



## SR-MPLS Handoff

---

Beginning with Release 5.0(1), segment routing (SR) Multiprotocol Label Switching (MPLS) handoff on the Cisco ACI border leaf switches is available as a new feature.



---

**Note** Procedures in this document describe how to configure SR-MPLS handoff using the GUI and REST API. You cannot configure SR-MPLS handoff through the NX-OS style CLI at this time.

---

- [Understanding ACI Handoffs, on page 1](#)
- [Understanding ACI Implementation of SR-MPLS Handoff, on page 6](#)
- [Understanding the SR-MPLS Configuration Model, on page 14](#)
- [Guidelines and Limitations for SR-MPLS, on page 18](#)
- [Configuring an SR-MPLS Infra L3Out Using the GUI, on page 26](#)
- [Configuring an SR-MPLS VRF L3Out Using the GUI, on page 33](#)
- [Creating SR-MPLS Custom QoS Policy Using the GUI, on page 36](#)
- [Displaying MPLS Statistics, on page 38](#)
- [Configuring SR-MPLS Global Block \(GB\), on page 40](#)
- [Migrating from IP Handoff Configuration to SR Handoff Configuration, on page 42](#)
- [About the BGP Domain-Path Feature for Loop Prevention, on page 49](#)

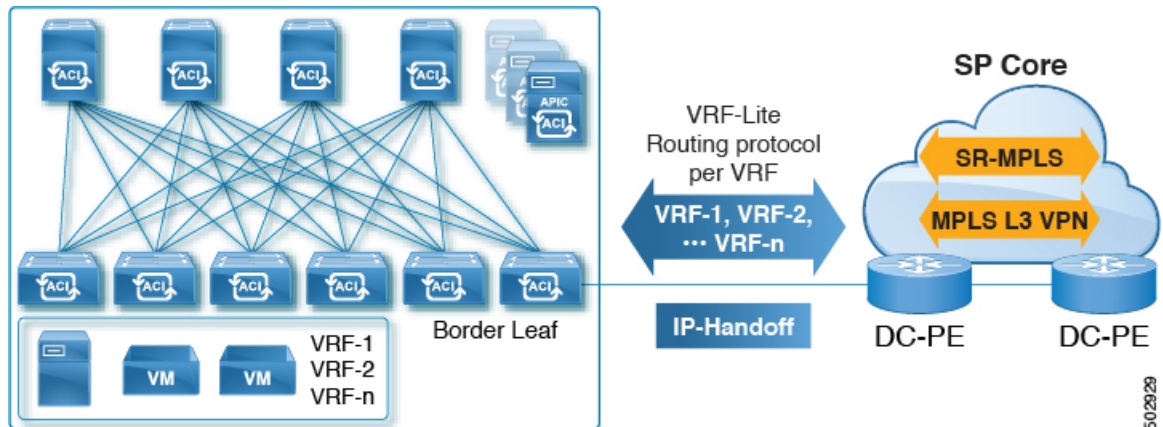
## Understanding ACI Handoffs

The following sections describe how ACI handoffs are handled for releases prior to Cisco APIC Release 5.0(1), using IP handoff, and how they are handled starting with Cisco APIC Release 5.0(1), using SR-MPLS handoff.

### ACI Handoffs Prior to Release 5.0(1): IP Handoff

Prior to Cisco APIC Release 5.0(1), when setting up an ACI fabric connecting the ACI border leaf nodes to a data center provider edge (DC-PE), if you have a configuration with a multi-tenant network, that means that you would need multiple VRFs, and you would need a routing protocol for each VRF. You would also need to dedicate an interface for each VRF, where the interface is either a physical interface or a logical interface. This configuration is typically called VRF-Lite, as shown in the following figure.

Figure 1: ACI Handoff to DC-PE Using IP Handoff (VRF-Lite)



In this configuration, the border leaf switch is connected to the DC-PE using VRF-Lite. The interface and routing protocol session configurations between the border leaf switch and the DC-PE is done using separate VRFs. Differentiated Services Code Point (DSCP) is configured on the border leaf switch for outgoing traffic. On the DC-PE, the DSCP is mapped to the segment routing for traffic engineering (SR-TE) policy, which is used to steer traffic through the transport network.

This configuration becomes cumbersome if you have a large number of sessions between the border leaf switch and the data center. Therefore, automation and scalability are key challenges when configuring using VRF-Lite.

## ACI Handoffs in Release 5.0(1): SR Handoff

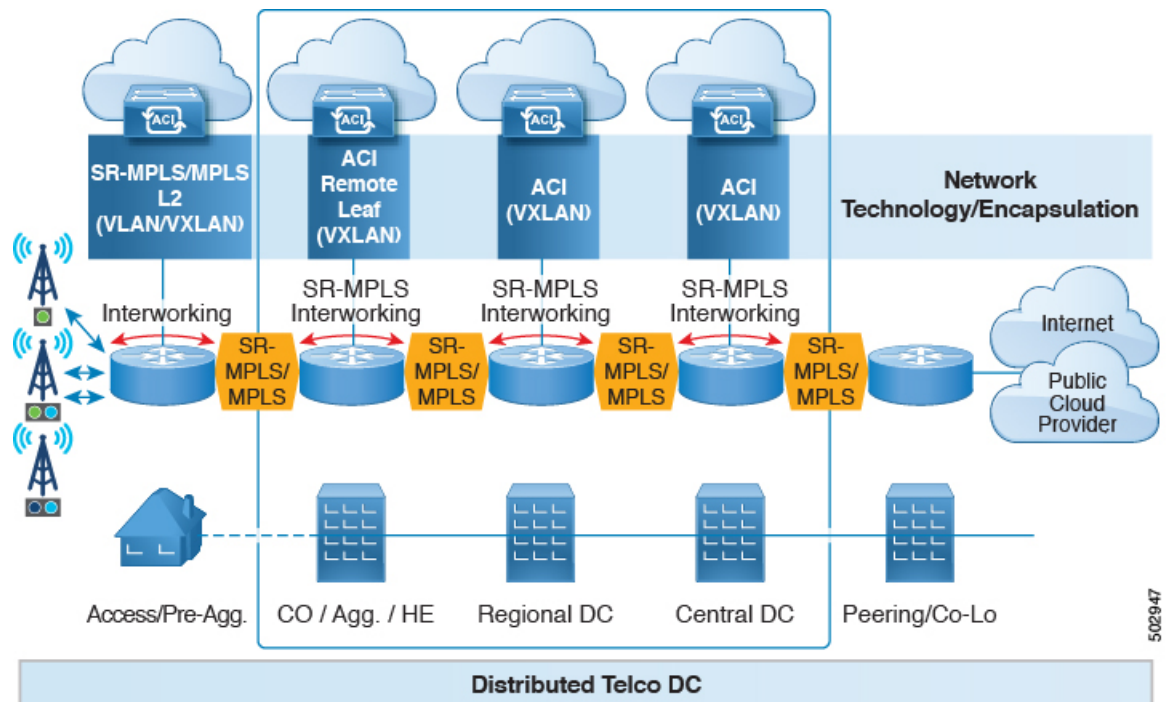
Beginning with Cisco APIC Release 5.0(1), you can now set up an ACI fabric connection between border leaf switches and DC-PE routers using SR-MPLS handoff. SR is a better solution than other options, because other options such as VXLAN may not be a common technology in an SP Core, whereas SR is a much more common and mature solution for transport devices.

The following scenarios show how configuring the ACI handoff to the DC-PE using SR-MPLS is beneficial:

- [Unified Segment Routing Transport, on page 2](#)
- [Monitoring DC-to-DC Flows in the Transport Network, on page 3](#)
- [Single Control Plane Session for Multiple VRFs, on page 4](#)
- [SR-TE/Flex Algo in Transport Using Color Community or Destination Prefix, on page 5](#)
- [DC and Transport QoS with SR or MPLS, on page 6](#)

### Unified Segment Routing Transport

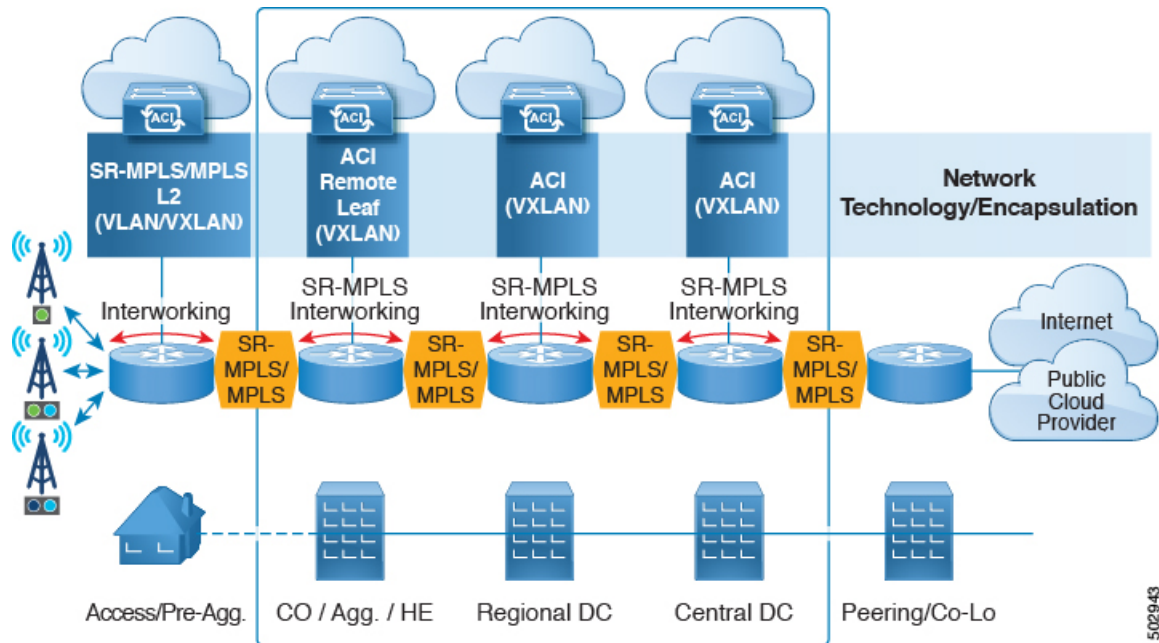
The following scenario highlights the deployment of a unified SR or MPLS transport network to interconnect different ACI DC networks. The VXLAN to SR-MPLS handoff is leveraged in each location between the ACI network and the DC-PE routers.



In this scenario, VXLAN is being used in the ACI fabric area, whereas segment routing is being used in the transport network. Rather than use VXLAN outside of the ACI fabric area, it would be preferable to use the same SR-based routing, where you would do an SR handoff or an MPLS handoff towards the transport device. By changing VXLAN to SR at the ACI border, the transport devices only need to run SR or MPLS and does not need to run VXLAN.

### Monitoring DC-to-DC Flows in the Transport Network

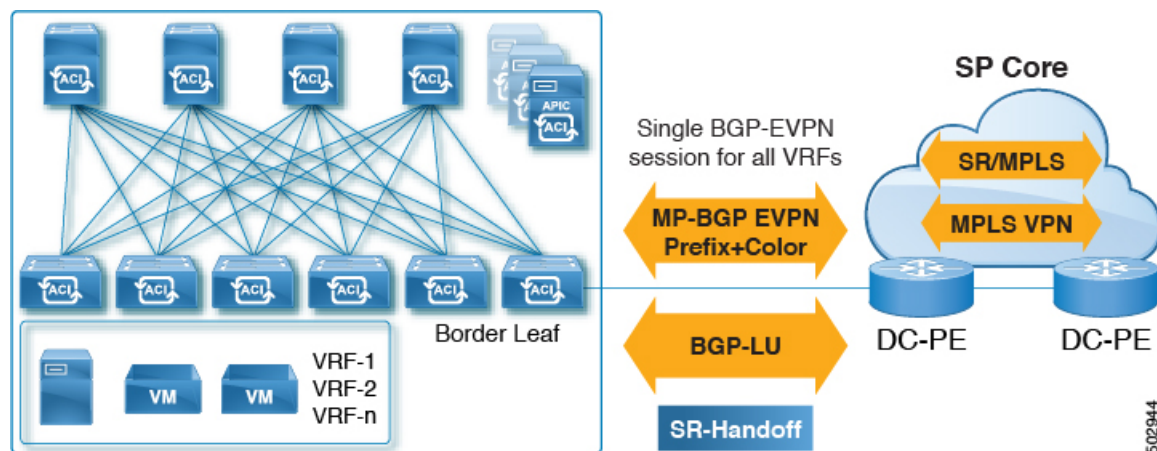
In the following scenario, DC-to-DC flows are encapsulated using segment routing instead of VXLAN.



In this scenario, the existing monitoring tools used for the transport network can monitor MPLS traffic, but cannot monitor VXLAN packets. By using ACI to SR-MPLS handoff, this allows the transport team to monitor the DC-to-DC flows using existing monitoring tools.

### Single Control Plane Session for Multiple VRFs

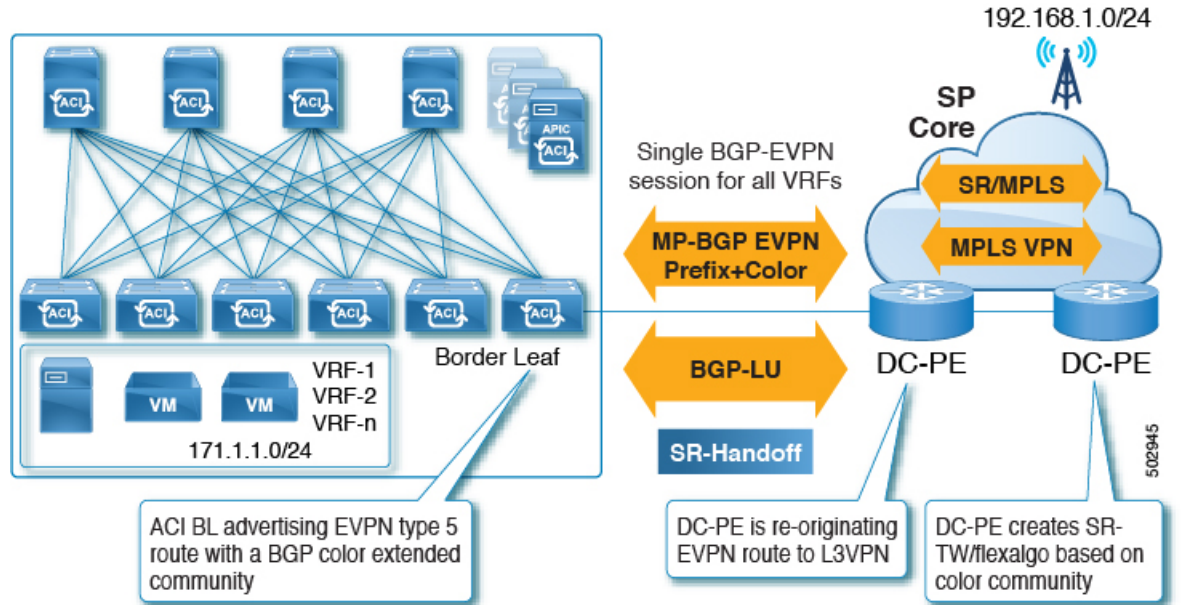
Using SR handoff, a single control plane session (MP-BGP EVPN) is used for all VRFs, rather than having per-VRF sessions that you would have to use in the IP handoff configuration. This provides better automation and scalability options for multiple VRFs between the ACI data center and the DC-PE.



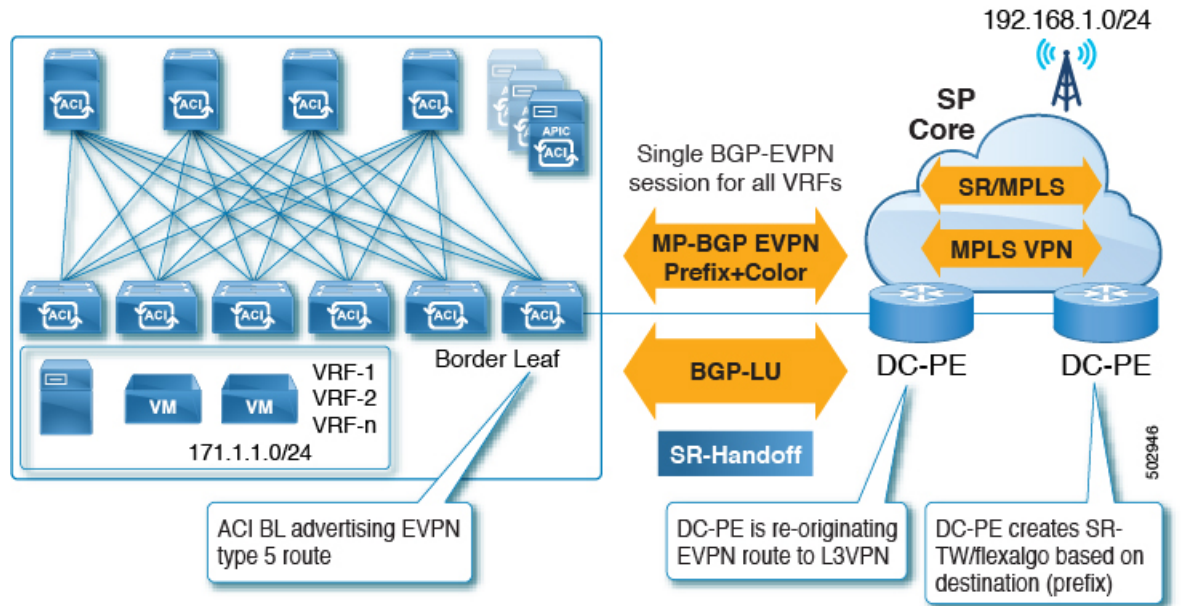
With SR handoff, a single control plane and data plane session is used instead of per-VRF control plane and data plane sessions, with a unified SR transport from the Cisco ACI fabric to the SP core. The BGP Label Unicast (BGP LU) address-family is used for the underlay label exchange. The MP-BGP EVPN address-family carries the prefix and MPLS label per VRF information.

**SR-TE/Flex Algo in Transport Using Color Community or Destination Prefix**

SR handoff is beneficial because it automates the signaling of SR in the SP core. In this situation, the ACI border leaf switch advertises an EVPN type 5 route with a BGP color extended community to the DC-PE. The DC-PE can then create a segment routing policy based on the color community or destination prefix received from the ACI border leaf switch. This functionality allows seamless integration between the DC and the transport network.



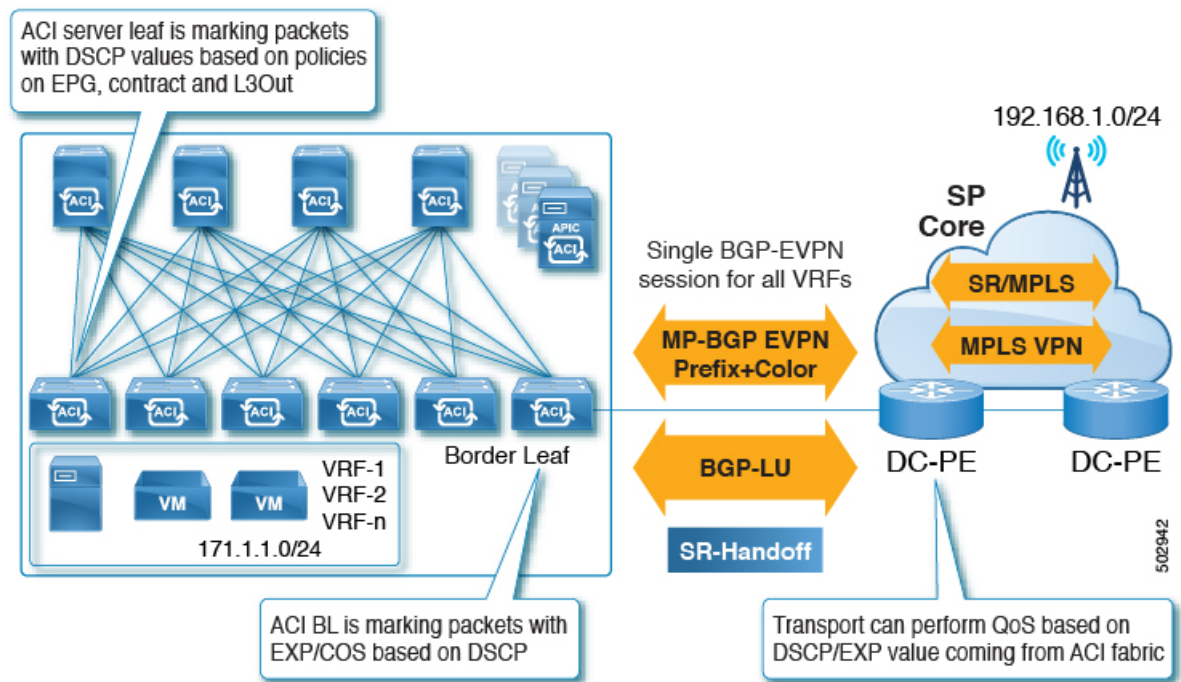
Similarly, you can advertise an EVPN type 5 prefix from the ACI border leaf switch and the DC-PE could create an SR-TE or Flex Algo routing policy based on the destination prefix, as shown in the following figure.



Of the two methods, we recommend using color community to reduce the configurations on the DC-PE. However, for either of these situations, you must verify that your DC-PE has the capability of supporting this functionality before utilizing SR-MPLS in this way.

### DC and Transport QoS with SR or MPLS

Within the ACI fabric, non-border leaf switches can mark packets with DSCP values using EPG, contract and L3Out QoS policies. Using these DSCP values, you can set MPLS egress rules on the ACI border leaf switch to then mark packets with experimental bits (EXP) or Class of Service (COS) values. The transport network can then perform QoS actions or pick different SR or MPLS paths, based on the DSCP or EXP values coming from the data center.



Similarly, using MPLS ingress rules, the ACI border leaf switch can mark the ingress packets coming into the fabric with COS, DSCP and QoS levels based on EXP values, where the QoS levels define the QoS actions within fabric.

## Understanding ACI Implementation of SR-MPLS Handoff

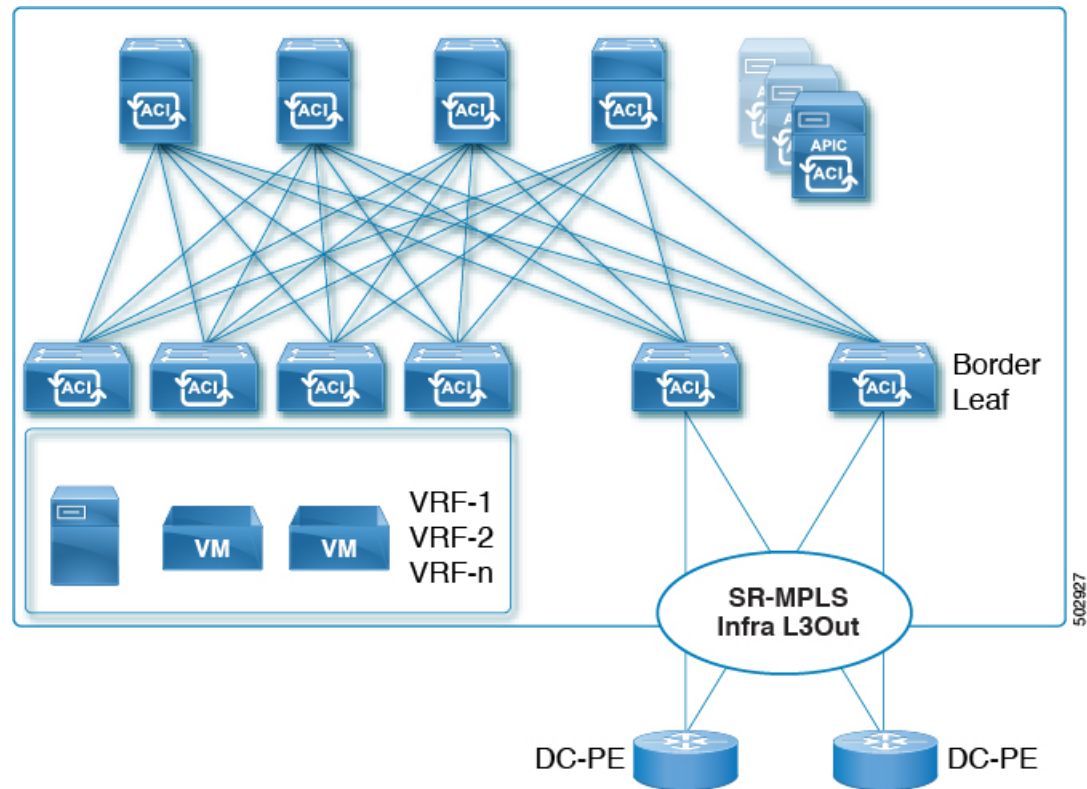
ACI implements SR-MPLS handoff using the following ACI components that have been introduced in Cisco APIC Release 5.0(1).

### SR-MPLS Infra L3Out

The SR-MPLS infra L3Out provides SR-MPLS connectivity. You will configure the SR-MPLS infra L3Out in the infra tenant on the border leaf switch to set up the underlay MP-BGP EVPN sessions for the SR-MPLS handoff. Tenant VRF instances are selectively mapped to the Cisco Application Centric Infrastructure (ACI) SR-MPLS infra L3Outs to advertise tenant subnets to the DC-PE routers and import MPLS VPN routes from

the DC-PE. An SR-MPLS infra L3Out is scoped to a pod or a remote leaf switch site, and is not extended across pods or remote leaf switch pairs.

**Figure 2: SR-MPLS Infra L3Out**



A pod or remote leaf switch site can have one or more SR-MPLS infra L3Outs.

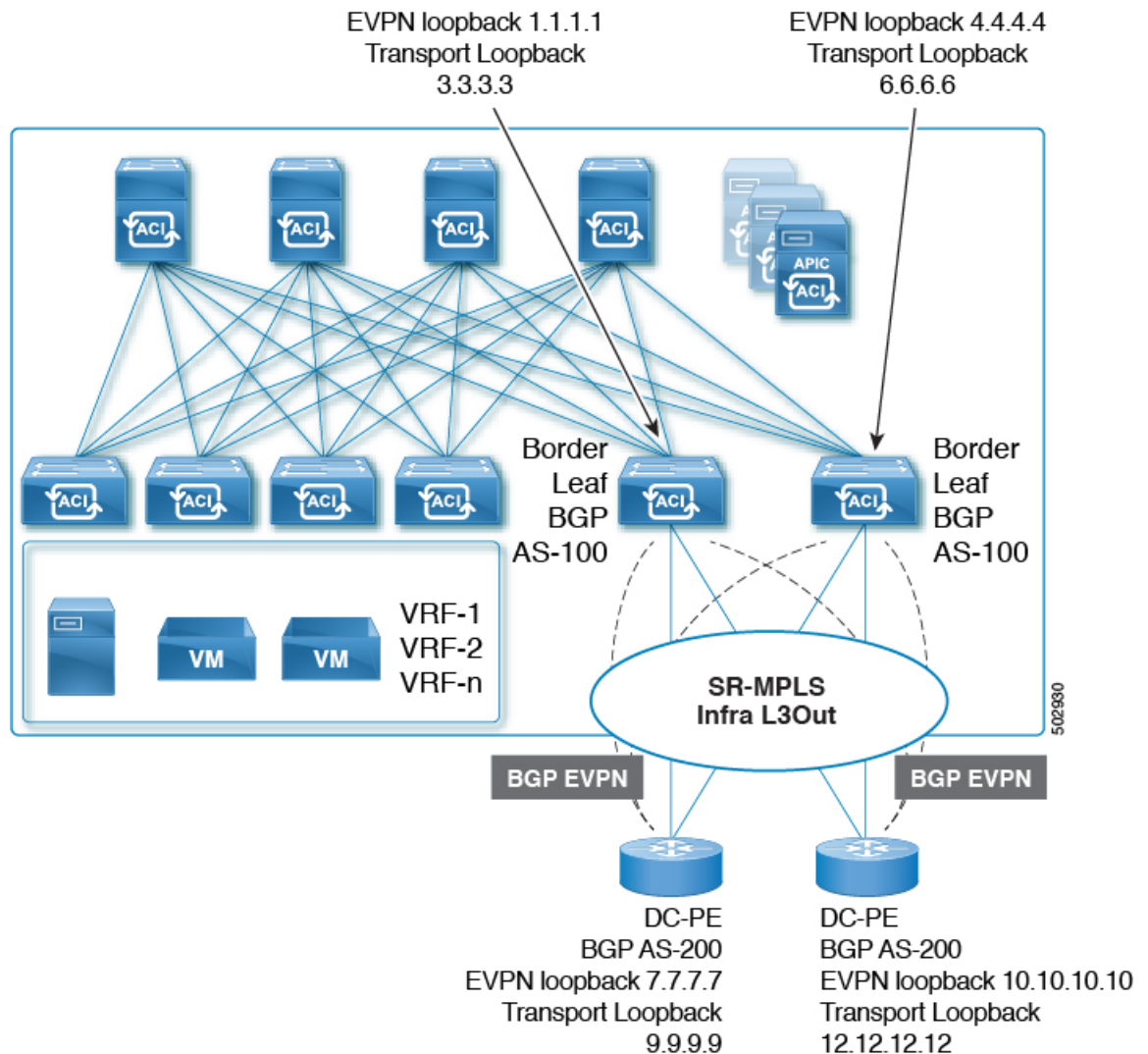
See [Configuring an SR-MPLS Infra L3Out Using the GUI](#), on page 26 for the procedures for configuring the SR-MPLS infra L3Out.

As part of the configuration process for the SR-MPLS infra L3Out, you will configure the following areas:

- [MP-BGP EVPN Session Between the Cisco ACI Border Leaf Switch and the DC-PE](#), on page 7
- [Multi-Hop BFD for BGP EVPN Session](#), on page 8
- [Underlay BGP Sessions \(BGP-Labeled Unicast and IPv4 Address-family\) On the Cisco ACI Border Leaf Switch and Next-Hop Router](#), on page 9
- [Single-Hop BFD for BGP-Labeled Unicast Session](#), on page 10

### **MP-BGP EVPN Session Between the Cisco ACI Border Leaf Switch and the DC-PE**

You will need to provide the necessary information to configure the MP-BGP EVPN sessions between the EVPN loopbacks of the border leaf switches and the DC-PE routers to advertise the overlay prefixes, as shown in the following figure.



The following configurations take place in this area:

- The label advertisement for the transport loopback using the BGP-labeled unicast address family.
- A unique router ID on the border leaf switch in the SR-MPLS infra VRF instance.
- The router ID should be different from the BGP-EVPN loopback and transport loopback addresses.

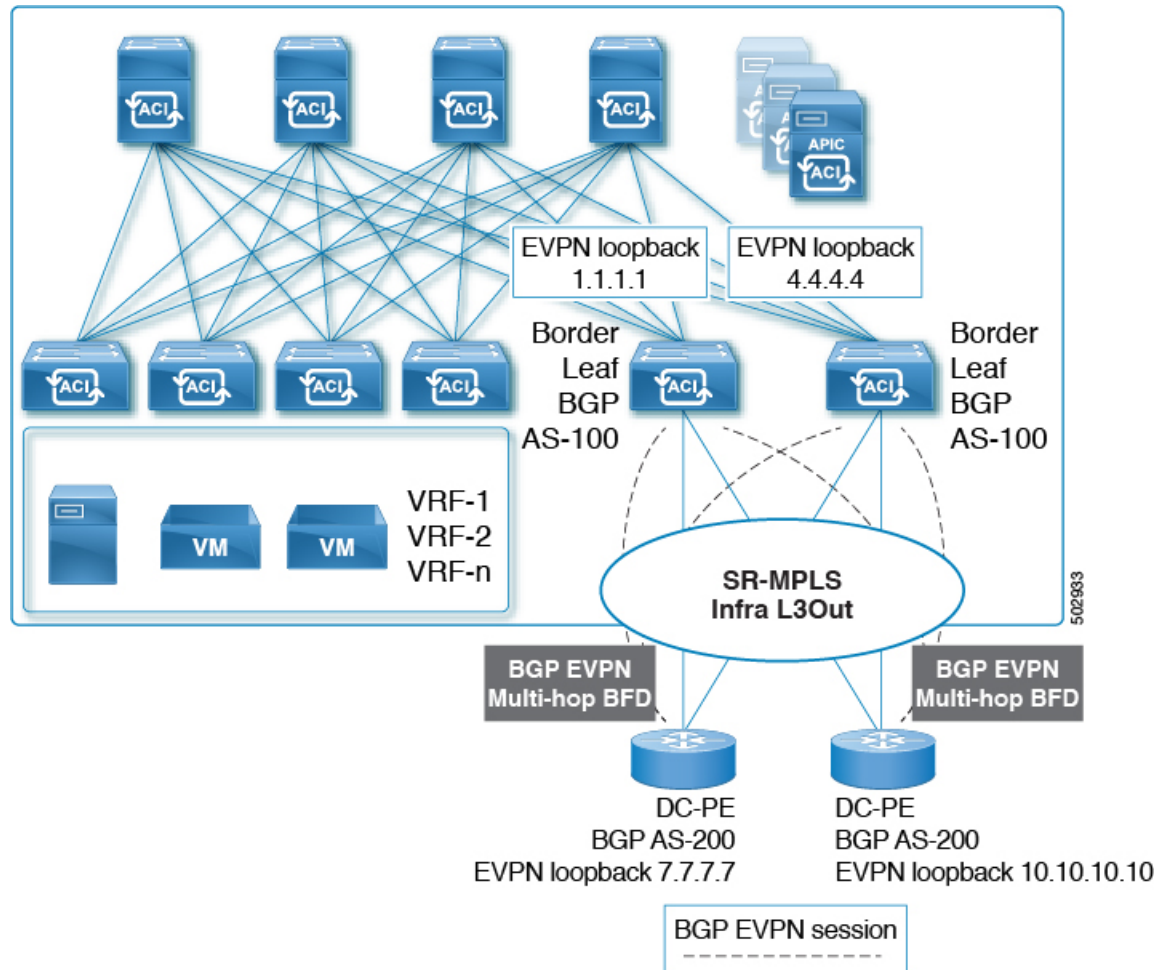
While you can use a different IP address for the loopback for the MP-BGP EVPN and the transport as shown in the figure, we recommend that you use the same loopback for the MP-BGP EVPN and the transport loopback on the Cisco ACI border leaf switch.

Only eBGP sessions are supported at this time.

### Multi-Hop BFD for BGP EVPN Session

Beginning with release 5.0(1), support is now available for multi-hop BFD, where you can configure multi-hop BFD EVPN sessions between EVPN loopbacks, as shown in the following figure.

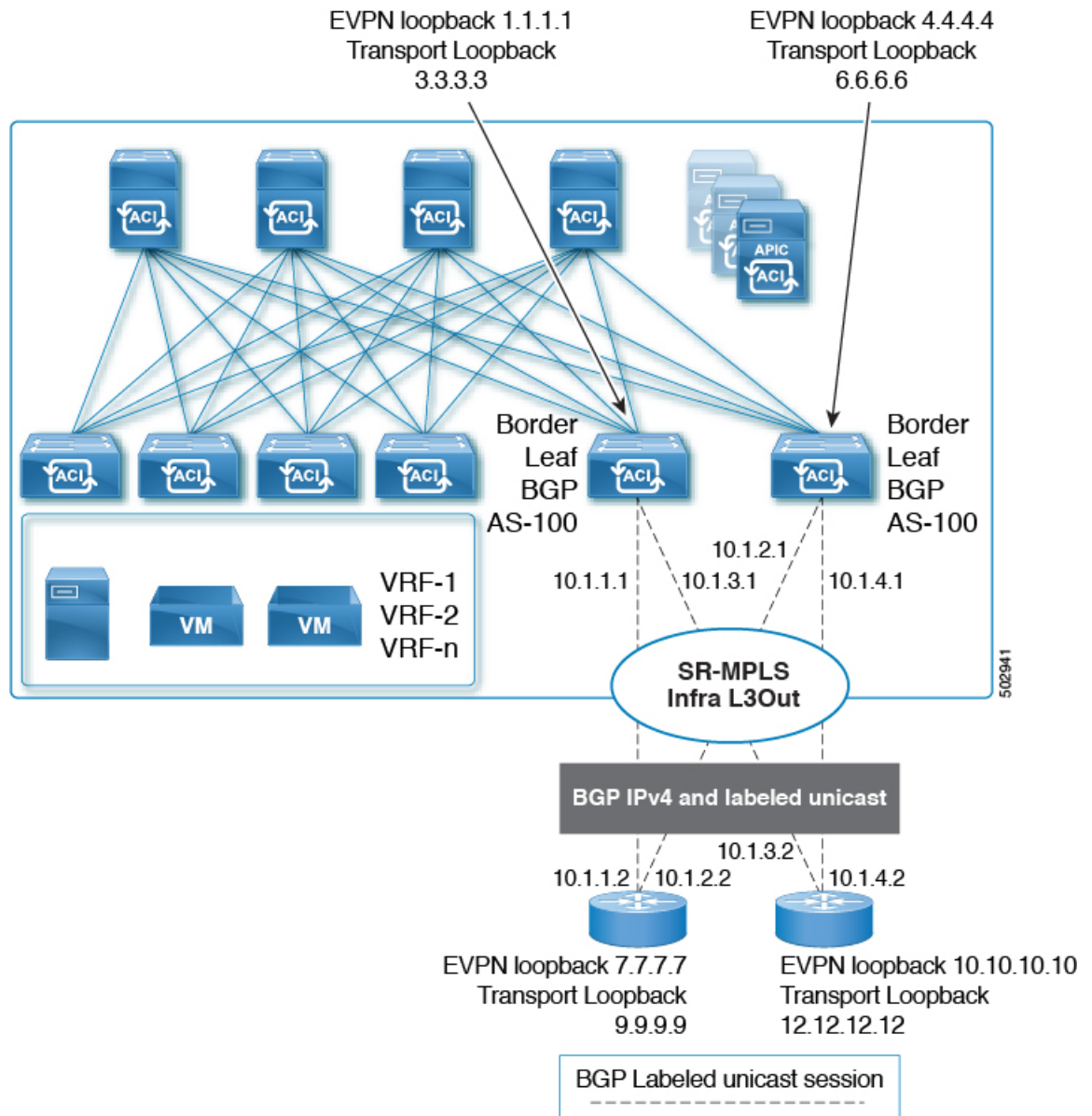




A multi-hop BFD with a minimum timer of 250 milliseconds and a detect multiplier of 3 is supported for the BGP EVPN session between the Cisco ACI border leaf switch and the DC-PE. You can modify this timer value based on your requirements.

### Underlay BGP Sessions (BGP-Labeled Unicast and IPv4 Address-family) On the Cisco ACI Border Leaf Switch and Next-Hop Router

You will also configure the BGP IPv4 and labeled unicast address-family per interface between the Cisco ACI border leaf switches and the DC-PE, as shown in the following figure.

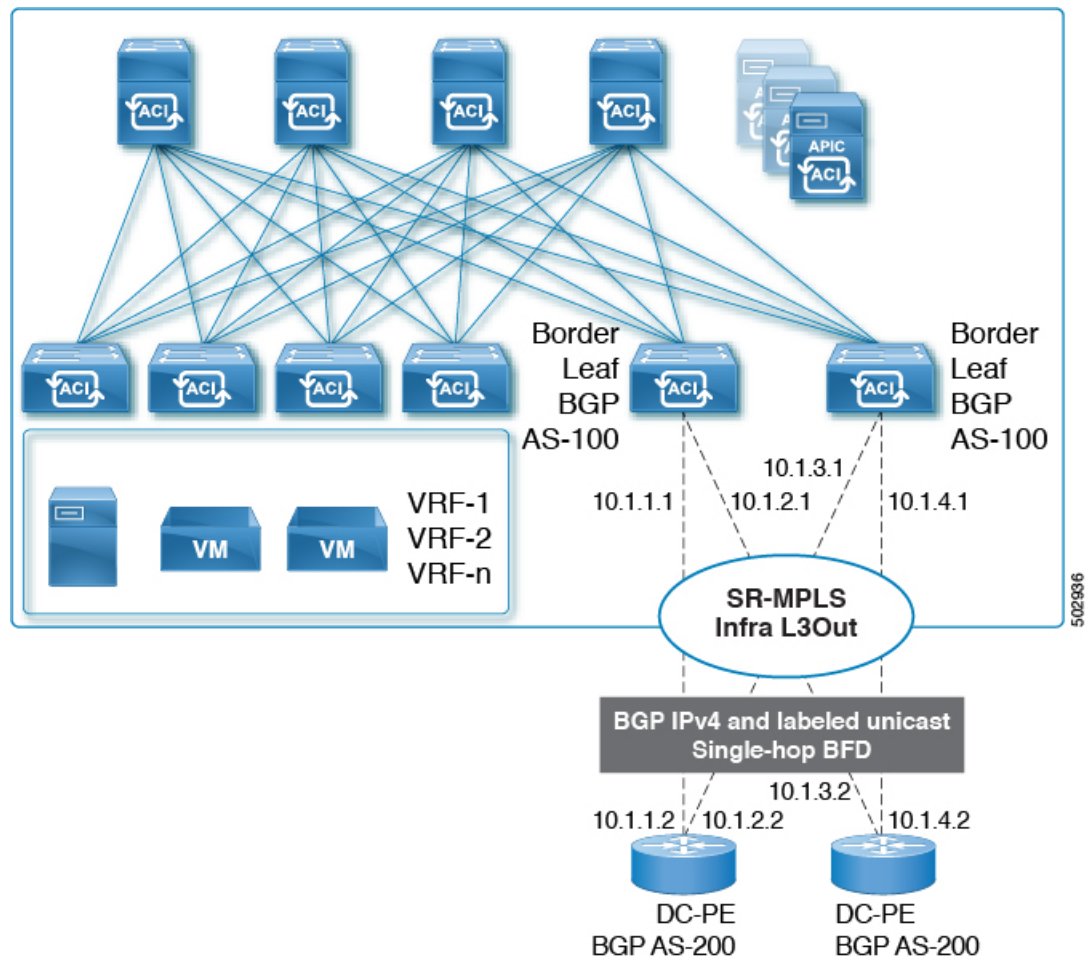


The BGP IPv4 address family automatically advertises the EVPN loopbacks, and the BGP-labeled unicast address family will automatically advertise the SR transport loopback with the SR-MPLS label.

Only eBGP sessions are supported at this time.

### Single-Hop BFD for BGP-Labeled Unicast Session

To prevent an issue related to soft failure, where the link remains up but the forwarding capability of the link is impacted, you can configure a single-hop BFD session for the underlay BGP session for the IPv4 and BGP-labeled unicast session, as shown in the following figure.



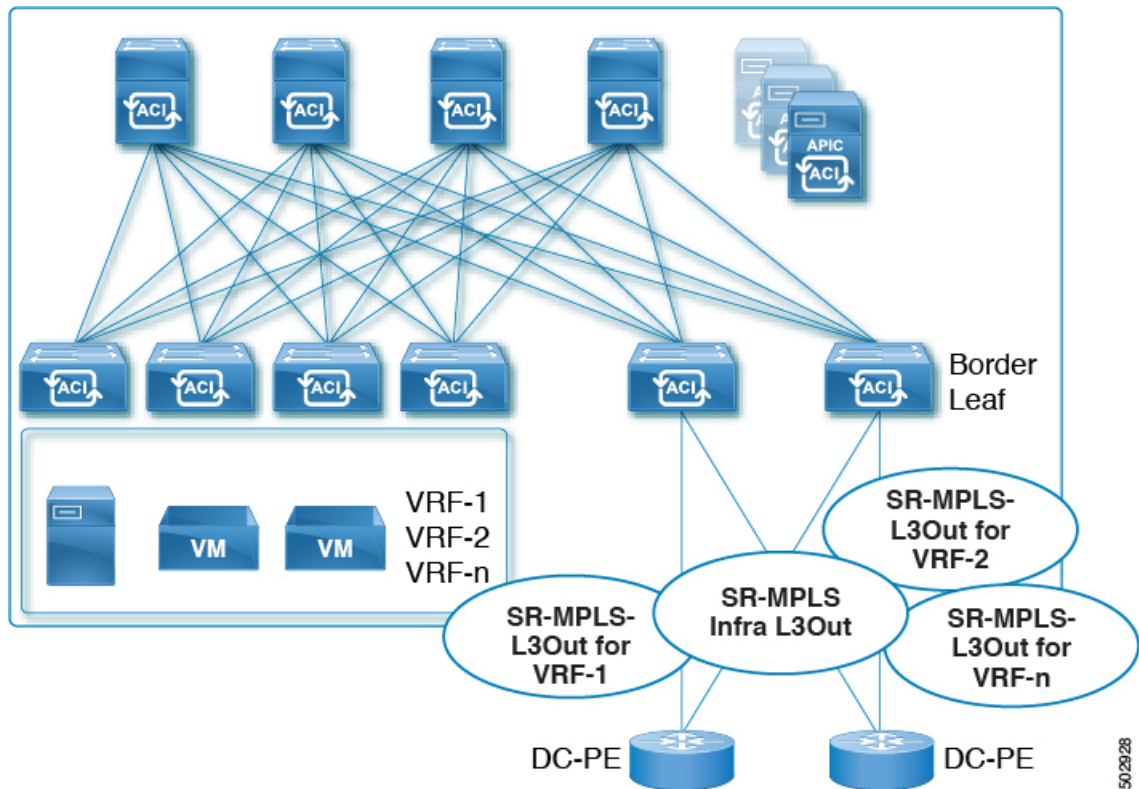
A single-hop BFD with a minimum timer of 50 milliseconds and a detect multiplier of 3 is supported for the BGP EVPN session between the Cisco ACI border leaf switch and the DC-PE. You can modify this timer value based on your requirements.

You can configure the BFD echo function on one or both ends of a BFD-monitored link. The echo function slows down the required minimum receive interval, based on the configured slow timer. The *RequiredMinEchoRx* BFD session parameter is set to *zero* if the echo function is disabled. The slow timer becomes the required minimum receive interval if the echo function is enabled.

## SR-MPLS VRF L3Out

Each VRF, whose prefixes need to be advertised towards an SR-MPLS transport, must be associated with the SR-MPLS infra L3Out. You will configure these associations using the SR-MPLS VRF L3Outs, which are attached to the SR-MPLS infra L3Out.

Figure 3: User Tenant SR-MPLS L3Out



You can attach one or more SR-MPLS VRF L3Outs to the same SR-MPLS infra L3Out. Through the SR-MPLS VRF L3Outs, you can configure import and export route maps to do the following things:

- Apply route policies based on prefixes and/or communities
- Advertise prefixes into the SR network
- Filter out prefixes received from the SR network

You will also configure an external EPG with one or more subnets on each SR-MPLS VRF L3Out tenant, which is used for the following:

- Security policies (contract)
- Policy-Based Redirect (PBR) policies
- Route leaking between VRFs

See [Configuring an SR-MPLS VRF L3Out Using the GUI, on page 33](#) for the procedures for configuring SR-MPLS VRF L3Outs.

## SR-MPLS Custom QoS Policy

You can use custom QoS policies to define how traffic coming from an MPLS network is prioritized within the ACI fabric. You can also use these policies to re-mark the traffic when it leaves the fabric via an MPLS L3Out.

When configuring a custom QoS policy, you define the following two rules that are applied on the border leaf switch:

- **Ingress rules:** Any traffic coming into the border leaf switch connected to the MPLS network will be checked for the MPLS experimental bits (EXP) value and if a match is found, the traffic is classified into an ACI QoS Level and marked with appropriate CoS and differentiated services code point (DSCP) values.

The values are derived at the border leaf using a custom QoS translation policy. The original DSCP values for traffic coming from SR-MPLS are retained without any remarking. If a custom policy is not defined or not matched, the default QoS Level (Level13) is assigned.

- **Egress rules:** When the traffic is leaving the fabric out of the border leaf's MPLS interface, it will be matched based on the DSCP value of the packet and if a match is found, the MPLS EXP and CoS values will be set based on the policy.

If the egress MPLS QoS policy is not configured, the MPLS EXP will default to zero. If they are configured based on the MPLS Custom QoS policy, it will remark the EXP.

The following two figures summarize when the ingress and egress rules are applied as well as how the internal ACI traffic may remark the packets' QoS fields while inside the fabric.

**Figure 4: Ingress QoS**

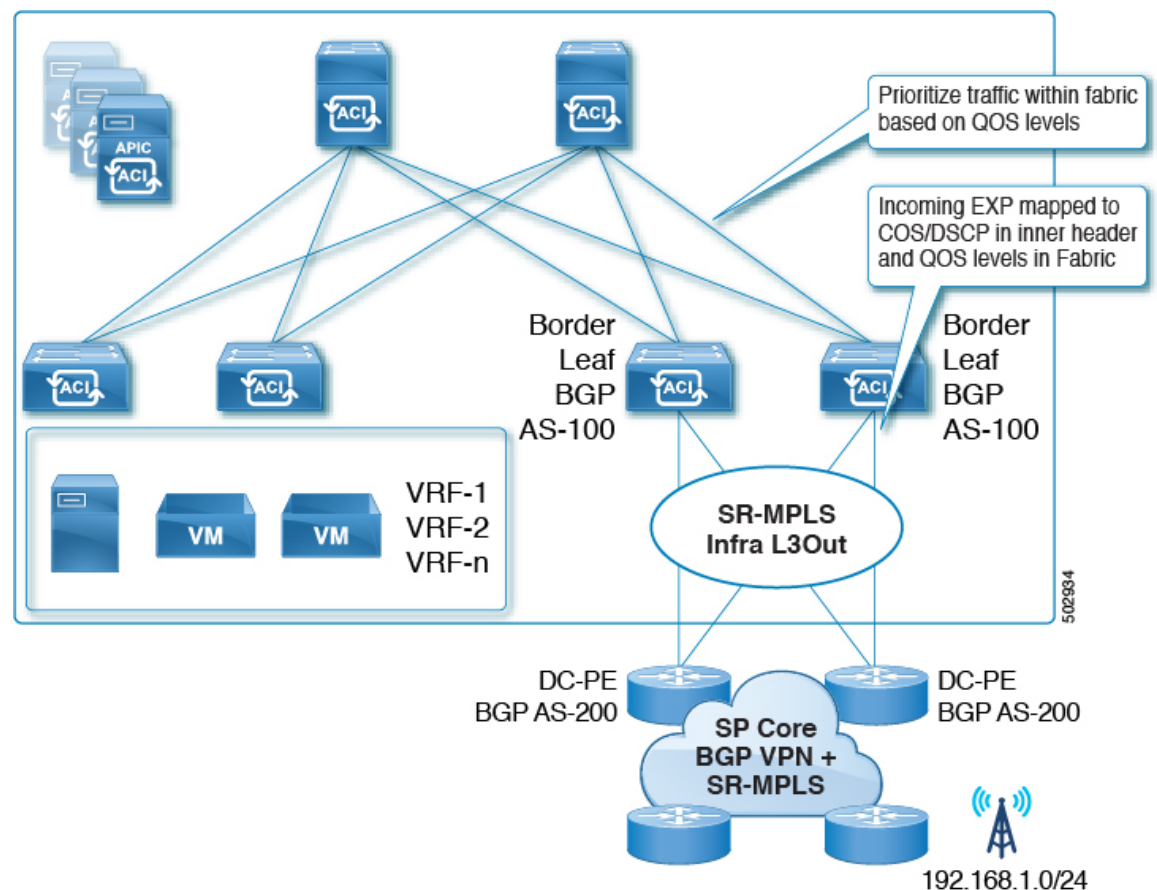
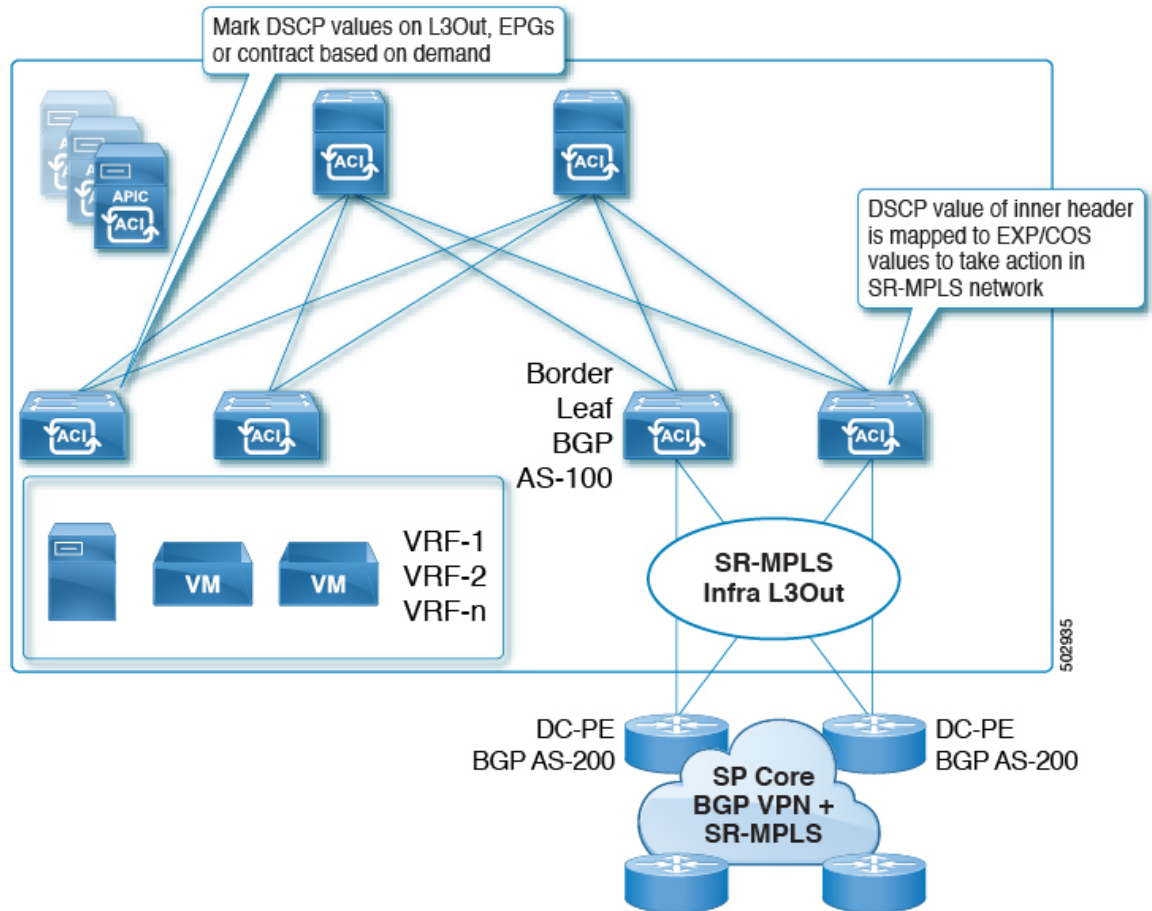


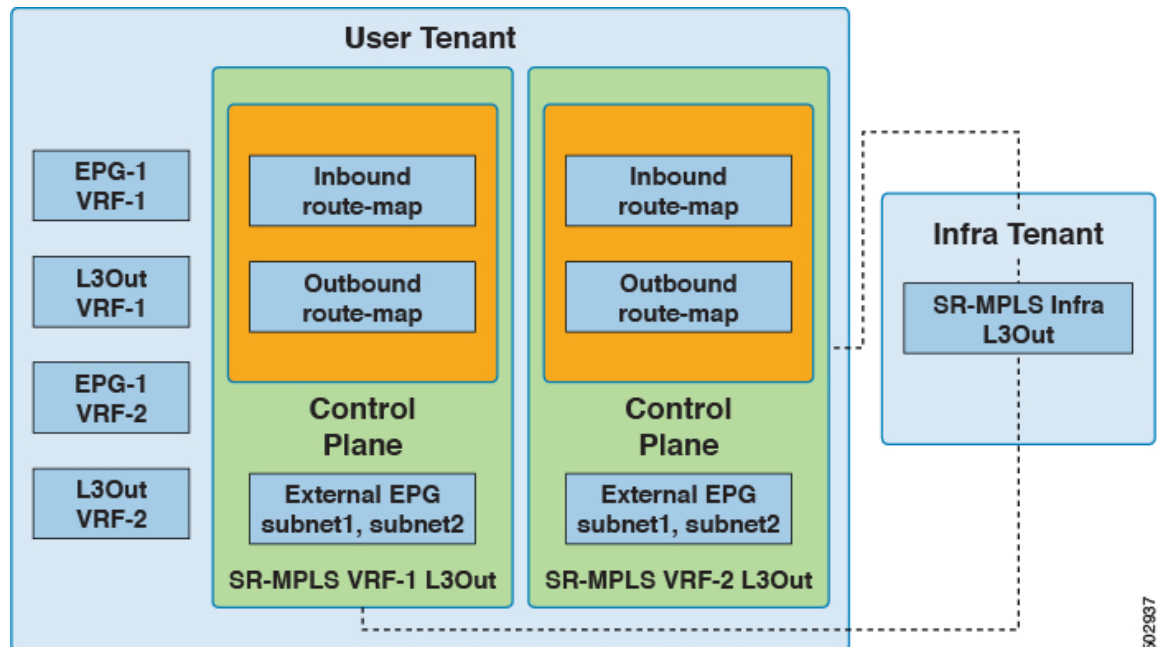
Figure 5: Ingress QoS



You can define multiple custom QoS policies and apply them to each SR-MPLS Infra L3Out you create, as described in [Creating SR-MPLS Custom QoS Policy Using the GUI](#), on page 36.

## Understanding the SR-MPLS Configuration Model

The following figure shows the configuration model for the ACI implementation of SR-MPLS handoff.



502937

Configuration of SR-MPLS handoff takes place within these tenants:

- **Infra Tenant:** Underneath the infra tenant, you will configure the SR-MPLS infra L3Out, as described in [SR-MPLS Infra L3Out](#), on page 6. The SR-MPLS infra L3Out is where you define the connectivity between the ACI fabric and the external devices connected to the border leaf switches. You will specify the overlay and underlay node path in the SR-MPLS infra L3Out.
- **User Tenant:** Underneath the user tenant, you may have multiple VRFs, EPGs, and L3Outs, as shown in the left area in the figure. Within the user tenant, you will configure the SR-MPLS VRF L3Out that you will use as part of the SR-MPLS handoff configuration, as described in [SR-MPLS VRF L3Out](#), on page 11.

Within the SR-MPLS VRF L3Out, you will also configure the following route maps:

- **Inbound route map:** By default, the policy for the inbound route map is to accept all prefixes. An explicit inbound route map can be configured to:
  - Match prefixes to selectively deny their advertisement inside the fabric
  - Match prefixes and community to selectively deny their advertisement inside the fabric
- **Outbound route map:** You must configure the policy for the outbound route map to advertise any prefix, including bridge domain subnets. By default, the policy for the outbound route map is to not advertise any prefix. An explicit outbound route map can be configured to:
  - Match prefixes to be advertised to the SR-MPLS network
  - Match prefixes and community to advertise prefixes to the SR-MPLS network
  - Set community, including color community, based on the prefix and/or community match

Both the inbound route map and the outbound route map are used for the control plane, to set which prefixes are permitted or denied in and out of the fabric.

Within the SR-MPLS VRF L3Out, you will also configure the external EPG and the subnets within this external EPG, which is used for the data plane. These subnets will be used to apply ACI security policies. The external EPG subnet is also used to leak prefixes in another VRF using flags. If you enable the route-leak and security flag on an external EPG subnet, then that subnet can be leaked to another VRF. You can also configure the external EPG subnet with the aggregated flag to leak prefixes to another VRF. In this case, you will need to define a contract to the leaf switch prefixes and allow communication across VRFs.



---

**Note** The external EPG on the SR-MPLS VRF L3Out is not used for routing policies, such as applying a route map to advertise or deny prefix advertisement.

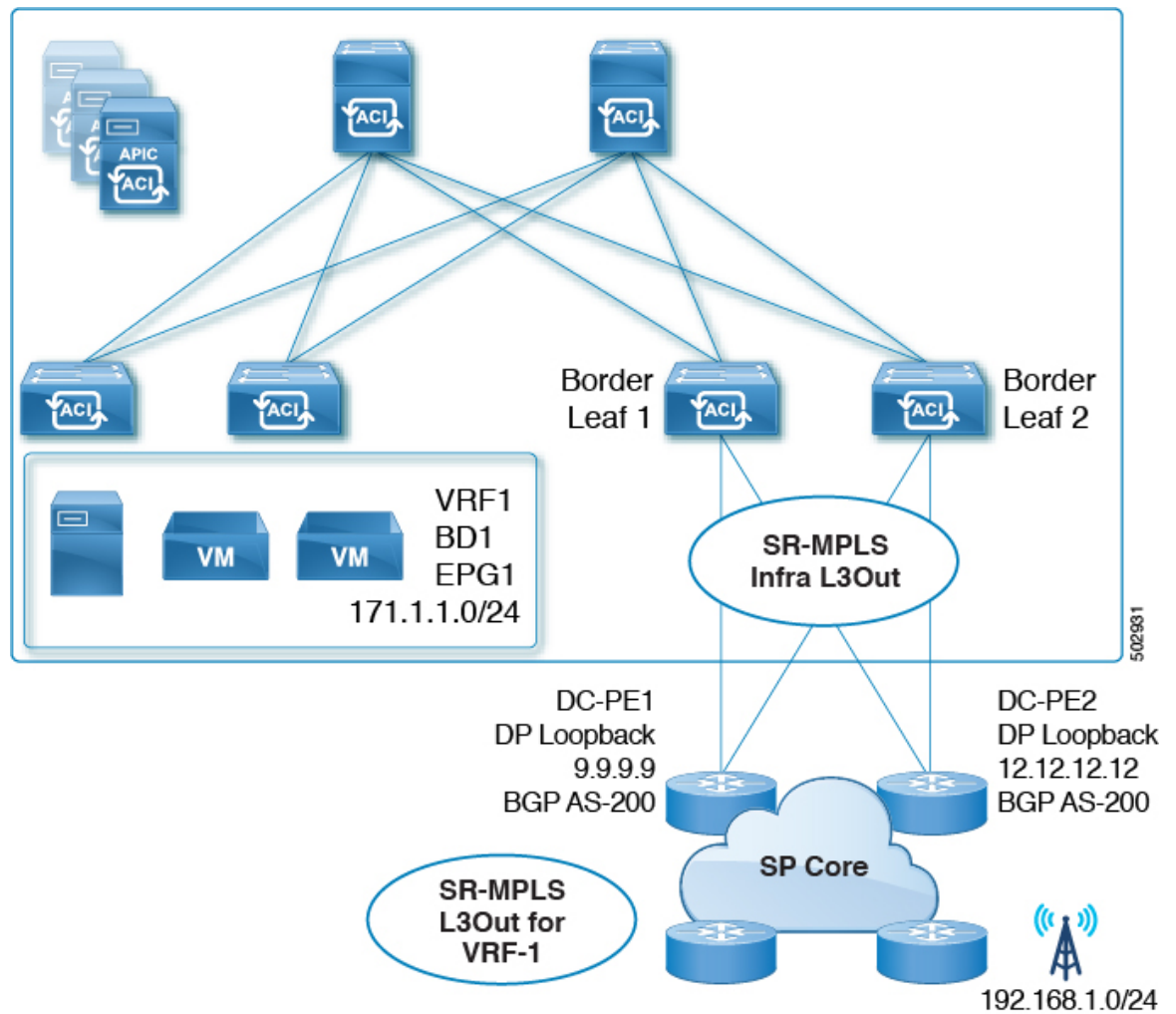
---

In this example, the SR-MPLS VRF-1 L3Out within the user tenant is attached to the SR-MPLS infra L3Out, and the SR-MPLS VRF-2 L3Out within the user tenant is also attached to the SR-MPLS infra L3Out.

#### **EPG to SR-MPLS L3Out**

The following figure shows an example of an EPG to SR-MPLS L3Out configuration.



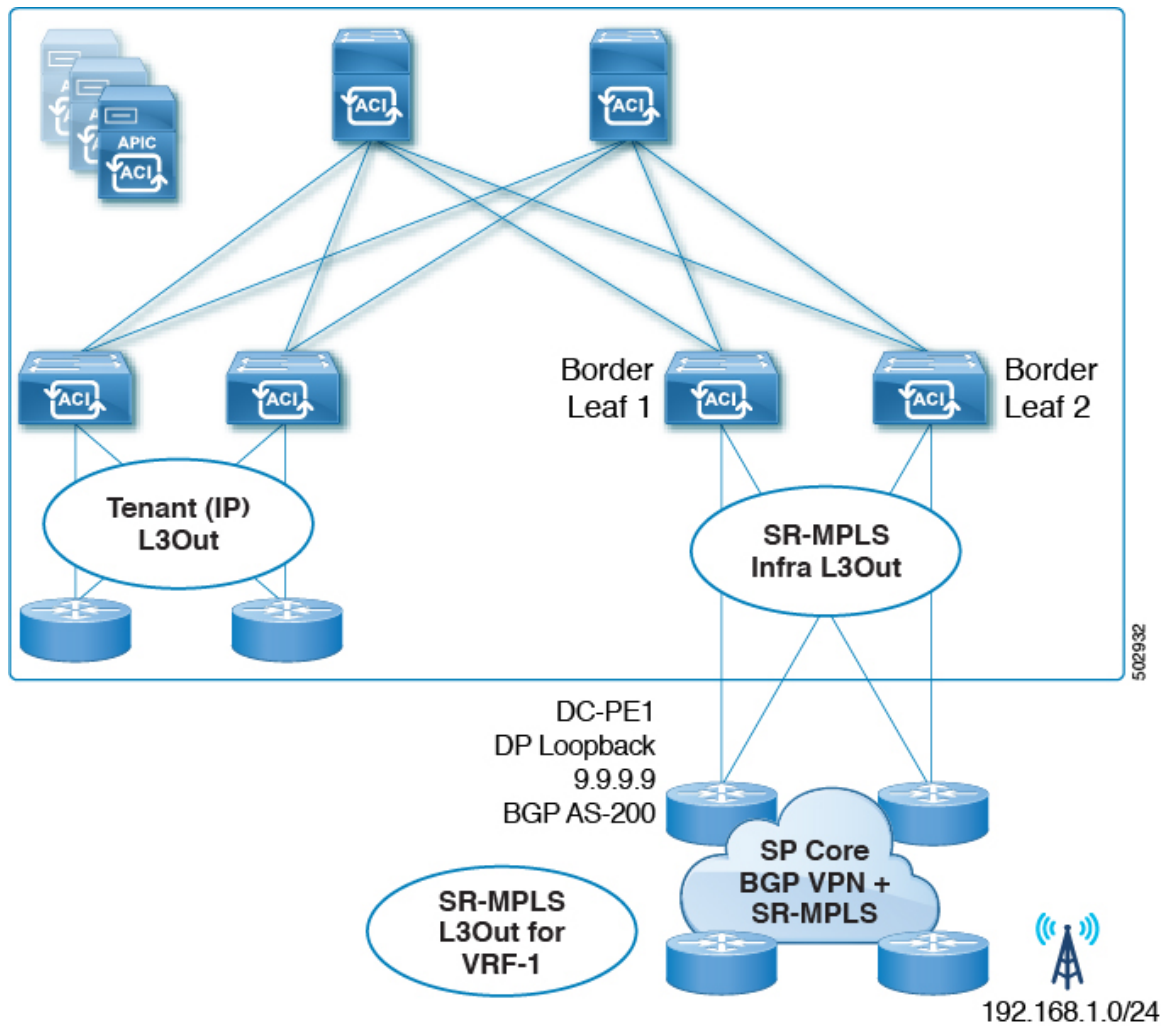


In this scenario, you would make the following configurations:

- Configure the SR-MPLS infra L3Out on the border leaf switches (BL1 and BL2 in the figure above)
- Configure the SR-MPLS VRF L3Out in the user tenant, along with the EPG, bridge domain and user VRFs
- Configure the route map for exporting and importing on prefixes and apply it to the SR-MPLS VRF L3Out
- Configure the contract and apply it between the EPG and the external EPG defined on the SR-MPLS VRF L3Out for traffic forwarding between the EPG and the SR-MPLS L3Out

#### IP L3Out to SR-MPLS L3Out

The following figure shows an example of a configuration to enable transit routing between a regular IP L3Out and an SR-MPLS L3Out.



In this scenario, you would make configurations similar to the EPG to SR-MPLS L3Out configuration described previously, with the differences highlighted below:

- Configure the SR-MPLS infra L3Out on the border leaf switches (BL1 and BL2 in the figure above)
- Configure the SR-MPLS VRF L3Out in the user tenant, along with the **IP L3Out** and user VRFs
- Configure the route map for exporting and importing on prefixes and apply it to the SR-MPLS VRF L3Out
- Configure the contract and apply it between the external EPGs associated to the IP L3Out and the SR-MPLS VRF L3Out for traffic forwarding between the **IP L3Out** and the SR-MPLS L3Out

## Guidelines and Limitations for SR-MPLS

Following are the guidelines and limitations for the SR-MPLS handoff feature.

- [Supported Platforms for SR-MPLS, on page 19](#)

- [Platform Limitations for SR-MPLS, on page 19](#)
- [Guidelines and Limitations for the SR-MPLS Infra L3Out, on page 20](#)
- [Guidelines and Limitations for the SR-MPLS VRF L3Out, on page 20](#)
- [Guidelines and Limitations for MPLS Custom QoS Policies, on page 25](#)
- [Guidelines and Limitations for SR-MPLS Statistics, on page 26](#)

### Supported Platforms for SR-MPLS

The SR-MPLS handoff feature is supported on the following platforms:

- **Border leaf switches:** -FX switch models and later (for example, switch models with "FX", "FX2", "FX3", "GX", and so on at the end of the switch name)
- **Spine switches:**
  - Modular spine switch models with "LC-EX", "LC-FX", and "GX" at the end of the linecard name
  - Fixed spine switches Cisco Nexus 9000 series N9K-C9332C and N9K-C9364C
- **DC-PE routers:**
  - Network Convergence System (NCS) 5500 Series
  - ASR 9000 Series
  - NCS 540 or 560 routers
  - ASR1000/ IOS-XE platforms
- The Cisco Application Centric Infrastructure (ACI)-to-SR-MPLS handoff solution uses a standards-based implementation with SR-MPLS, BGP-LU, BGP EVPN, and prefix re-origination between BGP EVPN and VPNv4/v6. Any DC-PE that supports these technologies should be able to support Cisco ACI to SR-MPLS handoff.



---

**Note** When the Cisco Application Centric Infrastructure (ACI) border leaf switch with the SR-MPLS handoff is connected to a PE device running IOS-XE software, the IOS-XE device should be configured with "neighbor <aci-leaf> next-hop-unchanged" under the BGP L2VPN EVPN address-family. With next-hop-unchanged configuration, Cisco ACI border leaf switch must learn the remote PE loopback.

---

### Platform Limitations for SR-MPLS

- On the FX platform, enabling the SR-MPLS feature would enable MPLS parsing on all ports, including ports where MPLS is not enabled or deployed. On FX2 platforms and later, MPLS parsing is enabled only on ports where SR-MPLS is enabled or deployed.
- On ports where MPLS parsing is enabled, pure Layer 2 switching of MPLS encapsulated packets are not supported. Non-MPLS Layer 2 traffic can use the Cisco ACI fabric as Layer 2 transit, without any issues.

### Guidelines and Limitations for the SR-MPLS Infra L3Out

- Even though a border leaf switch can be in multiple SR-MPLS infra L3Outs, a border leaf switch/provider edge router combination can only be in one SR-MPLS infra L3Out as there can be only one routing policy for a user VRF/border leaf switch/DC-PE combination.
- If there is a requirement to have SR-MPLS connectivity from multiple pods and remote locations, ensure that you have a different SR-MPLS infra L3Out in each of those pods and remote leaf locations with SR-MPLS connectivity.
- SR-MPLS infra L3Outs do not support multicast.

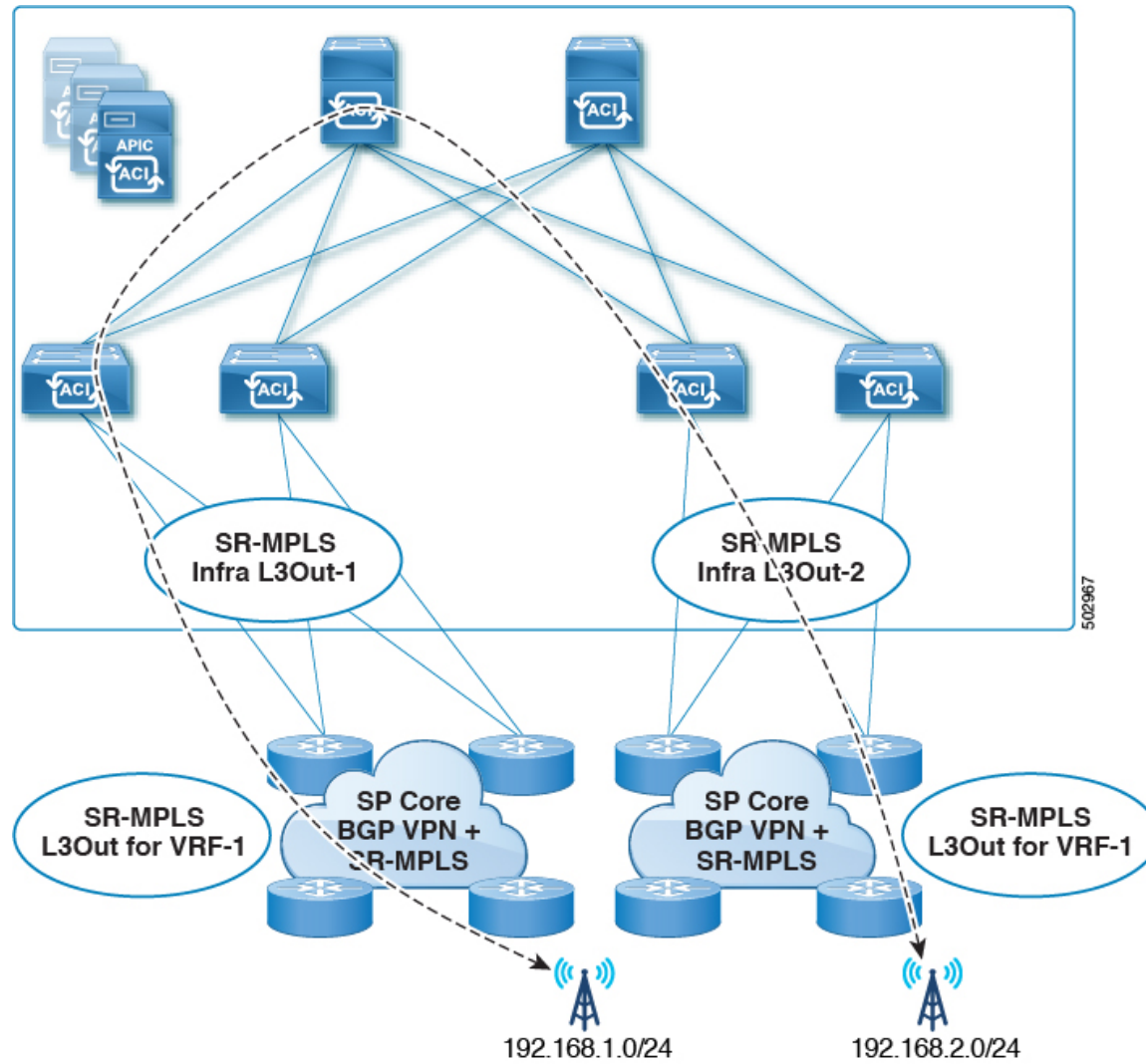
### Routing Policy

- **Supported:** Beginning with Cisco APIC release 6.1(1), fabric ports on a remote leaf can now be deployed on SRMPLS infra l3outs, as a routed sub interface.

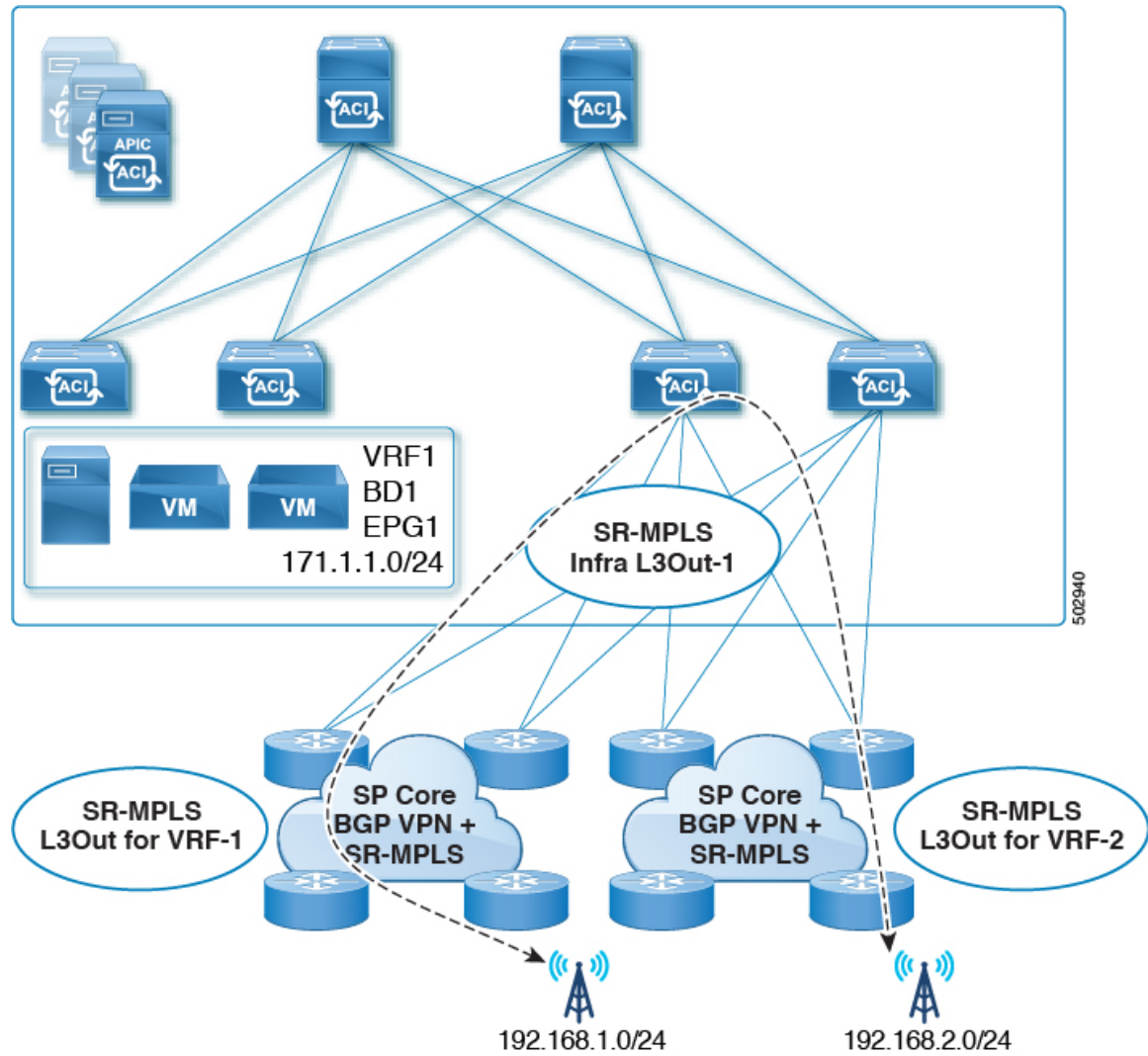
### Guidelines and Limitations for the SR-MPLS VRF L3Out

#### Routing Policy

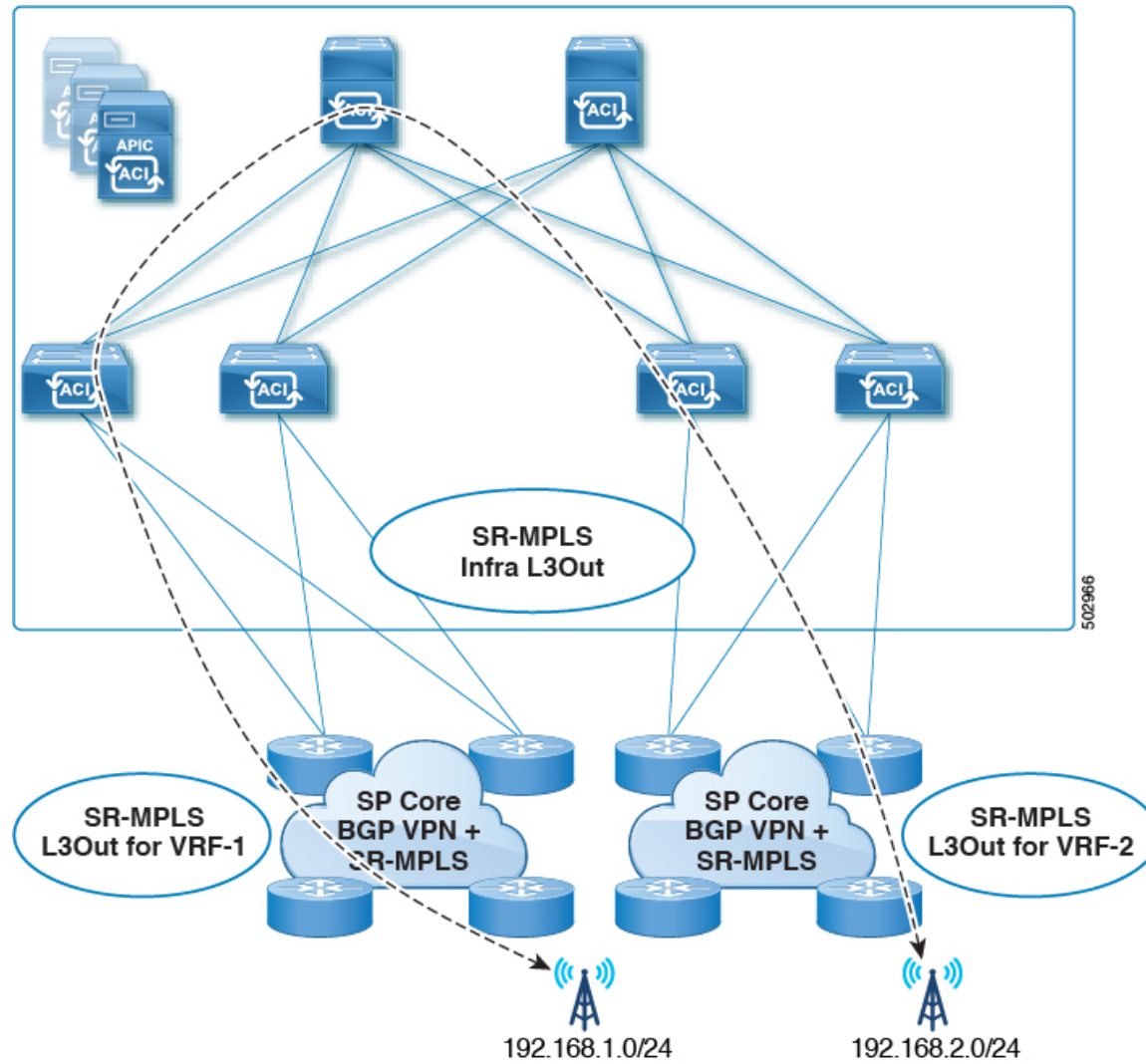
- Within each SR-MPLS VRF L3Out, defining the outbound route map (export routing policy) is mandatory, but defining the inbound route map (import routing policy) is optional.
- Routing policies associated with any SR-MPLS VRF L3Outs have to be a global type. In other words, you have to explicitly add all the routes, including bridge domain subnets.
- Host-based routing is not supported with SR-MPLS.
- Transit routing is supported, but with some restrictions:
  - **Supported:** Transit SR-MPLS traffic **with a single VRF using different border leaf pairs**, as shown in the following figure. For this configuration, you must advertise the unique prefixes range through each SR-MPLS infra L3out (border leaf pair). You must also ensure that there is no routing loop in the transport network (in other words, that the fabric is acting as a hub with the two-transport network acting as a spoke).



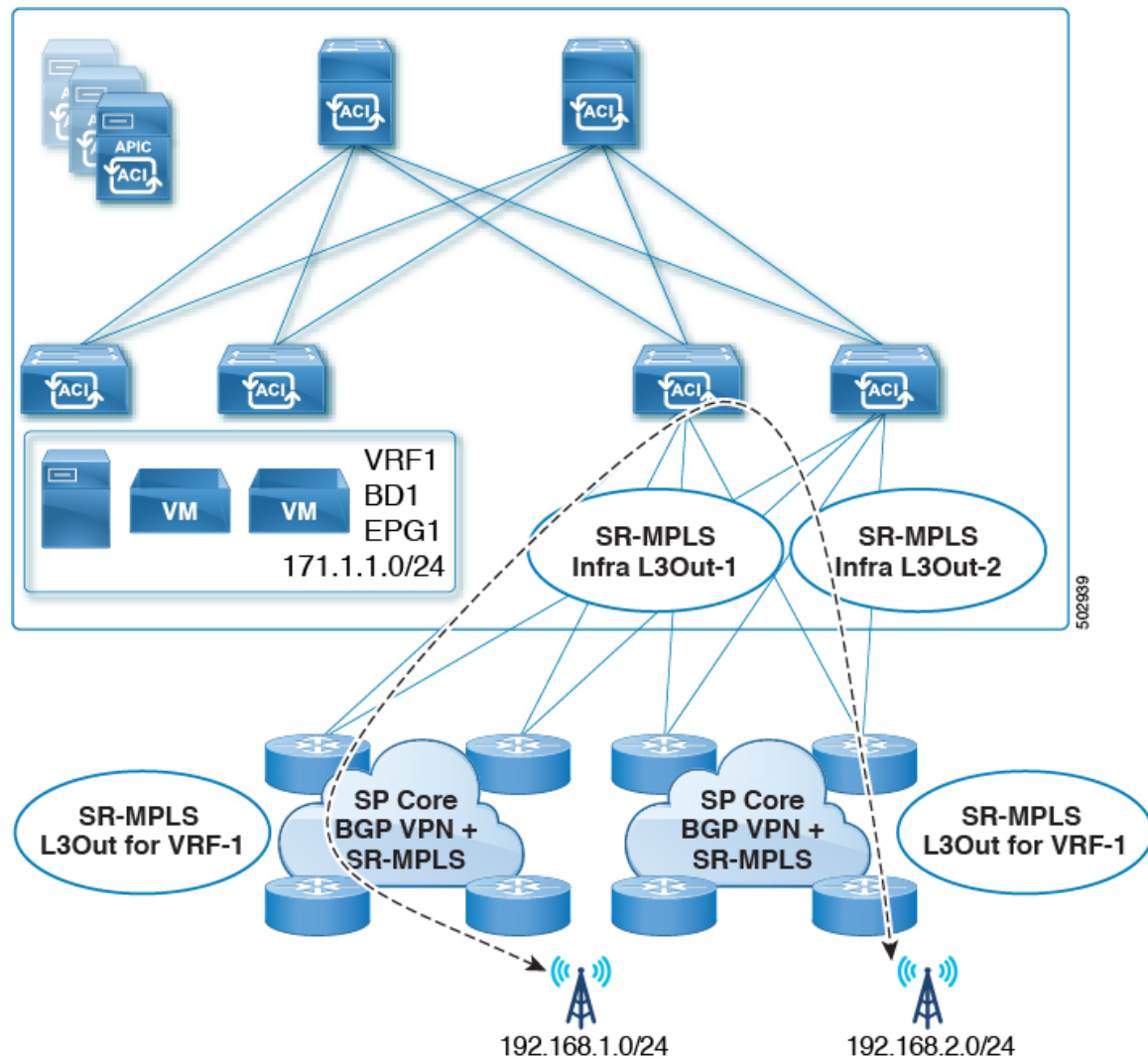
- **Supported:** Transit SR-MPLS traffic with the same border leaf pair and different VRFs, as shown in the following figure.



- **Supported:** Transit SR-MPLS traffic **with different border leaf pairs and different VRFs**, as shown in the following figure.



- Transit SR-MPLS traffic **within the same VRF and on the same border leaf pair**, as shown in the following figure:
  - **Unsupported** for releases prior to Release 5.1(1).
  - **Supported** for Release 5.1(1) and later, where re-originated routes are prevented from being advertised back into the same Infra L3Out peers to avoid transient loops in the system.



- If a leaf switch is configured on multiple SR-MPLS infra L3Outs, the same subnets can be advertised out of all the L3Outs if the prefixes are configured in a single prefix list (in one match rule), and the route map with that prefix list is then associated with all the SR-MPLS VRF L3Outs.

For example, consider the following configuration:

- A single prefix list P1, with subnets S1 and S2
- SR-MPLS VRF L3Out 1, which is associated with route map R1, with prefix list P1
- SR-MPLS VRF L3Out 2, which is associated with route map R2, with prefix list P1

Because the prefixes are configured in the same prefix list (P1), even though they are associated with different SR-MPLS VRF L3Outs, the same subnets within prefix list P1 are advertised out of both L3Outs.

On the other hand, consider the following configuration:

- Two prefix lists:
  - Prefix list P1, with subnets S1 and S2



- Prefix list P2, with subnets S1 and S2
- SR-MPLS VRF L3Out 1, which is associated with route map R1, with prefix list P1
- SR-MPLS VRF L3Out 2, which is associated with route map R2, with prefix list P2

Because the prefixes are configured in the two prefix lists (P1 and P2), and they are associated with different SR-MPLS VRF L3Outs, subnets S1 and S2 are not advertised out of both of the L3Outs.

- SR-MPLS VRF L3Outs do not support multicast.

### Security Policy

- You can configure a security policy through the external EPG instance profile, which is defined within an SR-MPLS VRF L3Out. The external EPG instance profile contains IP prefixes that are reachable through the SR-MPLS network from one or more SR-MPLS infra L3Outs and need the same security policy.
- You can configure 0/0 prefix in the external EPG instance profile to classify, as part of the external EPG, the inbound traffic flows originated from any external IP address.
- You can associate an external EPG in the external EPG instance profile with one or more SR-MPLS VRF L3Outs. When the external EPG instance profile is external to multiple SR-MPLS infra L3Outs, multiple SR-MPLS VRF L3Outs point to the same external EPG instance profile.
- You must configure contracts between local EPGs and external EPG instance profiles or between external EPGs associated to different VRF L3Outs (to enable transit routing).

### Guidelines and Limitations for MPLS Custom QoS Policies

Following is the default MPLS QoS behavior:

- Class of Service (COS) preservation is not supported for intra ToR MPLS egress QoS policies, where the destination port is an MPLS port.
- All incoming MPLS traffic on the border leaf switch is classified into QoS Level 3 (the default QoS level).
- The border leaf switch will retain the original DSCP values for traffic coming from SR-MPLS without any remarking.
- The border leaf switch will forward packets with the default MPLS EXP (0) to the SR-MPLS network.

Following are the guidelines and limitations for configuring MPLS Custom QoS policies:

- Data Plane Policers (DPP) are not supported at the SR-MPLS L3Out.
- Layer 2 DPP works in the ingress direction on the MPLS interface.
- Layer 2 DPP works in the egress direction on the MPLS interface in the absence of an egress custom MPLS QoS policy.
- VRF level policing is not supported.

### Guidelines and Limitations for SR-MPLS Statistics

Following are the guidelines and limitations for SR-MPLS statistics:

- To see the SR-MPLS statistics, you have to perform a one-time stateful reload when enabling the SR-MPLS configuration on any leaf switch.
- The SR-MPLS interface statistics are only supported only on border leaf switch models with "FX2" or "GX" at the end of the switch name.
- The SR-MPLS VRF instance statistics are supported on border leaf switch models with "FX," "FX2", or "GX" at the end of the switch name.
- For the 15 minute historic stats, it might take 20 minutes to update the 15 minute interval data.
- SR-MPLS interface statistics shown in a switch's CLI get cleared after an admin or operational down event.
- SR-MPLS interface statistics in a switch's CLI are reported every 10 seconds. If, for example, an interface goes down 3 seconds after the collection of the statistics, the CLI reports only 3 seconds of the statistics and clears all of the other statistics.

## Configuring an SR-MPLS Infra L3Out Using the GUI

- The SR-MPLS infra L3Out is configured on the border leaf switch, which is used to set up the underlay BGP-LU and overlay MP-BGP EVPN sessions that are needed for the SR-MPLS handoff.
- An SR-MPLS infra L3Out will be scoped to a pod or a remote leaf switch site.
- A pod or remote leaf switch site can have one or more SR-MPLS infra L3Outs.

You will configure the following pieces when configuring the SR-MPLS infra L3Out:

#### • Nodes

- Only leaf switches are allowed to be configured as nodes in the SR-MPLS infra L3Out (border leaf switches and remote leaf switches).
- Each SR-MPLS infra L3Out can have border leaf switches from one pod or remote leaf switch from the same site.
- Each border leaf switch or remote leaf switch can be configured in multiple SR-MPLS infra L3Outs if it connects to multiple SR-MPLS domains.
- You will also configure the loopback interface underneath the node, and a node SID policy underneath the loopback interface.

#### • Interfaces

- Supported types of interfaces are:
  - Routed interface or sub-interface
  - Routed port channel or port channel sub-interface

For sub-interfaces, any VLAN tag is supported.

- You will also configure the underlay BGP peer policy underneath the interfaces area in the SR-MPLS infra L3Out.

- **QoS rules**

- You can configure the MPLS ingress rule and MPLS egress rule through the MPLS QoS policy in the SR-MPLS infra L3Out.
- If you do not create an MPLS QoS policy, any ingressing MPLS traffic is assigned the default QoS level.

You will also configure the underlay and overlay through the SR-MPLS infra L3Out:

- **Underlay:** BGP peer IP (BGP LU and IPv4 peer) configuration as part of the interface configuration.
- **Overlay:** MP-BGP EVPN remote IPv4 address (MP-BGP EVPN peer) configuration as part of the logical node profile configuration.

### Before you begin

- Review the SR-MPLS guidelines and limitations provided in [Guidelines and Limitations for SR-MPLS, on page 18](#), especially the guidelines and limitations provided in [Guidelines and Limitations for the SR-MPLS Infra L3Out, on page 20](#).
- Configure an MPLS custom QoS policy using the procedures provided in [Creating SR-MPLS Custom QoS Policy Using the GUI, on page 36](#)

## Procedure

---

- Step 1** Navigate to **Tenants > infra > Networking > SR-MPLS Infra L3Outs**.
- Step 2** Right-click on **SR-MPLS Infra L3Outs** and choose **Create SR-MPLS Infra L3Out**.  
The **Connectivity** window appears.

Create SR-MPLS Infra L3Out

1. Connectivity 2. Nodes And Interfaces

Connectivity

SR-MPLS Infra Layer3 outside (L3out) is required to configure SR/MPLS handoff from ACI. SR-MPLS Infra L3out configures following important components of ACI to SR handoff:

- BGP EVPN session from ACI Border Leaf (BL) to remote BGP peer - ACI BL can advertise SR/MPLS labels for multiple VRFs using a single BGP session. BGP EVPN session is formed between EVPN loopbacks of BL and DC Provider Edge (DC-PE) router. By doing so removes the need for per VRF sub-interface and routing protocol session from BL to DC-PE. We recommend enabling Multi-hop BFD for faster convergence. Multiple BGP EVPN session can be configured on each BL.
- BGP-LU and IPv4 address family - To provide reachability to router (DC-PE) from BL. BGP-LU between ACI BL and router is used to advertise SR/MPLS labels for transport loopback of the ACI BL and router BGP-IPv4 address family between ACI BL and router is used to provide reachability between BGP-EVPN control plane loopbacks. We recommend enabling Single-hop BFD for faster convergence.
- MPLS custom QoS policy to the ingress and egress SR/MPLS QoS for the SR-MPLS handoff.

Prerequisites:

- Configure the node, port, functional profile, AEP, and Layer 3 domain.

Name:  Layer 3 Domain:

Pod:  MPLS Custom QoS Policy:

BGP-EVPN Connectivity

BFD Multihop Policy:

BGP-EVPN Remote IPv4 Address  Remote ASN  TTL

Previous Cancel Next

**Step 3** In the **Connectivity** window, enter the necessary information.

a) In the **Name** field, enter a name for the SR-MPLS Infra L3Out.

This will be the name for the policy controlling connectivity to the outside. The name can be up to 64 alphanumeric characters.

**Note** You cannot change this name after the object has been saved.

b) In the **Layer 3 Domain** field, choose an existing Layer 3 domain or choose **Create L3 Domain** to create a new layer 3 domain.

c) In the **Pod** field, choose a pod, if you have a Multi-Pod configuration.

If you do not have a Multi-Pod configuration, leave the selection at pod 1.

d) (Optional) In the **MPLS Custom QoS Policy** field, choose an existing QoS policy or choose **Create MPLS Custom QoS Policy** to create a new QoS policy.

For more information on creating a new QoS policy, see [Creating SR-MPLS Custom QoS Policy Using the GUI, on page 36](#).

If you do not create a custom QoS policy, the following default values are assigned:

- All incoming MPLS traffic on the border leaf switch is classified into QoS Level 3 (the default QoS level).
- The border leaf switch does the following:
  - Retains the original DSCP values for traffic coming from SR-MPLS without any remarking.

- Forwards packets to the MPLS network with the original COS value of the tenant traffic if the COS preservation is enabled.
  - Forwards packets with the default MPLS EXP value (0) to the SR-MPLS network.
- In addition, the border leaf switch does not change the original DSCP values of the tenant traffic coming from the application server while forwarding to the SR network.
- e) Navigate to the **BGP-EVPN Connectivity** area.
- f) (Optional) In the **BFD Multihop Policy** field, choose an existing BFD multihop policy or choose **Create BFD Multihop Node Policy** to create a new policy.
- If you have a MP-BGP EVPN multihop session from the border leaf switch to the DC-PE, when you enable the **BFD Multihop Policy** option, the BGP session will not depend on the regular BGP timer; instead, it will get terminated faster, based on the BFD timers. See [Multi-Hop BFD for BGP EVPN Session, on page 8](#) for more information.
- g) In the **BGP-EVPN Remote IPv4 Address** field, enter the MP-BGP EVPN remote IPv4 address.
- This BGP peer IP address is part of the overlay configuration. This is the loopback address of the DC-PE (one entry per remote DC-PE).
- h) In the **Remote ASN** field, enter a number that uniquely identifies the neighbor autonomous system of the DC-PE. The Autonomous System Number can be in 4-byte as plain format from 1 to 4294967295.
- Note** ACI does not support asdot or asdot+ format AS numbers. For more information on asdot or asdot+ format AS numbers, see the [Explaining 4-Byte Autonomous System \(AS\) ASPLAIN and ASDOT Notation for Cisco IOS](#) document.
- i) In the **TTL** field, enter the connection time to live (TTL).
- The range is from 2 to 255 hops.
- j) Click **Next**.
- The **Nodes and Interfaces** window appears.

Create SR-MPLS Infra L3Out

1. Connectivity 2. Nodes And Interfaces

Nodes and Interfaces

Select the Border leaf (BL) switches for the SR-MPLS configuration. Configure BGP EVPN control plane loopback, router id and transport loopback for each BL. Multiple interface can be configured for each BL, and for each interface of BL, BGP labeled unicast (BGP-LU) peer is configured. BGP IPv4 address family is automatically enabled once BGP-LU peer is configured. Single hop BFD can be enabled for each BGP-LU and IPv4 address family session.

Node Profile Name:

Interface Profile Name:

BFD Interface Policy:

Transport Data Plane:

Interface Types

Layer 3:

Layer 2:

Nodes

Node ID	Router ID	BGP-EVPN Loopback	MPLS Transport Loopback	Segment ID (SID) Index	+ Hide Interfaces	
<input type="text" value="select an option"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>		
Interface	VLAN Encap	MTU (bytes)	IPv4 Address	Peer IPv4 Address	Remote ASN	+
<input type="text" value="Select a port"/>	<input type="text"/>	9000	<input type="text"/>	<input type="text"/>	<input type="text"/>	
			BGP-Label Unicast Source address/mask	BGP-Label Unicast address	BGP-Label Unicast	

Previous Cancel Finish

**Step 4** In the **Nodes and Interfaces** window, enter the necessary information to configure the border leaf nodes and interfaces.

- a) In the **Node Profile Name** and **Interface Profile Name** fields, determine if you want to use the default naming convention for the node profile and interface profile names.

The default node profile name is *L3Out-name\_nodeProfile*, and the default interface profile name is *L3Out-name\_interfaceProfile*, where *L3Out-name* is the name that you entered in the **Name** field in the **Connectivity** page. Change the profile names in these fields, if necessary.

- b) (Optional) In the **BFD Interface Policy** field, choose an existing BFD interface policy or choose **Create BFD Interface Policy** to create a new BFD interface policy.
- c) In the **Transport Data Plane** field, determine the type of routing that you would like to use for the handoff on the Cisco ACI border leaf switches.

The options are:

- **MPLS**: Select this option to use Multiprotocol Label Switching (MPLS) for the handoff towards the transport device.
- **SR-MPLS**: Select this option to use segment routing (SR) Multiprotocol Label Switching (MPLS) for the handoff towards the transport device.

- d) In the **Interface Types** area, make the necessary selections in the Layer 3 and Layer 2 fields.

The options are:

- Layer 3:

- **Interface:** Choose this option to configure a Layer 3 interface to connect the border leaf switch to the external router.

When choosing this option, the Layer 3 interface can be either a physical port or a direct port-channel, depending on the specific option selected in the **Layer 2** field in this page.

- **Sub-Interface:** Choose this option to configure a Layer 3 sub-interface to connect the border leaf switch to the external router.

When choosing this option, a Layer 3 sub-interface is created for either a physical port or a direct port-channel, depending on the specific option selected in the **Layer 2** field in this page.

- Layer 2:
  - **Port**
  - **Direct Port Channel**

- e) From the **Node ID** field drop-down menu, choose the border leaf switch, or node, for the L3Out.

For multi-pod configurations, only the leaf switches (nodes) that are part of the pod that you selected in the previous screen are displayed.

You might see a warning message appear on your screen, describing how to configure the router ID.

- If you do not have a router ID already configured for this node, go to [4.f, on page 31](#) for instructions on configuring a router ID for this node.
- If you have a router ID already configured for this node (for example, if you had configured MP-BGP route reflectors previously), you have several options:
  - **Use the same router ID for the SR-MPLS configuration:** This is the recommended option. Make a note of the router ID displayed in this warning to use in the next step in this case, and go to [4.f, on page 31](#) for instructions on configuring a router ID for this node.
  - **Use a different router ID for the SR-MPLS configuration:** In this situation, you must first take the node out of the active path to avoid traffic disruption to the existing application before entering the router ID in the next step. To take the node out of the active path:
    1. Put the node in maintenance mode.
    2. Enter the different router ID for the SR-MPLS configuration, as described in [4.f, on page 31](#).
    3. Reload the node.

- f) In the **Router ID** field, enter a unique router ID (the IPv4 or IPv6 address) for the border leaf switch part of the infra L3Out.

The router ID must be unique across all border leaf switches and the DC-PE.

As described in [4.e, on page 31](#), if a router ID has already been configured on this node, you have several options:

- If you want to use the same router ID for the SR-MPLS configuration, enter the router ID that was displayed in the warning message in [4.e, on page 31](#).
- If you do not want to use the same router ID for the SR-MPLS configuration, or if you did not have a router ID already configured, enter an IP address (IPv4 or IPv6) in this field for the border leaf switch part of the infra L3Out, keeping in mind that it has to be a unique router ID.

Once you have settled on an entry for the **Router ID**, the entries in the **BGP-EVPN Loopback** and **MPLS Transport Loopback** fields are automatically populated with the entry that you provided in the **Router ID** field.

- g) (Optional) Enter an IP address in the **BGP-EVPN Loopback** field, if necessary.

For BGP-EVPN sessions, the BGP-EVPN loopback is used for the control plane session. Use this field to configure the MP-BGP EVPN session between the EVPN loopbacks of the border leaf switch and the DC-PE to advertise the overlay prefixes. The MP-BGP EVPN sessions are established between the BP-EVPN loopback and the BGP-EVPN remote peer address (configured in the **BGP-EVPN Remote IPv4 Address** field in the **Connectivity** window).

The **BGP-EVPN Loopback** field is automatically populated with the same entry that you provide in the **Router ID** field. Enter a different IP address for the BGP-EVPN loopback address, if you don't want to use the router ID as the BGP-EVPN loopback address.

Note the following:

- For BGP-EVPN sessions, we recommend that you use a different IP address in the **BGP-EVPN Loopback** field from the IP address that you entered in the **Router ID** field.
- While you can use a different IP address for the BGP-EVPN loopback and the MPLS transport loopback, we recommend that you use the same loopback for the BGP-EVPN and the MPLS transport loopback on the ACI border leaf switch.

- h) In the **MPLS Transport Loopback** field, enter the address for the MPLS transport loopback.

The MPLS transport loopback is used to build the data plane session between the ACI border leaf switch and the DC-PE, where the MPLS transport loopback becomes the next-hop for the prefixes advertised from the border leaf switches to the DC-PE routers. See [MP-BGP EVPN Session Between the Cisco ACI Border Leaf Switch and the DC-PE, on page 7](#) for more information.

Note the following:

- For BGP-EVPN sessions, we recommend that you use a different IP address in the **MPLS Transport Loopback** field from the IP address that you entered in the **Router ID** field.
- While you can use a different IP address for the BGP-EVPN loopback and the MPLS transport loopback, we recommend that you use the same loopback for the BGP-EVPN and the MPLS transport loopback on the ACI border leaf switch.

- i) In the **Segment ID (SID) Index** field, enter the SID index.

The SID index is configured on each node for the MPLS transport loopback. The SID index value is advertised using BGP-LU to the peer router, and the peer router uses the SID index to calculate the local label.

The supported value for the SID index entry is between 0 and 4294967295. The SID index has to be unique across the entire segment routing domain.

- j) If you selected **Port** in the Layer 2 area above, the **Interface** field appears. Choose a port from the drop-down list.
- k) If you selected **Direct Port Channel** in the Layer 2 area above, the **PC Paths** field appears. Choose the port channel from the drop-down list. This is the path to the port channel end points for the interface profile.
- l) If you selected **Sub-Interface** in the Layer 3 area above, the **VLAN Encap** field appears. Enter the encapsulation used for the layer 3 outside profile.
- m) In the **MTU (bytes)** field, enter the maximum transmit unit of the external network.
- Acceptable entries in this field are from 576-9216. To inherit the value, enter **inherit** in this field.
- n) In the **IPv4 Address** field, enter an IP address for the BGP-Label unicast source.



This is the IP address assigned to the Layer 3 interface/sub-interface/port channel that you configured in a previous step.

- o) In the **Peer IPv4 Address** field, enter the BGP-Label unicast peer IP address.

This is the interface's IP address of the router directly connected to the border leaf switch.

- p) In the **Remote ASN** field, enter the BGP-Label Autonomous System Number of the directly-connected router.

- q) Determine if you want to configure additional interfaces for this node for the SR-MPLS infra L3Out.

- If you do not want to configure additional interfaces for this node for this SR-MPLS infra L3Out, skip to [4.s, on page 33](#).
- If you want to configure additional interfaces for this node for this SR-MPLS infra L3Out, click + in the **Interfaces** area to bring up the same options for another interface for this node.

**Note** If you want to delete the information that you entered for an interface for this node, or if you want to delete an interface row that you added by accident, click the trash can icon for the interface row that you want to delete.

- r) Determine if you want to configure additional nodes for this SR-MPLS infra L3Out.

- If you do not want to configure additional nodes for this SR-MPLS infra L3Out, skip to [4.s, on page 33](#).
- If you want to configure additional nodes for this SR-MPLS infra L3Out, click + in the **Nodes** area to bring up the same options for another node.

**Note** If you want to delete the information that you entered for a node, or if you want to delete a node row that you added by accident, click the trash can icon for the node row that you want to delete.

- s) When you have entered the remaining additional information in the **Nodes and Interfaces** window, click **Finish** to complete the necessary configurations in the **Create SR-MPLS Infra L3Out** wizard.

---

### What to do next

Configure an SR-MPLS VRF L3Out using the procedures provided in [Configuring an SR-MPLS VRF L3Out Using the GUI, on page 33](#).

## Configuring an SR-MPLS VRF L3Out Using the GUI

Using the procedures in this section, you will configure a SR-MPLS VRF L3Out, which will be used to forward traffic from the SR-MPLS infra L3Out that you configured in the previous set of procedures.

- User tenant VRFs are mapped to the SR-MPLS infra L3Outs to advertise tenant bridge domain subnets to the DC-PE routers and import the MPLS VPN routes received from the DC-PE.
- You must specify routing and security policies in the SR-MPLS VRF L3Out for each VRF. These policies point to one or more SR-MPLS infra L3Outs.
- One SR-MPLS VRF L3Out is supported for each VRF.

**Before you begin**

- Review the SR-MPLS guidelines and limitations provided in [Guidelines and Limitations for SR-MPLS, on page 18](#), especially the guidelines and limitations provided in [Guidelines and Limitations for the SR-MPLS VRF L3Out, on page 20](#).
- Configure an SR-MPLS infra L3Out using the procedures provided in [Configuring an SR-MPLS Infra L3Out Using the GUI, on page 26](#).

**Procedure**

**Step 1** Configure the SR-MPLS VRF L3Out by navigating to the **Create SR-MPLS VRF L3Out** window for the tenant (**Tenants** > *tenant* > **Networking** > **SR-MPLS VRF L3Outs**).

**Step 2** Right-click on **SR-MPLS VRF L3Outs** and select **Create SR-MPLS VRF L3Out**.

The **Create SR-MPLS VRF L3Out** window appears.

**Figure 6: Create SR-MPLS VRF L3Out**

The screenshot shows the 'Create SR-MPLS VRF L3Out' configuration window. The window title is 'Create SR-MPLS VRF L3Out' with a help icon and a close icon. The form includes the following fields and sections:

- Name:** A text input field with a red error indicator.
- VRF:** A dropdown menu with 'select an option' and a red error indicator.
- SR-MPLS Infra L3Out:** A dropdown menu with 'select an option' and a red error indicator.
- External EPGs:** A section containing an 'External EPG Name' input field with a red error indicator, a minus icon, a plus icon, and a 'Hide Subnets and Contracts' button.
- Subnets and Contracts:** A section containing:
  - IP Prefix:** An input field with 'address/mask' below it.
  - Inter VRF Policy:** Two checkboxes: 'Route Leaking' and 'Security'.
  - A trash icon and a plus icon.
  - Provided Contract:** A dropdown menu with 'select a value'.
  - Consumed Contract:** A dropdown menu with 'select a value'.
- Route Maps:** A section containing:
  - Outbound:** A dropdown menu with 'select an option' and a red error indicator.
  - Inbound:** A dropdown menu with 'select an option'.

At the bottom right of the window are 'Cancel' and 'Submit' buttons.

**Step 3** In the **Name** field, enter a name for the SR-MPLS VRF L3Out.

This will be the name for the policy controlling connectivity to the outside. The name can be up to 64 alphanumeric characters.

**Note** You cannot change this name after the object has been saved.

**Step 4** In the **VRF** field, select an existing VRF or click **Create VRF** to create a new VRF.

**Step 5** In the **SR-MPLS Infra L3Out** field, select an existing SR-MPLS infra L3Out or click **Create SR-MPLS Infra L3Out** to create a new SR-MPLS infra L3Out.

For more information on creating an SR-MPLS infra L3Out, see [Configuring an SR-MPLS Infra L3Out Using the GUI, on page 26](#).

**Step 6** Navigate to the **External EPGs** area and, in the **External EPG Name** area, enter a unique name for the external EPG to be used for this SR-MPLS VRF L3Out.

**Step 7** Navigate to the **Subnets and Contracts** area and configure individual subnets within this EPG.

**Note** If you want to configure the subnet fields but you do not see the following fields, click **Show Subnets and Contracts** to display the following fields.

- a) In the **IP Prefix** field, enter an IP address and netmask for the subnet.
- b) In the **Inter VRF Policy** field, determine if you want configure inter-VRF policies.
  - If you do not want to configure inter-VRF policies, skip to [7.c, on page 35](#).
  - If you want to configure inter-VRF policies, select the appropriate inter-VRF policy that you want to use.

The options are:

- **Route Leaking.**

If you select **Route Leaking**, the **Aggregate** field appears. Click the box next to **Aggregate** if you also want to enable this option.

- **Security.**

Note that you can select one of the two options listed above or both options for the **Inter VRF Policy** field.

- c) In the **Provided Contract** field, select an existing provider contract or click **Create Contract** to create a provider contract.
- d) In the **Consumed Contract** field, select an existing consumer contract or click **Create Contract** to create a consumer contract.
- e) Determine if you want to configure additional subnets for this external EPG.
  - If you do not want to configure additional subnets for this external EPG, skip to [Step 8, on page 35](#).
  - If you want to configure additional subnets for this external EPG, click + in the **Subnet and Contracts** area to bring up the same options for another subnet.

**Note** If you want to delete the information that you entered for a subnet, or if you want to delete a subnet row that you added by accident, click the trash can icon for the subnet row that you want to delete.

**Step 8** Determine if you want to create additional external EPGs to be used for this SR-MPLS VRF L3Out.

- If you do not want to configure additional external EPGs to be used for this SR-MPLS VRF L3Out, skip to [Step 9, on page 36](#).
- If you want to configure additional external EPGs to be used for this SR-MPLS VRF L3Out, click + in the **External EPG Name** area to bring up the same options for another external EPG.

**Note** If you want to delete the information that you entered for an external EPG, or if you want to delete an external EPG area that you added by accident, click the trash can icon for the external EPG area that you want to delete.

**Step 9** In the **Route Maps** area, configure the outbound and inbound route maps.

Within each SR-MPLS VRF L3Out:

- Defining the outbound route map (export routing policy) is mandatory. This is needed to be able to advertise prefixes toward the external DC-PE routers.
  - Defining the inbound route map (import routing policy) is optional, because, by default, all the prefixes received from the DC-PE routers are allowed into the fabric.
- In the **Outbound** field, select an existing export route map or click **Create Route Maps for Route Control** to create a new export route map.
  - In the **Inbound** field, select an existing import route map or click **Create Route Maps for Route Control** to create a new import route map.

**Step 10** When you have completed the configurations in the **Create SR-MPLS VRF L3Out** window, click **Submit**.

## Creating SR-MPLS Custom QoS Policy Using the GUI

SR-MPLS Custom QoS policy defines the priority of the packets coming from an SR-MPLS network while they are inside the ACI fabric based on the incoming MPLS EXP values defined in the MPLS QoS ingress policy. It also marks the CoS and MPLS EXP values of the packets leaving the ACI fabric through an MPLS interface based on IPv4 DSCP values defined in MPLS QoS egress policy.

If no custom ingress policy is defined, the default QoS Level (`Level3`) is assigned to packets inside the fabric. If no custom egress policy is defined, the default EXP value of 0 will be marked on packets leaving the fabric.

### Procedure

- Step 1** From the top menu bar, navigate to **Tenants > infra**.
- Step 2** In the left pane, select **infra > Policies > Protocol > MPLS Custom QoS**.
- Step 3** Right click the **MPLS Custom QoS** folder and choose **Create MPLS Custom QoS Policy**.
- Step 4** In the **Create MPLS Custom QoS Policy** window that opens, provide the name and description of the policy you're creating.

Create MPLS Custom QoS Policy

Name:

Description: optional

MPLS IngressRule:

Priority	EXP Range From	EXP Range To	Target DSCP	Target CoS

MPLS EgressRule:

DSCP Range From	DSCP Range To	Target EXP	Target CoS

Cancel Submit

**Step 5** In the **MPLS Ingress Rule** area, click + to add an ingress QoS translation rule.

Any traffic coming into the border leaf (BL) connected to the MPLS network will be checked for the MPLS EXP value and if a match is found, the traffic is classified into an ACI QoS Level and marked with appropriate CoS and DSCP values.

Create MPLS Custom QoS Policy

Name: mpls-qos1

Description: optional

MPLS IngressRule:

Priority	EXP Range From	EXP Range To	Target DSCP	Target CoS
Unspecified	Unspecified	Unspecified	Unspecified	Unspecified

Update Cancel

a) In the **Priority** field, select the priority for the ingress rule.

This is the QoS Level you want to assign for the traffic within ACI fabric, which ACI uses to prioritize the traffic within the fabric.. The options range from Level1 to Level6. The default value is Level13. If you do not make a selection in this field, the traffic will automatically be assigned a Level13 priority.

b) In the **EXP Range From** and **EXP Range To** fields, specify the EXP range of the ingressing MPLS packet you want to match.

c) In the **Target DSCP** field, select the DSCP value to assign to the packet when it's inside the ACI fabric.

The DSCP value specified is set in the original traffic received from the external network, so it will be re-exposed only when the traffic is VXLAN decapsulated on the destination ACI leaf node.

The default is `Unspecified`, which means that the original DSCP value of the packet will be retained.

- d) In the **Target CoS** field, select the CoS value to assign to the packet when it's inside the ACI fabric.

The CoS value specified is set in the original traffic received from the external network, so it will be re-exposed only when the traffic is VXLAN decapsulated on the destination ACI leaf node.

The default is `Unspecified`, which means that the original CoS value of the packet will be retained, but only if the CoS preservation option is enabled in the fabric.

- e) Click **Update** to save the ingress rule.  
f) Repeat this step for any additional ingress QoS policy rules.

**Step 6** In the **MPLS Egress Rule** area, click + to add an egress QoS translation rule.

When the traffic is leaving the fabric out of the border leaf's MPLS interface, it will be matched based on the DSCP value of the packet and if a match is found, the MPLS EXP and CoS values will be set based on the policy.

- a) Using the **DSCP Range From** and **DSCP Range To** dropdowns, specify the DSCP range of the ACI fabric packet you want to match for assigning the egressing MPLS packet's priority.  
b) From the **Target EXP** dropdown, select the EXP value you want to assign to the egressing MPLS packet.  
c) From the **Target CoS** dropdown, select the CoS value you want to assign to the egressing MPLS packet.  
d) Click **Update** to save the ingress rule.  
e) Repeat this step for any additional egress QoS policy rules.

**Step 7** Click **OK** to complete the creation of the MPLS custom QoS Policy.

## Displaying MPLS Statistics

Following are the MPLS-specific statistics that you can choose to display in the statistics screens described in this topic:

- At the interface level, as described in [Displaying SR-MPLS Statistics for Interfaces, on page 39](#)
- At the VRF level, as described in [Displaying SR-MPLS Statistics for VRFs, on page 40](#)

To display statistics information for all the interfaces and VRFs in your system, navigate to:

**Tenant > infra > Networking > SR-MPLS Infra L3Outs**

The **SR-MPLS Infra L3Outs** panel is displayed, showing all of the SR-MPLS infra L3Outs configured on your system. Remaining at the upper-level **SR-MPLS Infra L3Outs** panel, navigate to the appropriate statistics page, depending on the type of statistics that you want to display:

- Click the **Interface Stats** tab to display a summary of the statistics for all of the MPLS interfaces on your system. Each row in this window displays MPLS statistics information for a specific interface on a specific node.



**Note** The interface statistics shown in the main SR-MPLS infra L3Outs page are for all the SR-MPLS-enabled interfaces only on border leaf switch models with "FX2" or "GX" at the end of the switch name.

To see other levels of MPLS interface statistics information, see [Displaying SR-MPLS Statistics for Interfaces, on page 39](#).

- Click the **VRF Stats** tab to display a summary of the statistics for all of the MPLS VRFs on your system. Each row in this window displays MPLS statistics information for a specific VRF configured on a specific node.

The VRF statistics provided in the SR-MPLS infra L3Out properties page are the individual VRF statistics on the given border leaf switch or remote leaf switch where the provider label of the SR-MPLS infra L3Out is consumed.

To see other levels of MPLS VRF statistics information, see [Displaying SR-MPLS Statistics for VRFs, on page 40](#).

## Displaying SR-MPLS Statistics for Interfaces

Following are the MPLS-specific interface statistics that you can choose to display in the statistics screens described in this topic:

- Mpls Egress Drop Bytes
- Mpls Egress Admit Bytes
- Mpls Egress Drop Packets
- Mpls Egress Admit Packets
- Mpls Ingress Drop Bytes
- Mpls Ingress Admit Bytes
- Mpls Ingress Drop Packets
- Mpls Ingress Admit Packets

To change the type of statistics that are shown on a statistics page, click the checkbox to bring up the **Select Stats** window, then move entries from the left column to the right column to show different statistics, and from the right column to the left column to remove certain statistics from view.

To change the layout of the statistics in this page to show statistics in a table format, click the icon with three horizontal bars and select **Table View**.

- To display detailed aggregate interface statistics for all of the interfaces in the SR-MPLS VRF L3Outs under an SR-MPLS infra L3Out, navigate to that SR-MPLS infra L3Out:

**Tenant > infra > Networking > SR-MPLS Infra L3Outs > SR-MPLS\_infra\_L3Out\_name**

Click the **Stats** tab to display detailed aggregate interface statistics for all of the interfaces in the SR-MPLS VRF L3Outs under that particular SR-MPLS infra L3Out.

- To display statistics for a specific interface on a specific leaf switch, navigate to that interfaces area on the leaf switch:

**Fabric > Inventory > Pod # > leaf\_switch > Interfaces**, then click either **Routed Interfaces** or **Encapsulated Routed Interfaces**.

Click on the specific interface that you want statistic information for, then click the **Stats** tab.

## Displaying SR-MPLS Statistics for VRFs

Following are the MPLS-specific VRF statistics that you can choose to display in the statistics screens described in this topic:

- Mpls Vrf Egress Drop Bytes
- Mpls Vrf Egress Admit Bytes
- Mpls Vrf Egress Drop Packets
- Mpls Vrf Egress Admit Packets
- Mpls Vrf Ingress Drop Bytes
- Mpls Vrf Ingress Admit Bytes
- Mpls Vrf Ingress Drop Packets
- Mpls Vrf Ingress Admit Packets

To change the type of statistics that are shown on a statistics page, click the checkbox to bring up the **Select Stats** window, then move entries from the left column to the right column to show different statistics, and from the right column to the left column to remove certain statistics from view.

To change the layout of the statistics in this page to show statistics in a table format, click the icon with three horizontal bars and select **Table View**.

- To display detailed aggregate VRF statistics for a specific VRF, navigate to that VRF:

**Tenant > tenant\_name > Networking > VRFs > VRF\_name**

Click the **Stats** tab to display the aggregate VRF statistics for this particular VRF. Note that this VRF is being used by one of the SR-MPLS L3Outs, and this SR-MPLS L3Out might have multiple leaf switches, with multiple interfaces for each leaf switch. The statistics shown in this window is an aggregate of all the interfaces in this SR-MPLS L3Out that is being used by this VRF.

- To display VRF statistics for a specific leaf switch, navigate to the VRF contexts for that leaf switch:

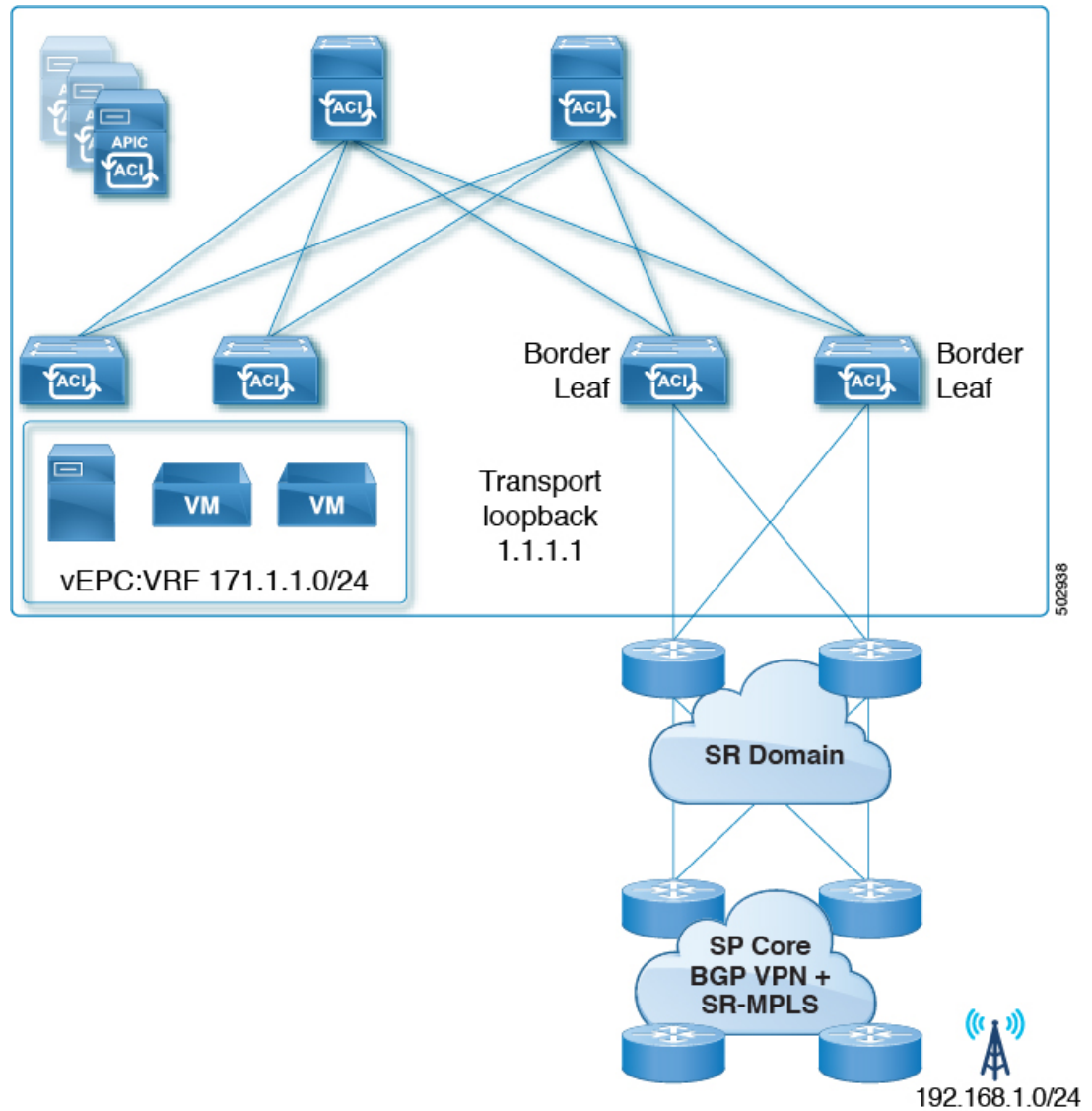
**Fabric > Inventory > Pod # > leaf\_switch > VRF Contexts > VRF\_context\_name**

Click the **Stats** tab to display the statistics for this VRF for this specific leaf switch.

## Configuring SR-MPLS Global Block (GB)

Configure SR-MPLS global block (GB) if you have an SR network between the border leaf switch in the ACI fabric and the DC-PE, as shown in the following figure.





We recommend that all nodes in an SR domain have the same SR-GB configuration.

Following are important guidelines to consider when configuring SR-MPLS global block:

- The allowed configurable SR-GB range is 16000-471804.
- The default SR-GB range in the ACI fabric is 16000-23999.
- ACI always advertises implicit null for the underlay label (transport loopback).

## Procedure

**Step 1** Navigate to the SR-MPLS Global Configurations window.

**Tenants > infra > Policies > Protocol > MPLS Global Configurations**

**Step 2** Access the default **MPLS Global Configurations** screen by double-clicking on **default** in the main SR-MPLS Global Configurations screen or by clicking on **default** in the left nav bar, under **Mpls Global Configurations**.

The default SR-MPLS Global Configurations window appears.

SR-MPLS Global Configurations

Policy History

Properties

Name: default

Description: optional

SR Global Block Minimum: 16001

SR Global Block Maximum: 23999

Show Usage Reset Submit

**Step 3** In the **SR Global Block Minimum** field, enter the minimum value for the SR-GB range.

The lowest allowable value in this field is 16000.

**Step 4** In the **SR Global Block Maximum** field, enter the maximum value for the SR-GB range.

The highest allowable value in this field is 471804.

**Step 5** Click **Submit**.

## Migrating from IP Handoff Configuration to SR Handoff Configuration

### Before you begin:

You have an previously-configured L3Out that is using a pre-Release 5.0(1) IP handoff configuration, as described in [ACI Handoffs Prior to Release 5.0\(1\): IP Handoff, on page 1](#)

### About this task:

These procedures provide instructions for migrating an L3Out that you configured previously with an IP handoff configuration (described in [ACI Handoffs Prior to Release 5.0\(1\): IP Handoff, on page 1](#)) to an SR handoff configuration using the new SR-MPLS components that have been introduced in Cisco APIC Release 5.0(1), as described in [ACI Handoffs in Release 5.0\(1\): SR Handoff, on page 2](#).

For these instructions, it is assumed that the two handoffs are used to connect to the same external network infrastructure, that an external device is able to access the ACI fabric using both L3Outs. The assumption is

that, currently, external clients are able to come in through the L3Outs used in the IP handoff configuration, but once you have completed the procedures in this section, the external clients can then come in through the L3Outs used in the SR-MPLS handoff configuration.



- Note** Throughout these procedures, the following terms are used to distinguish between the two types of L3Outs:
- **IP-based L3Out:** Used for the previously-configured user tenant L3Out that is using a pre-Release 5.0(1) IP handoff configuration.
  - **SR-MPLS L3Out:** Used for the newly-configured user tenant L3Out that has been configured using the new SR-MPLS components that have been introduced in Cisco APIC Release 5.0(1).

Following are the overall steps that you will go through as part of this process:

- Configure the external EPGs on the SR-MPLS VRF L3Out to mirror the IP-based L3Out configuration. This includes the subnets configuration for classification of inbound traffic and the contracts provided or consumed by the external EPGs.
- Redirect inbound and outbound traffic to ensure that it starts preferring the SR-MPLS L3Out.
- Disconnect the IP-based L3Out.

The following sections provide detailed instructions for each of the overall steps listed above.

## Configuring External EPGs on the SR-MPLS VRF L3Out

In this task, you will configure the external EPGs on the SR-MPLS VRF L3Out to mirror the IP-based L3Out configuration (the L3Out that you configured previously that is using a pre-Release 5.0(1) IP handoff configuration). This includes the subnets configuration for classification of inbound traffic and the contracts provided or consumed by the external EPGs.

### Before you begin

Review the information provided in [Migrating from IP Handoff Configuration to SR Handoff Configuration, on page 42](#).

### Procedure

- 
- Step 1** Create a new infra SR-MPLS L3Out, if you have not done so already.  
See [Configuring an SR-MPLS Infra L3Out Using the GUI, on page 26](#) for those instructions, then return here.
- Step 2** Create a new user tenant SR-MPLS L3Out, if you have not done so already.  
See [Configuring an SR-MPLS VRF L3Out Using the GUI, on page 33](#) for those instructions, then return here. Note that this L3Out should be associated to the same VRF of the previously-configured IP-based L3Out.  
As part of the process for creating the new user tenant SR-MPLS L3Out, you will be asked to configure the external EPG for this SR-MPLS L3Out.

- For the external EPG for the new SR-MPLS L3Out, enter the same IP prefix information that you currently have for your previously-configured IP-based L3Out.
- If you have more than one external EPG configured for your previously-configured IP-based L3Out, create additional external EPGs for the new SR-MPLS L3Out and match the same IP prefix information for each EPG.

In the end, the external EPG settings that you configure for the new SR-MPLS L3Out, with the accompanying subnet settings, should match the external EPG and subnet settings that you had previously configured for the IP-based L3Out.

Once you have completed the procedures for creating the new user tenant SR-MPLS L3Out, you should now have two L3Outs (two paths in BGP):

- The existing, previously-configured IP-based L3Out that is using a pre-Release 5.0(1) IP handoff configuration, as mentioned in the **Before you begin** area in [Migrating from IP Handoff Configuration to SR Handoff Configuration, on page 42](#).
- The new SR-MPLS L3Out that you created using the new SR-MPLS components that have been introduced in Cisco APIC Release 5.0(1).

**Step 3** Ensure the same security policy is applied to the external EPGs of the SR-MPLS L3Out as you had for the IP-based L3Out.

In the non-border leaf switches and the border leaf switches, the new security policy in the external EPG that you configured when you created the new SR-MPLS L3Out will result in a fault for every subnet whose prefix clashes with the subnet prefix in any EPG of the previously-configured IP-based L3Out. This is a fault that does not impact functionality, as long as the same security policies are applied to the same external EPGs of both L3Outs.

---

### What to do next

Redirect inbound and outbound traffic to ensure that it starts preferring the SR-MPLS L3Out using the procedures provided in [Redirecting Traffic to SR-MPLS L3Out, on page 44](#).

## Redirecting Traffic to SR-MPLS L3Out

In this task, you will redirect inbound and outbound traffic to ensure that it starts preferring the SR-MPLS L3Out.

### Before you begin

- Review the information provided in [Migrating from IP Handoff Configuration to SR Handoff Configuration, on page 42](#).
- Configure the external EPGs on the SR-MPLS VRF L3Out to mirror the IP-based L3Out configuration using the procedures provided in [Configuring External EPGs on the SR-MPLS VRF L3Out, on page 43](#).

### Procedure

---

**Step 1** Navigate to the BGP Peer Connectivity Profile for the previously-configured IP-based L3Out.

In the Navigation pane, navigate to **Tenants > *tenant\_name\_for\_IP\_handoff\_L3Out* > Networking > L3Outs > *L3Out\_name* > Logical Node Profiles > *logical\_profile\_name* > Logical Interface Profiles > *logical\_interface\_profile\_name* > *BGP\_peer\_connectivity\_profile* .**

**Step 2** Click on the BGP Peer Connectivity Profile in the left nav bar so that the **BGP Peer Connectivity Profile** page is displayed in the right main window.

**Step 3** Scroll down the page until you see the **Route Control Profile** area in the **BGP Peer Connectivity Profile** page.

**Step 4** Determine if route control policies were already configured for the existing IP-based L3Out.

You may or may not have had route control policies configured for the existing IP-based L3Out; however, for the new SR-MPLS L3Out, you will need to have route control policies configured. If you had route control policies configured for the existing IP-based L3Out, you can use those route control policies for the new SR-MPLS L3Out; otherwise, you will have to create new route control policies for the SR-MPLS L3Out.

- If you see two route control profiles displayed in the Route Control Profile table:
  - An export route control policy, shown with Route Export Policy in the Direction column in the table.
  - An import route control policy, shown with Route Import Policy in the Direction column in the table.

then route control policies have already been configured for the IP-based L3Out. Go to [Step 5, on page 46](#).

- If you do not see two route control profiles displayed in the Route Control Profiles table, then create a new route map that will be used for the SR-MPLS L3Out:
  - a) In the **Navigation** pane, expand the **Tenants > *tenant\_name\_for\_IP\_handoff\_L3Out* > Policies > Protocol**.
  - b) Right-click on **Route Maps for Route Control** and select **Create Route Maps for Route Control**.
  - c) In the **Create Route Maps for Route Control** dialog box, in the **Name** field, enter a route profile name.
  - d) In the **Type** field, you must choose **Match Routing Policy Only**.
  - e) In the **Contexts** area, click the + sign to open the **Create Route Control Context** dialog box and perform the following actions:
    1. Populate the **Order** and the **Name** fields as desired.
    2. In the **Match Rule** field, click **Create Match Rule**.
    3. In the **Create Match Rule** dialog box, in the **Name** field, enter a name for the match rule.
    4. Enter the necessary information in the appropriate fields (**Match Regex Community Terms**, **Match Community Terms** and **Match Prefix**), then click **Submit**.
    5. In the **Set Rule** field, click **Create Set Rules for a Route Map**
    6. In the **Create Set Rules for a Route Map** dialog box, in the **Name** field, enter a name for the action rule profile.
    7. Choose the desired attributes, and related community, criteria, tags, and preferences. Click **Finish**.
    8. In the **Create Route Control Context** window, click **OK**.
    9. In the **Create Route Maps for Route Control** dialog box, click **Submit**.
  - f) Navigate to the BGP Peer Connectivity Profile screen:
 

**Tenants > *tenant\_name\_for\_IP\_handoff\_L3Out* > Networking > L3Outs > *L3out-name* > Logical Node Profiles > *logical-node-profile-name* > Logical Interface Profiles > *logical-interface-profile-name* > *BGP\_peer\_connectivity\_profile***

- g) Click on the BGP Peer Connectivity Profile in the left nav bar so that the **BGP Peer Connectivity Profile** page is displayed in the right main window.
- h) Scroll down to the **Route Control Profile** field, then click + to configure the following:
  - **Name:** Select the route-map that you just configured for the route import policy.
  - **Direction:** Select **Route Import Policy** in the Direction field.

Repeat these steps to select the route-map for the route export policy and set the **Route Export Policy** in the Direction field.

**Step 5** Force the BGP to choose the new SR path by configuring the route policies for all the peers in the border leaf switches for the VRF that will be undergoing the migration.

- If the previously-configured IP-based L3Out was configured for **eBGP**, configure both the route import policy and the route export policy for the IP-based L3Out peer to have an additional AS path entry (for example, the same AS as local entry). This is the most typical scenario.

**Note** The following procedures assume you do not have set rules configured already for the route map. If you do have set rules configured already for the route map, edit the existing set rules to add the additional AS path entry (check the **Set AS Path** checkbox and select the criterion **Prepend AS**, then click + to prepend AS numbers).

- a. Navigate to **Tenant > tenant\_name\_for\_IP\_handoff\_L3Out > Policies > Protocol > Set Rules** and right click **Create Set Rules for a Route Map**.

The **Create Set Rules For A Route Map** window appears.

- b. In the **Create Set Rules For A Route Map** dialog box, perform the following tasks:
  1. In the **Name** field, enter a name for these set rules.
  2. Check the **Set AS Path** checkbox, then click **Next**.
  3. In the **AS Path** window, click + to open the **Create Set AS Path** dialog box.
- c. Select the criterion **Prepend AS**, then click + to prepend AS numbers.
- d. Enter the AS number and its order and then click **Update**.
- e. Click **OK**.
- f. In the **Create Set Rules For A Route Map** window, confirm the listed criteria for the set rule based on AS Path and click **Finish**.
- g. Navigate back to the BGP Peer Connectivity Profile screen for this existing IP-based L3Out:
 

**Tenants > tenant\_name\_for\_IP\_handoff\_L3Out > Networking > L3Outs > L3out-name > Logical Node Profiles > logical-node-profile-name > Logical Interface Profiles > logical-interface-profile-name > BGP\_peer\_connectivity\_profile**
- h. Scroll down to the Route Control Profile area and note the route profile names for both the export route control policy and the import route control policy that are being used for this existing IP-based L3Out.
- i. Navigate to **Tenants > tenant\_name\_for\_IP\_handoff\_L3Out > Policies > Protocol > Route Maps for Route Control**.

- j. First locate the **export** route control profile that is being used for this existing IP-based L3Out and click on that route profile.

The properties page for this route control profile appears in the main panel.
  - k. Locate the route control context entry in the page and double-click the route control context entry.

The properties page for this route control context appears.
  - l. In the **Set Rule** area, select the set rule that you created earlier in these procedures with the additional AS path entry, then click **Submit**.
  - m. Now locate the **import** route control profile that is being used for this existing IP-based L3Out and click on that route profile, then repeat these steps to use the set rule with the additional AS path entry for the import route control profile. Doing this will influence inbound traffic, where an external source should start preferring.
- If the previously-configured IP-based L3Out was configured for **iBGP**, due to the fact that SR-MPLS only supports eBGP, you will need to use the local preference setting to steer traffic to an eBGP-configured SR-MPLS L3Out, as described in the previous bullet. Configure both the route import policy and the route export policy for the IP-based L3Out peer to have a lower local preference value:
    - a. Navigate to **Tenant > tenant\_name\_for\_IP\_handoff\_L3Out > Policies > Protocol > Set Rules** and right click **Create Set Rules for a Route Map**.
    - The **Create Set Rules For A Route Map** window appears.
    - b. In the **Name** field, enter a name.
    - c. Check the **Set Preference** checkbox.
    - The Preference field appears.
    - d. Enter the BGP local preference path value.
    - The range is 0-4294967295.
    - e. Click **Finish**.
    - f. Navigate back to the BGP Peer Connectivity Profile screen for this existing IP-based L3Out:  
**Tenants > tenant\_name\_for\_IP\_handoff\_L3Out > Networking > L3Outs > L3out-name > Logical Node Profiles > logical-node-profile-name > Logical Interface Profiles > logical-interface-profile-name > BGP\_peer\_connectivity\_profile**
    - g. Scroll down to the Route Control Profile area and note the route profile names for both the export route control policy and the import route control policy that are being used for this existing IP-based L3Out.
    - h. Navigate to **Tenants > tenant\_name\_for\_IP\_handoff\_L3Out > Policies > Protocol > Route Maps for Route Control**.
    - i. First locate the **export** route control profile that is being used for this existing IP-based L3Out and click on that route profile.
    - The properties page for this route control profile appears in the main panel.
    - j. Locate the route control context entry in the page and double-click the route control context entry.
    - The properties page for this route control context appears.
    - k. In the **Set Rule** area, select the set rule that you created earlier in these procedures with the BGP local preference path, then click **Submit**.

1. Now locate the **import** route control profile that is being used for this existing IP-based L3Out and click on that route profile, then repeat these steps to use the set rule with the BGP local preference path entry for the import route control profile.

**Step 6** Confirm that traffic is now choosing the SR-MPLS path.

The routing/path selection should be through SR-MPLS (BGP should choose the SR-MPLS path over the IP path). You can monitor the traffic and routes in URIB for each VRF to verify that the SR-MPLS path is selected.

---

### What to do next

Disconnect the IP-based L3Out using the procedures provided in [Disconnecting the IP-Based L3Out, on page 48](#).

## Disconnecting the IP-Based L3Out

In this task, you will be disconnecting the IP-based L3Out.

### Before you begin

- Review the information provided in [Migrating from IP Handoff Configuration to SR Handoff Configuration, on page 42](#).
- Configure the external EPGs on the SR-MPLS VRF L3Out to mirror the IP-based L3Out configuration using the procedures provided in [Configuring External EPGs on the SR-MPLS VRF L3Out, on page 43](#).
- Redirect inbound and outbound traffic to ensure that it starts preferring the SR-MPLS L3Out using the procedures provided in [Redirecting Traffic to SR-MPLS L3Out, on page 44](#).

### Procedure

---

**Step 1** Clean up the IP paths.

You can clean up the IP paths using one of the following methods:

- Remove one subnet at a time in the external EPG in the previously-configured IP-based L3Out.
- Remove the external EPGs in the previously-configured IP-based L3Out.

Either of the methods above will result in the fault being cleared, and the external EPG in the SR-MPLS L3Out will now be deployed.

As part of the process of changing the security policy from the IP-based L3Out to the SR-MPLS L3Out, there might be up to a 15-second drop. After that period, the outbound traffic from ACI to outside will take the SR-MPLS path.

If you see that the previously-configured IP-based L3Out was migrated successfully to the new SR-MPLS L3Out, you can then delete the previously-configured IP-based L3Out.

**Step 2** Determine if you have additional L3Outs/VRFs that you want to migrate to SR-MPLS.



Repeat the procedures in [Migrating from IP Handoff Configuration to SR Handoff Configuration, on page 42](#) to migrate other user L3Outs and VRFs to SR-MPLS.

The same procedures in [Migrating from IP Handoff Configuration to SR Handoff Configuration, on page 42](#) can also be used to migrate between a tenant GOLF L3Out and a tenant SR-MPLS L3Out.

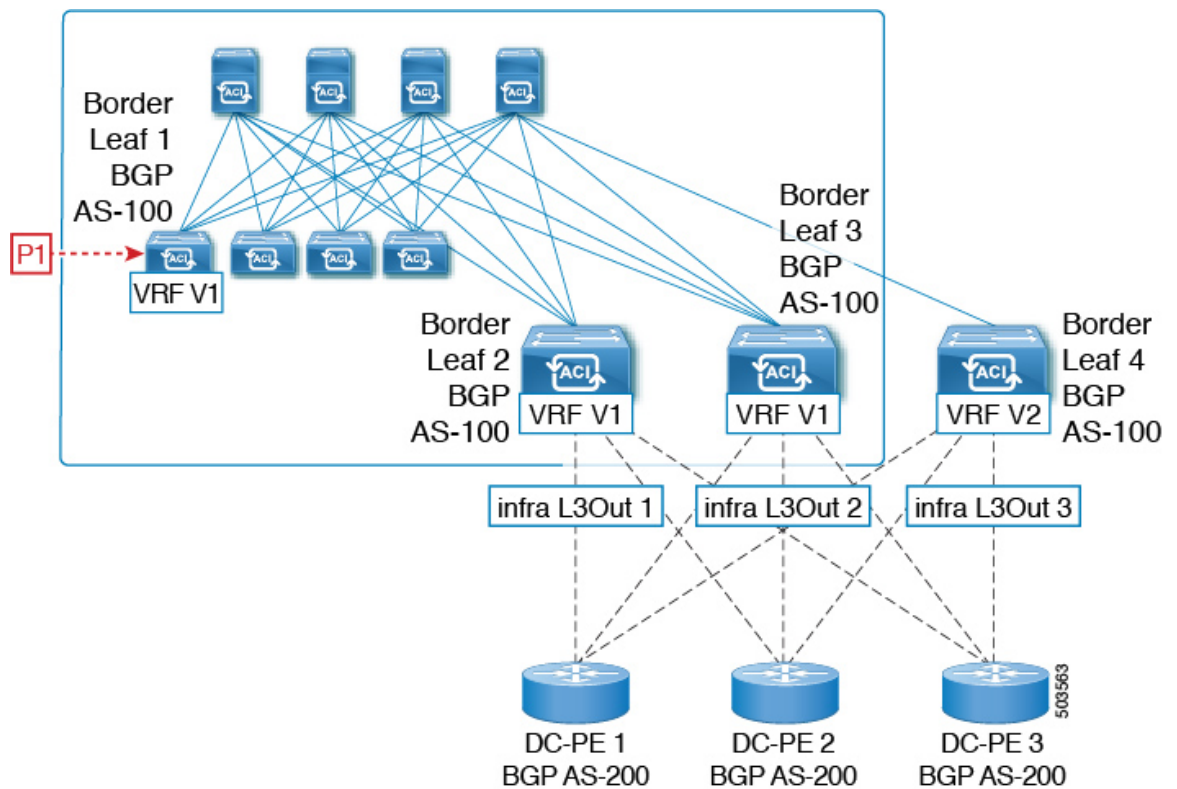
## About the BGP Domain-Path Feature for Loop Prevention

BGP routing loops might occur in certain situations due to various conditions, such as:

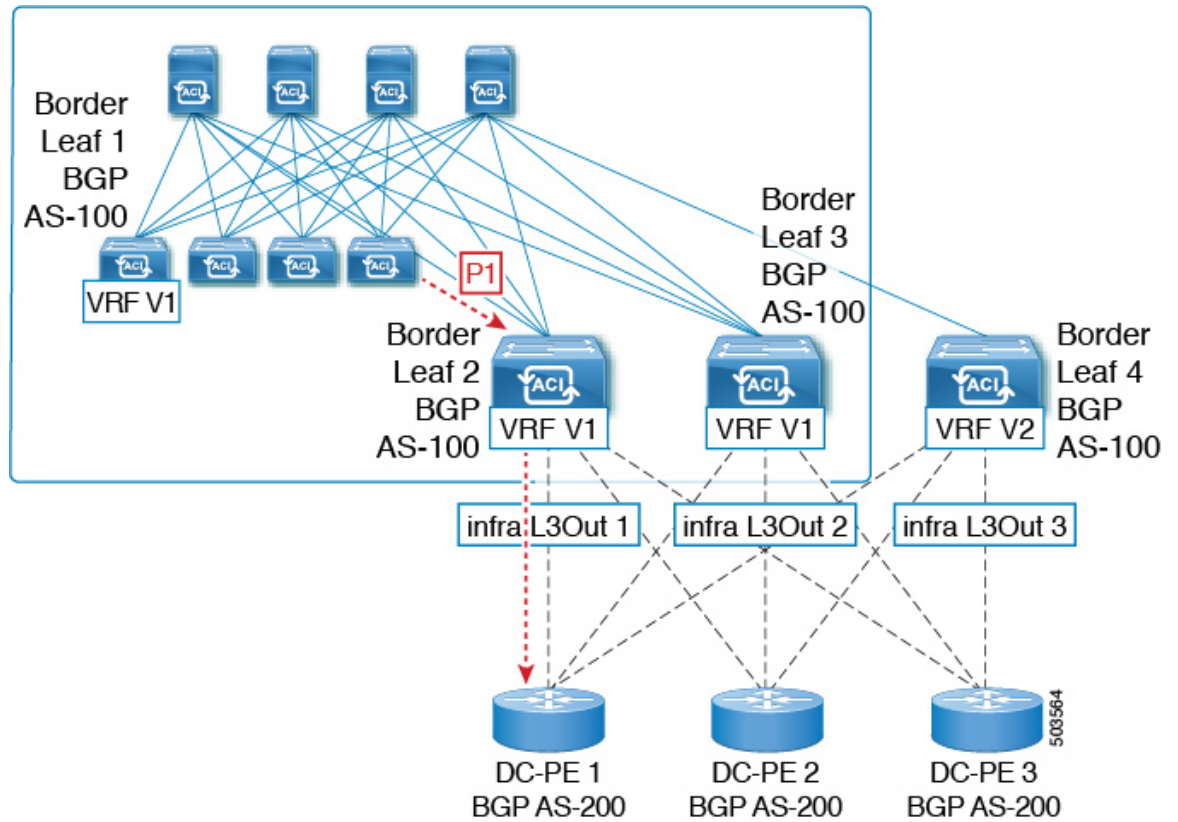
- Intentional disabling of existing BGP loop prevention mechanisms, such as AS Path checks
- Route leaks across different VRFs or VPNs

Following is an example scenario where a BGP routing loop might occur:

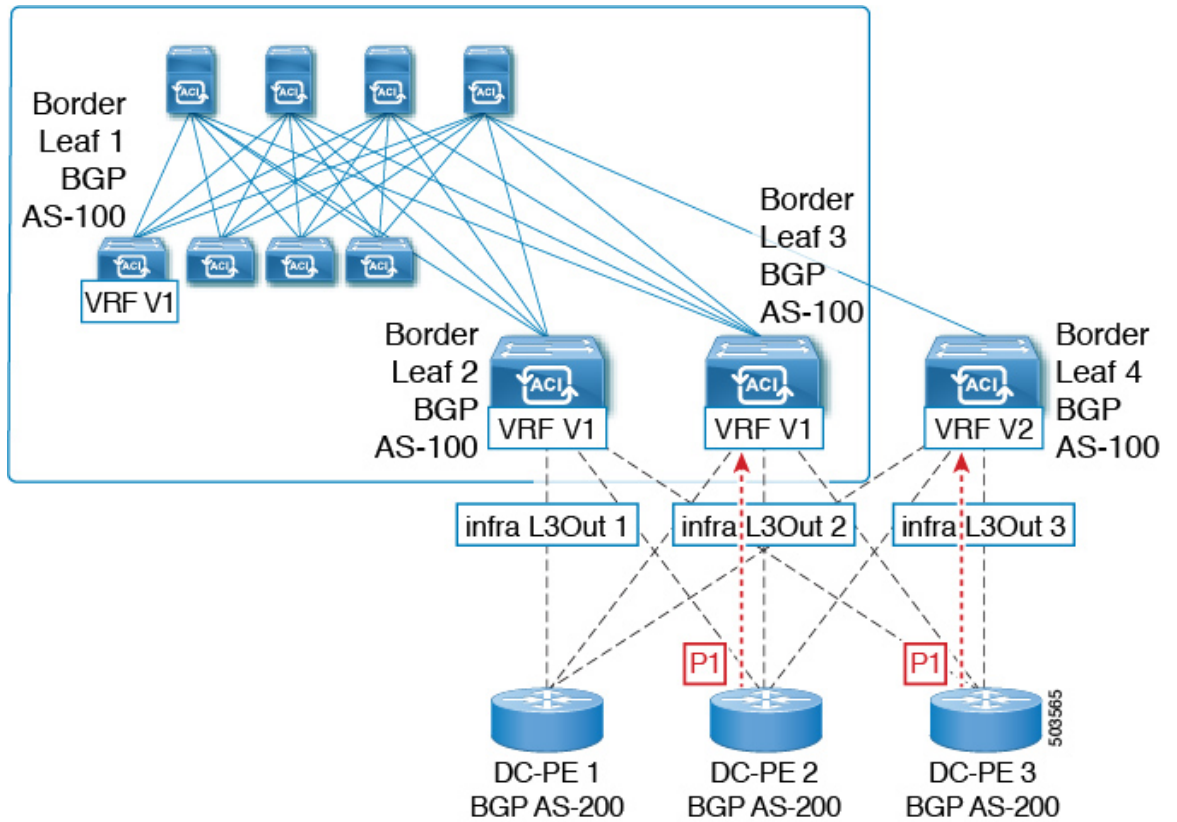
1. A prefix P1 received from a BGP IP L3Out peer is advertised in the ACI fabric using the Multiprotocol Border Gateway Protocol (MP-BGP).



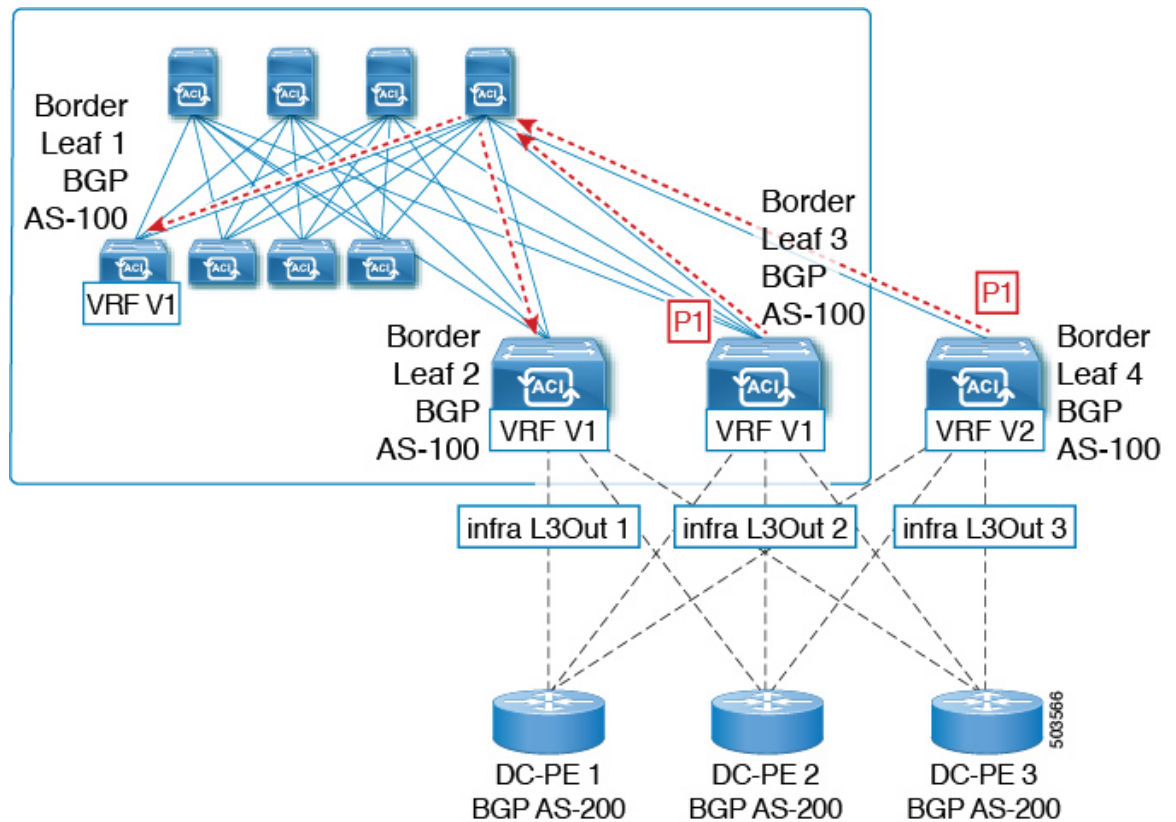
2. As a transit case, this prefix can be advertised out externally through an SR-MPLS infra L3Out.



3. This prefix could then be imported back into the ACI fabric from the core, either in the same VRF or in a different VRF.



- A BGP routing loop would occur when this imported prefix is then advertised back to the originating switch, either from the same VRF or through a leak from a different VRF.



Beginning with Release 5.1(3), the new BGP Domain-Path feature is available, which helps with BGP routing loops in the following ways:

- Keeps track of the distinct routing domains traversed by a route within the same VPN or extended VRFs, as well as across different VPNs or VRFs
- Detects when a route loops back to a VRF in a domain where it has already traversed (typically at a border leaf switch that is the stitching point between domains, but also at an internal switch, in some cases)
- Prevents the route from getting imported or accepted when it would lead to a loop

Within an ACI fabric, the VRF scope is global and is extended to all switches where it is configured. Therefore, a route that is exported out of a domain in a VRF is blocked from being received back into the VRF on any other switch.

The following components are used with the BGP Domain-Path feature for loop prevention:

- **Routing domain ID:** Every tenant VRF in an ACI site is associated with one internal fabric domain, one domain for each VRF in each SR-MPLS infra L3Out, and one domain for each IP L3Out. When the BGP Domain-Path feature is enabled, each of these domains is assigned a unique routing domain ID, in the format *Base*:<variable>, where:
  - *Base* is the non-zero value that was entered in the **Domain ID Base** field in the **BGP Route Reflector Policy** page
  - <variable> is a randomly-generated value specifically for that domain

- **Domain path:** The domain segments traversed by a route are tracked using a BGP domain path attribute:
  - The domain ID of the VRF for the source domain where the route is received is prepended to the domain path
  - The source domain ID is prepended to the domain path while re-originating a route across domains on the border leaf switches
  - An external route is not accepted if any of the local domain IDs for the VRFs is in the domain path
  - The domain path is carried as an optional and transitive BGP path attribute with each domain segment, represented as <Domain-ID:SAFI>
  - The ACI border leaf switches prepend the VRF internal domain ID for both locally originated and external routes to track leaks within the domain
  - A route from the internal domain can be imported and installed in a VRF on a node with a conflicting external domain ID to provide an internal backup or transit path
  - For infra L3Out peers, the advertisement of a route to a peer is skipped if the domain ID of the peer domain is present in the domain path of the route (outbound check is not applicable for IP L3Out peers)
  - The border leaf switches and non-border leaf switches will both process the domain path attribute



---

**Note** You can configure the BGP Domain-Path feature for loop prevention, or simply enable the configuration to send a received domain path, through the GUI or REST API. You cannot configure the BGP Domain-Path feature for loop prevention or enable the configuration to send a received domain path through the NX-OS style CLI.

---



---

**Note** When upgrading to Release 5.1(3) from a previous release, if you have contracts configured for inter-VRF shared services, those contracts might not work as expected with the BGP Domain-Path feature for loop prevention because the BGP domain ID would not have been set in those contracts that were configured before you upgraded to Release 5.1(3). In those situations, delete the contract and then add the contract back, which will allow the BGP domain update to occur. This is only an issue when you have contracts that were configured prior to your upgrade to Release 5.1(3); this is not an issue when you create new contracts after you've completed the upgrade to Release 5.1(3).

---

## Configuring the BGP Domain-Path Feature for Loop Prevention Using the GUI

### Before you begin

Become familiar with the BGP Domain-Path feature using the information provided in [About the BGP Domain-Path Feature for Loop Prevention, on page 49](#).

## Procedure

**Step 1** If you want to use the BGP Domain-Path feature for loop prevention, set the BGP Domain-Path attribute on the BGP route reflector.

**Note** If you do not want to use the BGP Domain-Path feature for loop prevention but you still want to send a received domain path, do not enable the BGP Domain-Path feature on the BGP route reflector in this step. Instead, go directly to [Step 2, on page 54](#) to only enable the **Send Domain Path** field in the appropriate BGP connectivity window.

a) Navigate to **System > System Settings > BGP Route Reflector**.

The **BGP Route Reflector** window appears. Verify that the **Policy** page tab is selected in this window.

b) Locate the **Domain ID Base** field.

c) Enter a number in the **Domain ID Base** field.

- Enter a value between 1-4294967295 to enable the BGP Domain-Path feature. If your ACI fabric is part of a Multi-Site environment, make sure that you use a unique value that will be specific for this ACI fabric in this **Domain ID Base** field.
- To disable the BGP Domain-Path feature, enter 0 in this **Domain ID Base** field.

When the BGP Domain-Path feature for loop prevention is enabled, an implicit routing domain ID of the format `Base:<variable>` will be allocated, where:

- `Base` is the non-zero value that you entered in this Domain ID Base field
- `<variable>` is a randomly-generated value specifically for the VRF or L3Out that will be used for the BGP Domain-Path feature for loop prevention

This routing domain ID is passed to BGP to identify the following domains:

- **VRF**: Identified by an internal domain ID using a randomly-generated value specifically for each VRF, as shown in the **Routing Domain ID** field in the **Policy** tab in the VRF window for that tenant
- **IP L3Out**: Identified by an external domain ID using a randomly-generated value specifically for each IP L3Out, as shown in the **Routing Domain ID** field in the **BGP Peer Connectivity Profile** window for that IP L3Out
- **SR-MPLS infra L3Out**: Identified by an external domain ID using a randomly-generated value specifically for each VRF in each SR-MPLS infra L3Out, as shown in the **Routing Domain ID** column in the **SR-MPLS Infra L3Outs** table in the window for each SR-MPLS VRFL3Out

The Domain-Path attribute is processed on the inbound directions to check for loops based on the routing domain IDs in the path. The Domain-Path attribute is sent to a peer, which is controlled separately through the BGP peer-level **Send Domain Path** field in the IP L3Out or in the SR-MPLS infraL3Out, as described in the next step.

**Step 2** To send the BGP domain path attribute to a peer, enable the **Send Domain Path** field in the appropriate BGP connectivity window.

If you want to use the BGP Domain-Path feature for loop prevention, first set the **Domain Base ID** in [Step 1, on page 54](#), then enable the **Send Domain Path** field here. If you do not want to use the BGP Domain-Path feature for loop prevention but you still want to send a received domain path, only enable the **Send Domain Path** field here (do not set the **Domain Base ID** in [Step 1, on page 54](#) in that case).

- To enable the **Send Domain Path** field for a IP L3Out peer:
  - a. Navigate to the **BGP Peer Connectivity Profile** window for the IP L3Out peer:  
**Tenant > tenant\_name > Networking > L3Outs > L3Out\_name > Logical Node Profile > log\_node\_prof\_name > Logical Interface Profile > log\_int\_prof\_name > BGP Peer <address>-Node-<node\_ID>**  
 The **BGP Peer Connectivity Profile** window for this configured L3Out appears.
  - b. Locate the **BGP Controls** area in the **BGP Peer Connectivity Profile** window.
  - c. In the **BGP Controls** area, click the box next to the **Send Domain Path** field.
  - d. Click **Submit**.  
 This action sends the BGP domain path attribute to a peer.
- To enable the **Send Domain Path** field for a SR-MPLS infra L3Out peer:
  - a. Navigate to **Tenant > infra > Networking > SR-MPLS Infra L3Outs > SR-MPLS-infra-L3Out\_name > Logical Node Profiles > log\_node\_prof\_name**.  
 The **Logical Node Profile** window for this configured SR-MPLS infra L3Out appears.
  - b. Locate the **BGP-EVPN Connectivity Profile** area, then determine if you want to create a new BGP-EVPN connectivity policy or if you want to enable the **Send Domain Path** field in an existing BGP-EVPN connectivity policy.
    - If you want to create a new a create a new BGP-EVPN connectivity policy, click + above the table in the **BGP-EVPN Connectivity Profile** area. The **Create BGP-EVPN Connectivity Policy** window appears.
    - If you want to enable the **Send Domain Path** field in an existing BGP-EVPN connectivity policy, double-click on that policy in the table in the **BGP-EVPN Connectivity Profile** area. The **BGP-EVPN Connectivity Policy** window appears.
  - c. Locate the **BGP Controls** area in the window.
  - d. In the **BGP Controls** area, click the box next to the **Send Domain Path** field.
  - e. Click **Submit**.  
 This action sends the BGP domain path attribute to a peer.

**Step 3** Navigate to the appropriate areas to see the routing IDs assigned to the various domains.

- To see the routing ID assigned to the VRF domain, navigate to:  
**Tenants > tenant\_name > Networking > VRFs > VRF\_name**, then click on the **Policy** tab for that VRF and locate the entry in the **Routing Domain ID** field in the **VRF** window.
- To see the routing ID assigned to the IP L3Out domain, navigate to:  
**Tenants > tenant\_name > Networking > L3Outs > L3Out\_name > Logical Node Profiles > log\_node\_prof\_name > BGP Peer**, then locate the entry in the **Routing Domain ID** field in the **BGP Peer Connectivity Profile** window.
- To see the routing ID assigned to the SR-MPLS infra L3Out domain, navigate to:

**Tenants** > *tenant\_name* > **Networking** > **SR-MPLS VRF L3Outs** > *SR-MPLS\_VRF\_L3Out\_name*, then locate the entry in the **Routing Domain ID** column in the **SR-MPLS Infra L3Outs** table in the window for that SR-MPLS VRF L3Out.

---