8287 are not supported.
- OSPF ping and traceroute is not supported in Cisco NX-OS Release 9.3(1).
- Beginning with Cisco NX-OS Release 9.3(3), MPLS OAM Nil FEC is supported on Cisco Nexus 9300-GX platform switches.
- A maximum of 4 labels can be specified in the **ping mpls nil-fec** and **traceroute mpls nil-fec** commands. This value is enforced by querying the platform and currently Cisco Nexus 9000 Series switches limit the label stack to 5. It means that for a Nil FEC echo request, you can specify a maximum of 4 labels because internally an extra explicit-null is added.
- The nexthop specified in the ping and traceroute commands must be a connected nexthop on the originator and it should not be a recursive nexthop.
- There is no support for treetrace.
- Nil FEC does not carry any information to identify the intended target. The packet may mis-forward at an incorrect node but the validation may return success if the packet ends up at a node after popping the non-null labels.
- Nil FEC operates on forwarding the information alone. It cannot detect the inconsistencies between the control plane and the forwarding plane by definition.
- Nil FEC ping and traceroute is not supported for deaggregator (per-VRF) labels. This includes the BGP EVPN-Layer 3 deaggregator labels.
- On Cisco Nexus 9000 Series switches that use Broadcom chipsets, there is no support to allow the software to send a query to determine which ECMP a packet takes. It means that for MPLS traceroutes that traverse one of these switches may display an error at the next hop if there is more than one ECMP as displayed in the following example:

  ```
  D 2 6.0.0.2 MRU 1496 [Labels: 2003/explicit-null Exp: 0/0] 4 ms
  ```

- When you use OAM to test a BGP EPE LSP (for example, the last label in the ping/traceroute label stack is an EPE label), OAM only returns success if the final router has OAM enabled and MPLS is enabled on the incoming interface.

  For example, if you have a setup as A---B---C, A and B are in the SR network, and B acts like a PE and C acts like a CE, B is configured with C as a BGP EPE peer (using egress-engineering on B), then C must have OAM and MPLS forwarding enabled on the incoming interface.

MPLS Ping and Traceroute for Nil FEC

The Nil FEC LSP ping and traceroute operations are extensions of regular MPLS ping and traceroute. The Nil FEC LSP ping and traceroute functionality supports segment routing and MPLS Static. It also acts as an additional diagnostic tool for all other LSP types.

Unlike the other FEC types, Nil FEC does not provide control plane validation. Nil FEC ping or traceroute probes can reach any switch on which the MPLS OAM functionality is enabled.

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 case of segment routing, each segment nodal label and adjacent 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.

Use the [**ping**|**traceroute**] **mpls nil-fec labels** *comma-separated-labels* [**output** {**interface** *tx-interface*} [**nexthop** *nexthop-ip-addr*]] CLI command to execute a ping or a traceroute.

If you have configured an SR-TE policy name or the color and the endpoint, you can use the following CLI command to execute a ping or a traceroute to use the existing SR-TE policy information.:

[**ping**|**traceroute**] **mpls nil-fec** [**policy name** *name*] [**endpoint** *nexthop-ip-addr*] [**on-demand color** *color-num*] CLI command to execute a ping or a traceroute.

MPLS Ping and Traceroute for BGP and IGP Prefix SID

MPLS ping and traceroute operations for Prefix SID are supported for the following BGP and IGP scenarios:

- Within an IS-IS level
- Across IS-IS levels
- BGP SR underlay

These FEC types perform an additional control plane check to ensure that the packets are not mis-routed. This validation ensures that the pinged FEC type is connected to the switch and is distributed to the other nodes. Nil FEC does not provide this validation.

MPLS echo request packets carry Target FEC Stack sub-TLVs. The Target FEC sub-TLVs are used by the responder for FEC validation. The IGP/BGP IPv4 prefix sub-TLV has been added to the Target FEC Stack sub-TLV. The IGP/BGP IPv4 prefix sub-TLV contains the prefix SID, the prefix length, and the protocol (IS-IS).

Use the **ping**|**traceroute sr-mpls** A.B.C.D/LEN fec-type [*bgp* | *igp {isis}* | *generic*] CLI command to execute a traceroute.

Verifying Segment Routing OAM

This section provides information on the CLI commands that can be used to verify the segment routing OAM features.

- Verifying Segment Routing OAM IS-IS, on page 4

Verifying Segment Routing OAM IS-IS

The following ping commands are used to display SR OAM when the underlying network is IS-IS:

```
switch# ping sr-mpls 11.1.1.3/32 fec-type igp isis

Sending 5, 100-byte MPLS Echos to IGP Prefix SID(IS-IS) FEC 11.1.1.3/32,
```

```
     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 label entry,
  '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 Ctrl-C to abort.
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 2/2/3 ms
 Total Time Elapsed 18 ms
```

```
switch# traceroute sr-mpls 11.1.1.3/32 fec-type igp isis


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,
  'X' - unknown return code, 'x' - return code 0

Type Ctrl-C to abort.
  0 172.18.1.2 MRU 1500 [Labels: 16103 Exp: 0]
L 1 172.18.1.1 MRU 1504 [Labels: implicit-null Exp: 0] 4 ms
! 2 172.18.1.10 3 ms
```

```
switch# ping sr-mpls 11.1.1.3/32 fec-type igp isis verbose

Sending 5, 100-byte MPLS Echos to IGP Prefix SID(IS-IS) FEC 11.1.1.3/32,
     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 label entry,
  '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 Ctrl-C to abort.
!    size 100, reply addr 172.18.1.10, return code 3
!    size 100, reply addr 172.18.1.10, return code 3
!    size 100, reply addr 172.18.1.10, return code 3
!    size 100, reply addr 172.18.1.10, return code 3
!    size 100, reply addr 172.18.1.10, return code 3

Success rate is 100 percent (5/5), round-trip min/avg/max = 2/2/3 ms
 Total Time Elapsed 17 ms
```

```
switch# ping sr-mpls 11.1.1.3/32 fec-type igp isis destination 127.0.0.1 127.0.0.2 repeat
1 verbose

Sending 1, 100-byte MPLS Echos to IGP Prefix SID(IS-IS) FEC 11.1.1.3/32,
     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 label entry,
  '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 Ctrl-C to abort.
Destination address 127.0.0.1
!    size 100, reply addr 172.18.1.10, return code 3

Destination address 127.0.0.2
!    size 100, reply addr 172.18.1.22, return code 3

Success rate is 100 percent (2/2), round-trip min/avg/max = 3/3/3 ms
 Total Time Elapsed 8 ms
```

Examples for using Ping and Traceroute CLI commands

Examples for IGP or BGP SR Ping and Traceroute

Using CLI to Execute a Ping with Explicit Outgoing Information

Use the **ping sr-mpls** *fec* **fec-type igp isis** CLI command to execute an IS-IS SR ping and the **ping sr-mpls** *fec* **fec-type bgp** CLI command to execute a BGP ping.

```
switch# ping sr-mpls 11.1.1.3/32 fec-type igp isis

Sending 5, 100-byte MPLS Echos to IGP Prefix SID(IS-IS) FEC 11.1.1.3/32,
     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 label entry,
  '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 Ctrl-C to abort.
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 2/2/3 ms
 Total Time Elapsed 18 ms
```

```
switch# ping sr-mpls 11.1.1.3/32 fec-type igp isis verbose

Sending 5, 100-byte MPLS Echos to IGP Prefix SID(IS-IS) FEC 11.1.1.3/32,
     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 label entry,
  '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 Ctrl-C to abort.
!    size 100, reply addr 172.18.1.10, return code 3
!    size 100, reply addr 172.18.1.10, return code 3
!    size 100, reply addr 172.18.1.10, return code 3
!    size 100, reply addr 172.18.1.10, return code 3
!    size 100, reply addr 172.18.1.10, return code 3
```

```
Success rate is 100 percent (5/5), round-trip min/avg/max = 2/2/3 ms
 Total Time Elapsed 17 ms
```

Examples for Nil FEC Ping and Traceroute

Using CLI to Execute a Ping with Explicit Outgoing Information

Use the **ping sr-mpls nil-fec labels** *comma-separated-labels* [**output** {**interface** *tx-interface*} [**nexthop** *nexthop-ip-addr*]] CLI command to execute a ping.

For example, the following command sends an MPLS packet with the outermost two labels in the label stack being 2001 and 2000 out the interface Ethernet 1/1 with a nexthop IP address of 4.0.0.2:

```
switch# ping mpls nil-fec labels 2001,2000 output interface e1/1 nexthop 4.0.0.2
```

It is mandatory that the nexthop is a connected nexthop; it is not recursively resolved.

The above CLI format is a simplified version. The [**output** {**interface** *tx-interface*} [**nexthop** *nexthop-ip-addr*]] is mandatory to be present in the VSH server. For example:

```
switch# ping mpls nil-fec labels 1,2 ?
output Output options
switch# ping mpls nil-fec labels1,2
^
% Invalid command at '^' marker.
```

Using CLI to Execute a Ping with Outgoing Information from an SRTE Policy

Use the following CLI command to execute a ping:

```
switch# ping mpls nil-fec policy name policy1
switch# ping mpls nil-fec policy endpoint 2.0.0.1 color 16
```

Using CLI to Execute a Traceroute with Explicit Outgoing Information

Use the following CLI command to execute a traceroute:

```
switch# ping mpls nil-fec labels 2001,2000 output interface e1/1 nexthop 4.0.0.2
```

Using CLI to Execute a Traceroute with Outgoing Information from an SRTE Policy

Use the following CLI command to execute a traceroute:

```
switch# traceroute mpls nil-fec policy name policy1
switch# traceroute mpls nil-fec policy endpoint 2.0.0.1 color 16
```

Displaying Show Statistics

Use the following command to display the statistics about the echo requests sent by the local MPLS OAM service:

```
show mpls oam echo statistics
```