

Multi-Site and SR-MPLS L3Out Handoff

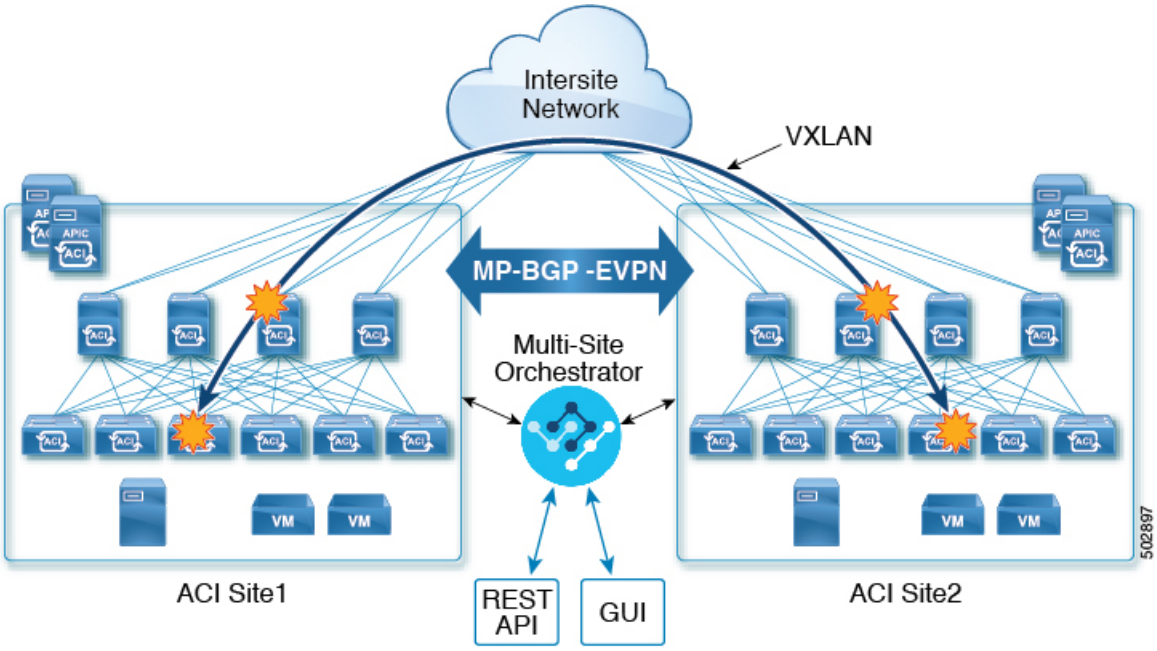
- Overview and Use Cases, on page 1
- SR-MPLS Infra Requirements and Guidelines, on page 5
- SR-MPLS Tenant Requirements and Guidelines, on page 7
- Creating Custom QoS Policy for SR-MPLS, on page 9
- Creating SR-MPLS Infra L3Out, on page 11
- Creating SR-MPLS Route Map Policy, on page 14
- Configure EPG-to-External-EPG (North-South) Communication, on page 16

Overview and Use Cases

Starting with Nexus Dashboard Orchestrator release 3.0(1) and APIC Release 5.0(1), the Multi-Site architecture provides better hand-off functionality between ACI border leaf (BL) switches and SR-MPLS networks.

In a typical Multi-Site deployment, traffic between sites is forwarded over an intersite network (ISN) via VXLAN encapsulation:

Figure 1: Multi-Site and ISN

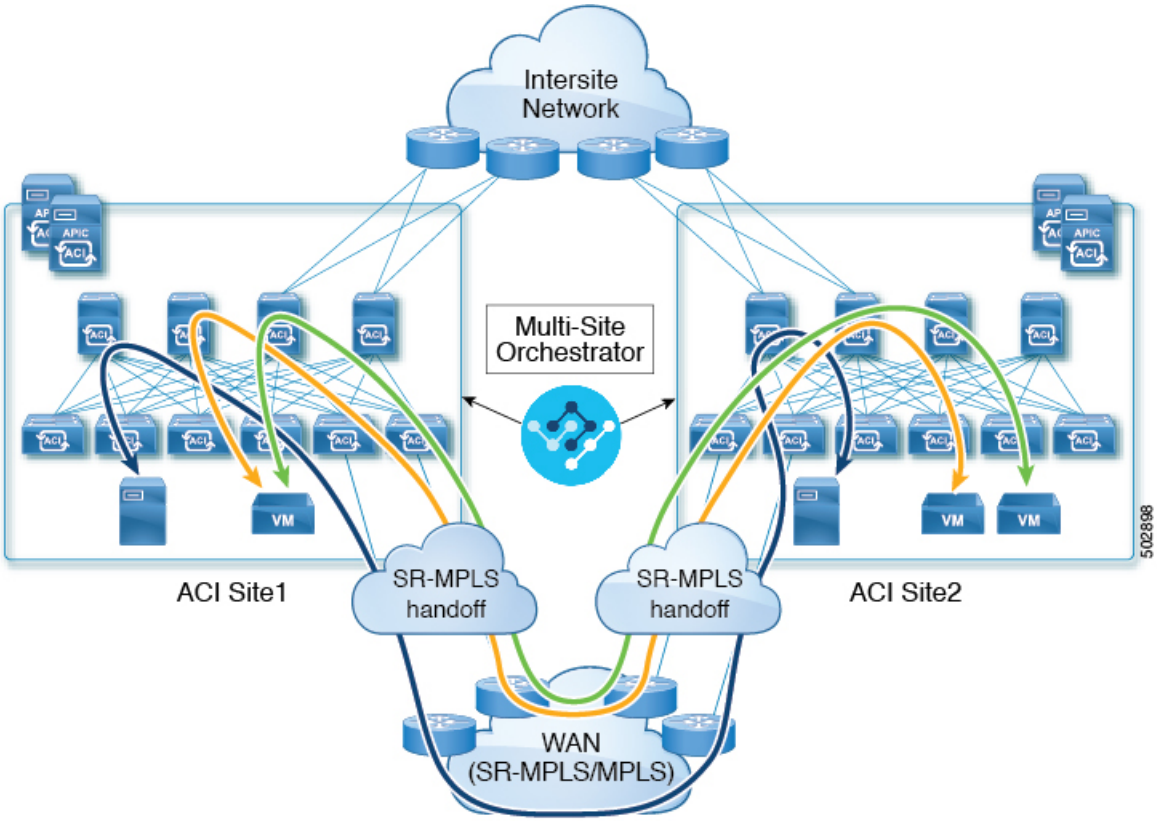


With Release 3.0(1), MPLS network can be used in addition to or instead of the ISN allowing inter-site communication via WAN, as shown in the following figure. In order to force East-West Layer 3 communication to follow the SR-MPLS L3Out data path (instead of the VXLAN data path across the ISN), several restrictions had to be applied to this SR-MPLS hand-off use case:

- The VRF to which the SR-MPLS L3Out belongs must not be stretched across sites.
- Because of the above restriction, every site must deploy one (or more) local SR-MPLS L3Outs for each defined site-local VRF.
- Contracts must not be applied between site-local EPGs belonging to different VRFs.

  This forces the communication to follow the SR-MPLS L3Out data path.

Figure 2: Multi-Site and MPLS



Additional Use Cases in Release 4.0(2) and Later

Prior to release 4.0(2), if you wanted to deploy the SR-MPLS use case, you would define a special "SR-MPLS" template that could be associated with only a single site and not stretched across multiple sites. In this case, if you had two sites managed by your Nexus Dashboard Orchestrator and connected via an SR-MPLS network, and you wanted to establish communication between an EPG in `site1` and another EPG in `site2`, you had to deploy two separate SR-MPLS L3Outs (one in each site) associated with two separate VRFs and establish contracts between the EPG in each site and that site's SR-MPLS L3Out (instead of directly between the EPGs). In other words, the EPGs' traffic would always use the SR-MPLS L3Out data path even for EPG-to-EPG communication across sites without integrating with the traditional Multi-Site data plane for East-West traffic.

Beginning with release 4.0(2), the SR-MPLS L3Outs can function similar to the traditional IP-based L3Outs which allows you to use the SR-MPLS L3Out hand-offs exclusively for North-South connectivity between a site and an external network, while all East-West traffic is handled in the traditional Multi-Site manner using VXLAN-encapsulated data plane across the ISN. This means that the SR-MPLS hand-offs can now be treated as traditional IP-based hand-offs and the same VRF can deploy a mix of IP and SR-MPLS L3Outs. These changes add support for the following specific use cases:

- Centralized deployment in a site of an SR-MPLS L3Out which belongs to a specific VRF.

  All traffic from endpoints that are part of that VRF (connected to the same site or to different sites) can leverage that centralized SR-MPLS L3Out for North-South connectivity. Note that this requires the VRF to be stretched across sites.

- Deployment of multiple sites each with their own local SR-MPLS L3Outs and intra-VRF traffic using the local L3Out if it is available or a remote SR-MPLS L3Out from another site (intersite L3Out).

  In this case, the remote SR-MPLS L3Out can be used as a simple backup or to reach unique external prefixes received on the remote SR-MPLS L3Out. Traffic will transit from a local EPG to the local SR-MPLS L3Out and if that path is down or the route is unavailable it can take another site's remote SR-MPLS L3Out.

- Similar use cases are supported for shared services, where application EPG in one VRF can use SR-MPLS L3Out in a different VRF, either in the local or remote site.

  In this case, the EPGs can be in a different tenant as well. For example, Tenant1 in Site1 can contain the application EPGs which will use an SR-MPLS L3Out in Tenant2 in Site2.

- Ability to combine IP-based and SR-MPLS hand-offs.

- SR-MPLS hand-off can now be used to connect to Provider Edge (PE) devices that are part of an SR-MPLS core network as well as PEs that are part of a regular MPLS network, in which case the hand-off can be considered as a simple MPLS hand-off.

Using SR-MPLS L3Outs (instead of traditional IP-based L3Outs) allows for operational simplification at higher scale by removing the need for the VRF-Lite configuration which requires creation of separate BL nodes, BL logical interfaces, and routing peering for each VRF that needs to be connected to the external network. With the SR-MPLS L3Outs, the logical nodes and logical interfaces are defined once in the infra tenant, together with a single MP-BGP EVPN peering with the external devices. This infra L3Out construct can then be used to provide external connectivity to multiple tenant VRFs and all the VRFs' prefixes are exchanged using the common MP-BGP EVPN control plane.

The following sections describe guidelines, limitations, and configurations specific to managing Schemas that are deployed to sites from the Nexus Dashboard Orchestrator. Detailed information about MPLS hand off, supported individual site topologies (such as remote leaf support), and policy model is available in the *Cisco APIC Layer 3 Networking Configuration Guide*.

Configuration Workflow

Other sections in this document detail the required configurations, but in short you will go through the following workflow:

- Create an SR-MPLS QoS policy.

  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.

  This is an optional step and 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.

- Create an SR-MPLS Infra L3Out.

  This configures an L3Out for traffic leaving a site that is connected to an SR-MPLS network.

- Create SR-MPLS route map policy.

Route maps are sets of `if-then` rules that enable you to specify which routes are advertised out of the Tenant SR-MPLS L3Out. Route maps also enable you to specify which routes received from the DC-PE routers will be injected into the BGP VPNv4 ACI control plane.

- If you want to deploy a use case similar to releases prior to release 4.0(2), create the VRF, SR-MPLS L3Out, and SR-External EPG for each site connected via an SR-MPLS network and establish a contract within each site between that site's tenant EPG and SR-External EPG.

  In this case, all communication from one site will follow the North-South route egressing your Multi-Site domain towards the external SR-MPLS network. If the traffic is destined to an EPG in another site managed by your Orchestrator, it will ingress the other fabric from the external network using that site's SR-MPLS L3Out.

- If you want to use the SR-MPLS L3Outs in the same way as the standard IP-based L3Out exclusively for North-South communication, you can create the VRFs, SR-MPLS L3Outs, EPGs, and contracts as you typically would for all existing EPG-to-EPG communication use cases.

SR-MPLS Infra Requirements and Guidelines

If you want to use your Nexus Dashboard Orchestrator to manage SR-MPLS L3Out hand-offs for an ACI fabric connected to an SR-MPLS network:

- Any changes to the topology, such as node updates, are not reflected in the Orchestrator configuration until site configuration is refreshed, as described in Refreshing Site Connectivity Information.
- Tenants deployed to a site that is connected via an SR-MPLS network will have a set of unique configuration options specifically for SR-MPLS configuration. Tenant configuration is described in the "Tenants Management" chapter of the *Multi-Site Configuration Guide, Release 3.1(x)*.
- Remote leaf switches are not supported as source or destination for any Multi-Site traffic flows.
- SR-External EPGs that are part of preferred group cannot be the providers of a shared service (inter-VRF) contract.
- Preferred Group is not supported for Intersite SR-MPLS L3Outs.
- vzAny is not supported as a shared service provider.
- VRF that is enabled for Preferred Group cannot be a vzAny consumer.
- We recommend configuring tenant contract objects under dedicated template to avoid circular dependencies with other configuration objects that use the same contracts
- We recommend keeping SR L3Outs and External EPGs in a dedicated template.

  This will allow you to stretch other objects (such as VRFs, BDs, EPGs) to sites which do not have an infra SR L3Out.

- When using SR-MPLS L3Out instead of traditional IP-based L3Outs:
  - Host-based routing advertisement is not supported for bridge domains that are stretched across sites.
  - Tenant Routed Multicast (TRM) is not supported with SR-MPLS L3Outs, so they can only be used for establishing Layer 3 unicast communication with the external network domain.

Supported Hardware

The SR-MPLS hand-off is supported for the following platforms:

- **Border Leaf switches**: The "FX", "FX2", and "GX" switch models.
- **Spine switches**:
  - Modular spine switch models with "LC-EX", "LC-FX", and "GX" at the end of the linecard names.
  - The Cisco Nexus 9000 series N9K-C9332C and N9K-C9364C fixed spine switches.
- **DC-PE routers**:
  - Network Convergence System (NCS) 5500 Series
  - ASR 9000 Series
  - NCS 540 or 560 routers

SR-MPLS Infra L3Out

You will need to create an SR-MPLS Infra L3Out for the fabrics connected to SR-MPLS networks as described in the following sections. When creating an SR-MPLS Infra L3Out, the following restrictions apply:

- Each SR-MPLS Infra L3Out must have a unique name.

  The SR-MPLS Infra L3Out allows you to establish the control plane and data plane connectivity between the ACI border leaf switches and the external Provider Edge (PE) devices. SR-MPLS L3Outs that belong to various tenant VRFs can then leverage that Infra L3Out connectivity to establish communication with the external network domain.
- You can have multiple SR-MPLS infra L3Outs connecting to different routing domains, where the same border leaf switch can be in more than one L3Out, and you can have different import and export routing policies for the VRFs toward each routing domain.
- 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.
- If you have a multi-pod or remote leaf topology where one of the pods is not connected directly to the SR-MPLS network, that pod's traffic destined for the SR-MPLS network will use standard IPN path to another pod, which has an SR-MPLS L3Out. Then the traffic will use the other pod's SR-MPLS L3Out to reach its destination across SR-MPLS network.
- Routes from multiple VRFs can be advertised from one SR-MPLS Infra L3Out to provider edge (PE) routers connected to the nodes in this SR-MPLS Infra L3Out.

  PE routers can be connected to the border leaf directly or through other provider (P) routers.
- The underlay configuration can be different or can be the same across multiple SR-MPLS Infra L3Outs for one location.

For example, assume the same border leaf switch connects to PE-1 in domain 1 and PE-2 in domain 2, with the underlay connected to another provider router for both. In this case, two SR-MPLS Infra L3Outs will be created: one for PE-1 and one for PE-2. But for the underlay, it's the same BGP peer to the provider router. Import/export route-maps will be set for EVPN session to PE-1 and PE-2 based on the corresponding route profile configuration in the user VRF.

MPLS Custom QoS Policies

Following is the default MPLS QoS behavior:

- 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.

SR-MPLS Tenant Requirements and Guidelines

While the Infra MPLS configuration and requirements are described in the Day-0 operations chapter, the following restrictions apply for any user Tenants you will deploy to sites that are connected to SR-MPLS networks.

- In case when traffic between two EPGs in the fabric needs to go through the SR-MPLS network:
  - Contracts must be assigned between each EPG and the SR-External EPG defined on the local Tenant SR-MPLS L3Out.
  - If both EPGs are part of the same ACI fabric but separated by an SR-MPLS network (for example, in Multi-Pod or remote leaf cases), the EPGs must belong to different VRFs and not have a contract between them nor route-leaking configured.
  - If EPGs are in different sites, they can be in the same VRF, but there must **not** be a contract configured directly between the EPGs and any other remote EPG that is part of the same VRF.
- When configuring a route map policy for the SR-MPLS L3Out:
  - Each L3Out must have a single export route map. Optionally, it can also have a single import route map.
  - Routing maps associated with any SR-MPLS L3Out must explicitly define all the routes, including bridge domain subnets, which must be advertised out of the SR-MPLS L3Out.

- If you configure a `0.0.0.0/0` prefix and choose to not aggregate the routes, it will allow the default route only.

  However, if you choose to aggregate routes for the `0.0.0.0/0` prefix, it will allow all routes.

- You can associate any routing policy with any tenant L3Out.

- Beginning with Nexus Dashboard release 4.0(1), transit routing between SR-MPLS networks is supported using the same or different VRFs for fabrics running Cisco APIC release 5.1(1) or later.

Figure 3: Transit Routing Configuration Using Single VRF

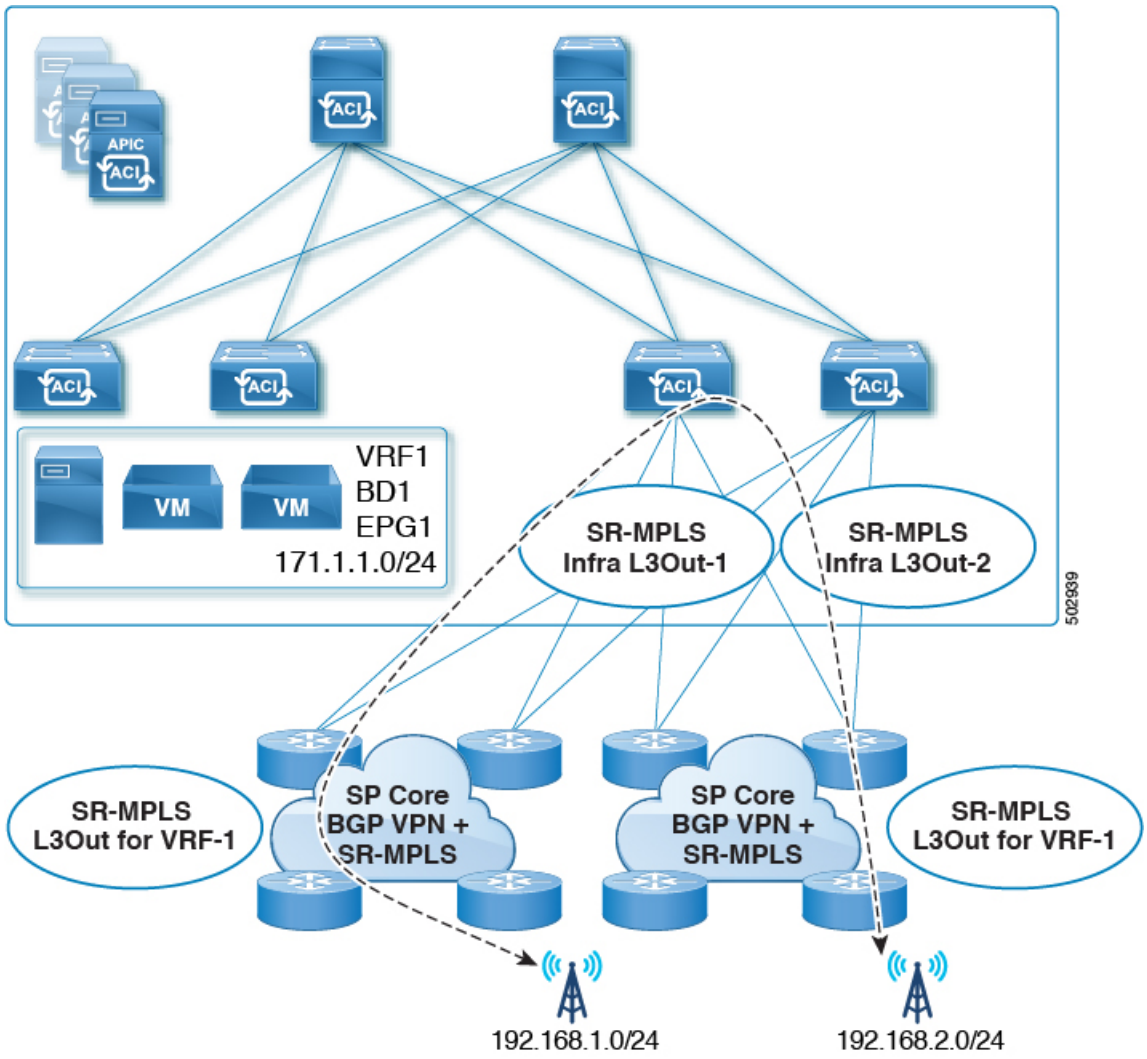
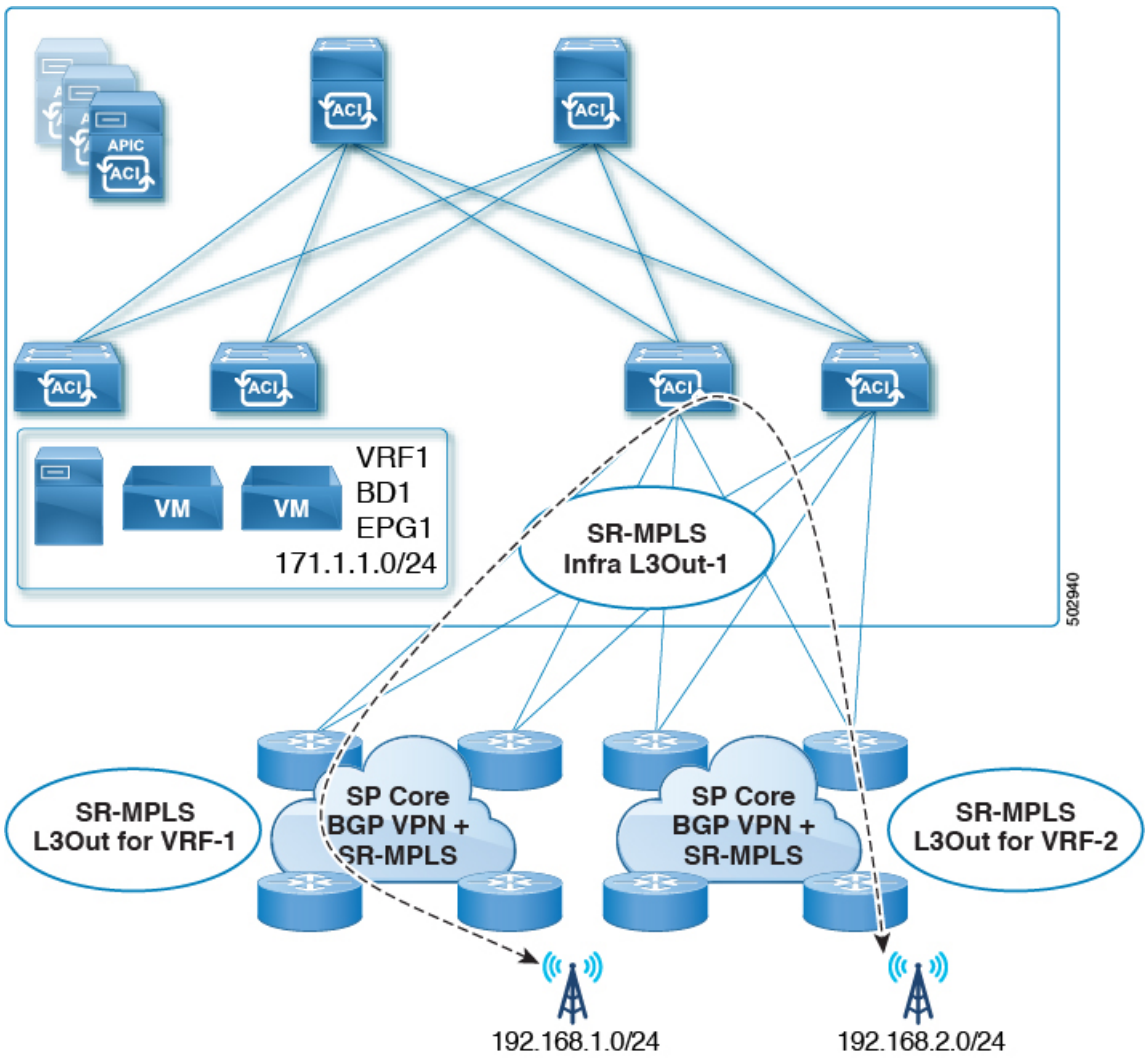


Figure 4: Transit Routing Configuration Using Different VRFs



Prior releases supported transit routing using different VRFs only.

Creating Custom QoS Policy for SR-MPLS

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.

Note Creating custom QoS policy is optional. 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 Log in to your Nexus Dashboard and open the Nexus Dashboard Orchestrator service.

Step 2 Create a new Fabric Policy.

a) From the left navigation pane, choose **Fabric Management** > **Fabric Policies**.
b) On the **Fabric Policy Templates** page, click **Add Fabric Policy Template**.
c) From the **+Create Object** dropdown, select **QoS SR-MPLS**.
d) In the right properties sidebar, provide the **Name** for the policy.
e) (Optional) Click **Add Description** and provide a description for the policy.

Step 3 Click **Add Ingress Rule** to add an ingress QoS translation rule.

These rules are applied for traffic that is ingressing the ACI fabric from an MPLS network and are used to map incoming packet's experimental bits (EXP) values to ACI QoS levels, as well as set DSCP and/or CoS values that the packet should have set when forwarded to an endpoint connected to the fabric.

The values are derived at the border leaf using a custom QoS translation policy. If a custom policy is not defined or not matched, default QoS Level (`Level3`) is assigned

a) In the **Match EXP From** and **Match EXP To** fields, specify the EXP range of the ingressing MPLS packet you want to match.
b) From the **Queuing Priority** dropdown, select the ACI QoS Level to map.

   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 `Level3`. If you do not make a selection in this field, the traffic will automatically be assigned a `Level3` priority.

c) From the **Set DSCP** dropdown, select the DSCP value to be used when sending the un-encapsulated packet to an endpoint connected to the 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.

   If you set the value to `Unspecified`, the original DSCP value of the packet will be retained.

d) From the **Set CoS** dropdown, select the CoS value to be used when sending the un-encapsulated packet to an endpoint connected to the fabric.

   The CoS value specified will be re-exposed only when the traffic is VXLAN decapsulated on the destination ACI leaf node.

   If you set the value to `Unspecified`, the original CoS value of the packet will be retained.

   In both of the above cases, the CoS preservation option must be enabled in the fabric. For more information about CoS preservation, see *Cisco APIC and QoS*.

e) Click the checkmark icon to save the rule.
f) Repeat this step for any additional ingress QoS policy rules.

Step 4 Click **Add Egress Rule** to add an egress QoS translation rule.

These rules are applied for the traffic that is leaving the ACI fabric via an MPLS L3Out and are used to map the packet's IPv4 DSCP value to the MPLS packet's EXP value as well as the internal Ethernet frame's CoS value.

The setting of the packet's IPv4 DSCP value is done at the non-border leaf switch based on existing policies used for EPG and L3Out traffic. If a custom policy is not defined or not matched, the default EXP value of `0` is marked on all labels. EXP values are marked in both, default and custom policy scenarios, and are done on all MPLS labels in the packet.

Custom MPLS egress policy can override existing EPG, L3Out, and Contract QoS policies.

a) Using the **Match DSCP From** and **Match DSCP 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 **Set MPLS EXP** dropdown, select the EXP value you want to assign to the egressing MPLS packet.
c) From the **Set CoS** dropdown, select the CoS value you want to assign to the egressing MPLS packet.
d) Click the checkmark icon to save the rule.
e) Repeat this step for any additional egress QoS policy rules.

Step 5 From the **Actions** menu, select **Sites Association** and choose the SR-MPLS site with which to associate this template.

Step 6 Click **Save** to save the template policy.

Step 7 Click **Deploy** to deploy the fabric policy to the sites.

What to do next

After you have created the QoS policy, enable MPLS connectivity and configure MPLS L3Out as described in Creating SR-MPLS Infra L3Out, on page 11.

Creating SR-MPLS Infra L3Out

This section describes how to configure SR-MPLS Infra L3Out settings for a site that is connected to an SR-MPLS network.

- 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 hand-off.
- An SR-MPLS infra L3Out will be scoped to a pod or a remote leaf switch site.
- Border leaf switches or remote leaf switches in one SR-MPLS infra L3Out can connect to one or more provider edge (PE) routers in one or more routing domains.
- A pod or remote leaf switch site can have one or more SR-MPLS infra L3Outs.

Before you begin

You must have:

- Added a site that is connected via SR-MPLS network as described in Adding Cisco ACI Sites.
- If necessary, created SR-MPLS QoS policy as described in Creating Custom QoS Policy for SR-MPLS, on page 9.

Procedure

Step 1 Ensure that SR-MPLS Connectivity is enabled for the site.

a) In the main navigation menu, select **Infrastructure** > **Site Connectivity**.
b) In the **Site Connectivity** page, click **Configure**.
c) In the left pane, under **Sites**, select the specific site that is connected via SR-MPLS.
d) In the right ***<Site>*** **Settings** pane, enable the **SR-MPLS Connectivity** and provide the SR-MPLS information.

- The **Segment Routing Global Block (SRGB) Range** is the range of label values reserved for Segment Routing (SR) in the Label Switching Database (LSD). The Segment ID (SID) is a unique identifier for a specific segment and is configured on each node for the MPLS transport loopback. The SID value, which you will configure later as part of the border leaf configuration, is advertised using BGP-LU to the peer router, and the peer router uses the SID index to calculate the local label.

  The default range is `16000-23999`.

- The **Domain ID Base** enables the BGP Domain-Path feature. For more information, see *Cisco APIC Layer 3 Networking Configuration Guide*.

  If you choose to provide a value in this field to enable the Domain-Path feature, ensure that you use a unique value for each SR-MPLS site in your Multi-Site domain, which will be specific to this ACI fabric.

Step 2 In the main pane, click **+Add SR-MPLS L3Out** within a pod.

Step 3 In the right **Properties** pane, provide a name for the SR-MPLS L3Out.

Step 4 (Optional) From the **QoS Policy** dropdown, select a QoS Policy you created for SR-MPLS traffic.

Select the QoS policy you created in Creating Custom QoS Policy for SR-MPLS, on page 9.

Otherwise, if you do not assign 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.

Step 5 From the **L3 Domain** dropdown, select the Layer 3 domain.

Step 6 Configure settings for border leaf switches and ports connected to the SR-MPLS network.

You need to provide information about the border leaf switches as well as the interface ports which connect to the SR-MPLS network.

a) Click **+Add Leaf** to add a leaf switch.
b) In the **Add Leaf** window, select the leaf switch from the **Leaf Name** dropdown.

c) In the **SID Index** field, provide a valid segment ID (SID) offset.

When configuring the interface ports later in this section, you will be able to choose whether you want to enable segment routing. 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. If you plan to enable segment routing, you must specify the segment ID for this border leaf.

- The value must be within the SRGB range you configured earlier.
- The value must be the same for the selected leaf switch across all SR-MPLS L3Outs in the site.
- The same value cannot be used for more than one leaf across all sites.
- If you need to update the value, you must first delete it from all SR-MPLS L3Outs in the leaf and re-deploy the configuration. Then you can update it with the new value, followed by re-deploying the new configuration.

d) Provide the local **Router ID**.

Unique router identifier within the fabric.

e) Provide the **BGP EVPN Loopback** address.

The BGP-EVPN loopback is used for the BGP-EVPN 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 BGP-EVPN loopback and the BGP-EVPN remote peer address (configured in the **MPLS BGP-EVPN Peer IPv4 Address** field in the **BGP Connectivity** step before).

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.

f) Provide the **MPLS Transport Loopback** address.

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.

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.

g) Click **Add Interface** to provide switch interface details.

From the **Interface Type** dropdown, select whether it is a Layer 3 physical interface or a port channel interface. If you choose to use a port channel interface, it must have been already created on the APIC.

Then provide the interface, its IP address, and MTU size. If you want to use a subinterface, provide the **VLAN ID** for the sub-interface, otherwise leave the VLAN ID field blank.

In the **BGP-Label Unicast Peer IPv4 Address** and **BGP-Label Unicast Remote AS Number**, specify the BGP-LU peer information of the next hop device, which is the device connected directly to the interface. The next hop address must be part of the subnet configured for the interface.

Choose whether you want to enable an MPLS or an SR-MPLS hand-off.

(Optional) Choose to enable the additional BGP options based on your deployment.

Finally, click the checkmark to the right of the **Interface Type** dropdown to save interface port information.

h) Repeat the previous sub-step for all interfaces on the switch that connect to the MPLS network.

i) Click **Save** to save the leaf switch information.
j) Repeat this step for all leaf switches connected to MPLS networks.

Step 7 Configure BGP settings.

You must provide BGP connectivity details for the BGP EVPN connection between the site's border leaf (BL) switches and the provider edge (PE) router.

a) Click **+Add BGP-EVPN Connectivity**.
b) In the **Add MPLS BGP-EVPN Connectivity** window, provide the details.

For the **MPLS BGP-EVPN Peer IPv4 Address** field, provide the loopback IP address of the DC-PE router, which is not necessarily the device connected directly to the border leaf.

For the **Remote AS Number**, 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. Keep in mind, ACI supports only `asplain` format and not `asdot` or `asdot+` format AS numbers. For more information on ASN formats, see Explaining 4-Byte Autonomous System (AS) ASPLAIN and ASDOT Notation for Cisco IOS document.

For the **TTL** field, specify a number large enough to account for multiple hops between the border leaf and the DC-PE router, for example `10`. The allowed range is `2-255` hops.

(Optional) Choose to enable the additional BGP options based on your deployment.

c) Click **Save** to save BGP settings.
d) Repeat this step to for any additional BGP connections.

Typically, you would be connecting to two DC-PE routers, so provide BGP peer information for both connections.

Step 8 Deploy the changes to sites.

What to do next

After you have enabled and configured MPLS connectivity, you can create and manage Tenants, route maps, and schemas as described in the Creating SR-MPLS Route Map Policy, on page 14.

Creating SR-MPLS Route Map Policy

This section describes how to create a route map policy. Route maps are sets of `if-then` rules that enable you to specify which routes are advertised out of the Tenant SR-MPLS L3Out. Route maps also enable you to specify which routes received from the DC-PE routers will be injected into the BGP VPNv4 ACI control plane.

You will use the SR-MPLS route map policy in the next section when defining site-local settings for the SR-MPLS L3Out.

Procedure

Step 1 Log in to your Nexus Dashboard and open the Nexus Dashboard Orchestrator service.

Step 2 Create a new Tenant Policy.

a) From the left navigation pane, choose **Application Management** > **Tenant Policies**.

b) On the **Tenant Policy Templates** page, click **Add Tenant Policy Template**.
c) In the Tenant Policies page's right properties sidebar, provide the **Name** for the tenant.
d) From the **Select a Tenant** dropdown, choose the tenant with which you want to associate this template.

All the policies you create int his template as described in the following steps will be associated with the selected tenant deployed to it when you push the template to a specific site.

By default, the new template is empty, so you need to add one or more tenant policies as described in the following steps. Note that you don't have to create every policy available in the template – you can create a template with just a single route map policy for your SR-MPLS use case.

Step 3 Create a Route Map Policy for Route Control.

a) From the **+Create Object** dropdown, select **Route Map Policy for Route Control**.
b) In the right properties sidebar, provide the **Name** for the policy.
c) (Optional) Click **Add Description** and provide a description for the policy.
d) Click **+Add Entry** and provide the route map information.

For each route map, you need to create one or more context entries. Each entry is a rule that defines an action based on one or more matching criteria based on the following information:

- **Context Order** – Context order is used to determine the order in which contexts are evaluated. The value must be in the `0-9` range.
- **Context Action** – Context action defines the action to perform (`permit` or `deny`) if a match is found.

Once the context order and action are defined, choose how you want to match the context:

- Click **+Add Attribute** to specify the action that will be taken should the context match.

  You can choose one of the following actions:

  - `Set Community`
  - `Set Route Tag`
  - `Set Dampening`
  - `Set Weight`
  - `Set Next Hop`
  - `Set Preference`
  - `Set Metric`
  - `Set Metric Type`
  - `Set AS Path`
  - `Set Additional Community`

  After you have configured the attribute, click **Save**.

- If you want to match an action based on an IP address or prefix, click **Add IP Address**.

  In the **Prefix** field, provide the IP address prefix. Both IPv4 and IPv6 prefixes are supported, for example `2003:1:1a5:1a5::/64` or `205.205.0.0/16`.

If you want to aggregate IPs in a specific range, check the **Aggregate** checkbox and provide the range. For example, you can specify `0.0.0.0/0` prefix to match any IP or you can specify `10.0.0.0/8` prefix to match any `10.x.x.x` addresses.

- If you want to match an action based on community lists, click **Add Community**.

  In the **Community** field, provide the community string. For example, `regular:as2-nn2:200:300`.

  Then choose the **Scope**: `Transitive` means the community will be propagated across eBGP peering (across autonomous systems) while `Non-Transitive` means the community will not be propagated.

e) Repeat the previous substeps to create any additional route map entries for the same policy.
f) Click **Save** to save the policy and return to the template page.
g) Repeat this step to create any additional Route Map for Route Control policies.

Step 4 From the **Actions** menu, select **Sites Association** and choose the SR-MPLS site with which to associate this template.

Step 5 Click **Deploy** to deploy the tenant policy to the sites.

Configure EPG-to-External-EPG (North-South) Communication

This section describes how to establish North-South communication between an application EPG and an external SR-MPLS network. You can also use this approach to enable EPG-to-EPG communication across sites via the SR-MPLS L3Out data path (leveraging the external SR-MPLS network).

If instead you want to establish EPG-to-EPG intersite connectivity via the VXLAN data plane across the ISN which is supported starting with release 4.0(2), you can simply establish a contract relationship between those EPGs as you typically would.

Procedure

Step 1 Choose the template or create a new one.

You can select the template as you typically would for other ACI fabric use cases:

a) In the main navigation menu, select **Application Management** > **Schemas**.
b) Select an existing schema or create a new one.
c) Select an existing template or click **Add New Template** and select `ACI Multi-Cloud` for template type.
d) Select the tenant for the new template.

   The tenant must be associated with the SR-MPLS site.

e) (Optional) Enable the **Autonomous** option for the template if you plan to deploy this template only to sites that do not have any intersite connectivity to other sites.

Step 2 Create a VRF.

a) From the **+Create Object** menu, choose **VRF**.
b) In the right properties sidebar, provide the name for the VRF.

Step 3 Create an SR-External EPG.

Note

If you assign the template that contains SR-External EPG to multiple sites, the EPG will be stretched to all of those sites. In this case each site must have a local SR-MPLS L3Out or you will not be allowed to deploy that template to all associated sites.

a) From the **+Create Object** menu, choose **SR-External EPG**.
b) In the right properties sidebar, provide the name for the external EPG.
c) From the **Virtual Routing & Forwarding** dropdown, select the VRF you created in the previous step.

Step 4 Create an SR-MPLS L3Out.

Note
The tenant SR-MPLS L3Out must be defined in the same template as the SR-External EPG you created in the previous step.

a) From the **+Create Object** menu, choose **SR-L3Out**.
b) In the right properties sidebar, provide the name for the L3Out.
c) From the **Virtual Routing & Forwarding** dropdown, select the same VRF you selected for the external EPG in the previous step.
d) From the **SR-External EPGs** dropdown, select the SR-External EPG you created in the previous step.

Step 5 Assign the template to a single site or to multiple sites, depending on the specific use case you need to configure.

Step 6 Select the site-local settings for the template you are configuring.

In the following few steps, you will configure site local settings for the VRF, SR-External EPG, and SR-MPLS L3Out you created in the previous steps.

Step 7 Configure site-local settings for the VRF.

You must provide BGP route information for the VRF used by the SR-MPLS L3Out.

a) In the main pane, scroll down to **VRF** area and select the VRF you created in the previous step.
b) From the **Address Family** dropdown, select whether it is IPv4 or IPv6 address.
c) In the **Route Target** field, provide the route string.

   For example, `route-target:ipv4-nn2:1.1.1.1:1901`.

d) From the **Type** dropdown, select whether to import or export the route.
e) Click **Save** to save the route information.
f) (Optional) Repeat this step to add any additional BGP route targets.

Step 8 Configure site-local settings for the SR-MPLS L3Out.

a) In the main pane, scroll down to **SR-MPLS L3Out** area and select the MPLS L3Out.
b) Click **+Add SR-MPLS Infra L3Out** to add the Infra SR-MPLS L3Out information.
c) In the **Add SR-MPLS Infra L3Out** window, select the Infra SR-MPLS L3Out you created when configuring Infra for that site.
d) Click **Add Route Map Policy** and select the route map policy you created in the tenant policy template in previous section, then specify whether you want to import or export the routes.

   You must configure a single export route map policy. Optionally, you can configure an additional import route map policy.

e) Click **Ok** to save the changes.

Step 9 Create and configure an application EPG as you typically would.

Note

The EPG can be in the same or different template and schema.

Step 10 Create a contract between the application EPG and the SR-External EPG.

Step 11 Deploy the configuration

a) In the main pain of the **Schemas** view, click **Deploy to Sites**.

b) In the **Deploy to Sites** window, verify the changes that will be pushed to the site and click **Deploy**.

Note

Starting from release 4.0(2), it is possible to use the EPG-to-SR-External-EPG contracts exclusively for North-South traffic (communication with resources external to the ACI fabrics), similar to the traditional IP-based L3Outs. In that case, EPG-to-EPG intersite communication can be enabled via the VXLAN data path across the ISN by simply creating a contract relationship between those EPGs.

However, if you want to establish EPG-to-EPG (East-West) communication between EPGs in different sites across the external SR-MPLS network, you can do that as outlined in the next step.

Step 12 If you want to use the SR-MPLS L3Out data path for EPG-to-EPG traffic across sites (leveraging the SR-MPLS external network instead of the VXLAN data path across the ISN), you can establish contracts between each site-local EPG and the SR-External EPG associated to the tenant SR-MPLS L3Out.

The SR-External EPG can be deployed as a site-local object in each site or as a stretched object across sites. Note that using the SR-MPLS L3out data path for EPG-to-EPG traffic across sites is only possible if there are no direct contract relationships between those EPGs or between each EPG and any other remote EPG

a) Create two application EPGs as you typically would in templates associated to different sites.

For example, `epg1` and `epg2`.

These EPGs can be in the same or different VRFs or Tenants.

b) Create two separate site-local SR-External EPGs or a single stretched SR-External EPG.

If you are creating separate SR-External EPGs, they can be in the same or different VRFs or Tenants and the same template or different templates depending on the specific deployment scenario.

For example, `mpls-extepg-1` and `mpls-extepg-2`

c) Configure two separate site-local tenant SR-MPLS L3Outs or a single stretched tenant SR-MPLS L3Out.

For example, `mpls-l3out-1` and `mpls-l3out-2`

d) Create a contract you will use to allow traffic between each site local EPG and SR-MPLS L3Out local connection.

You will need to create and define a filter for the contract as you typically would.

e) Assign the contracts to the appropriate EPGs.

In order to allow traffic between the two application EPGs you created, you will actually need to assign the contract twice: once between `epg1` and its `mpls-extepg-1` and then again between `epg2` and its `mpls-extepg-2`. Note that it's possible to have the same SR-External EPG instead of two separate ones if it is stretched across sites.

As an example, if you want `epg1` to provide a service to `epg2`, you would:

- Assign the contract to `epg1` with type `provider`.
- Assign the contract to `mpls-l3out-1` with type `consumer`.
- Assign the contract to `epg2` with type `consumer`.

- Assign the contract to `mpls-extepg-2` with type `provider`.