Routed Connectivity to External Networks

This chapter contains the following sections:

About Routed Connectivity to Outside Networks

A Layer 3 outside network configuration (L3Out) defines how traffic is forwarded outside of the fabric. Layer 3 is used to discover the addresses of other nodes, select routes, select quality of service, and forward the traffic that is entering, exiting, and transiting the fabric.

Note For guidelines and cautions for configuring and maintaining Layer 3 outside connections, see Guidelines for Routed Connectivity to Outside Networks, on page 14.

For information about the types of L3Outs, see External Layer 3 Outside Connection Types.

Create L3Out Wizard

A new Create L3Out wizard is introduced in APIC release 4.2(1) that provides a straightforward walk-through for configuring an L3Out.

The Create L3Out wizard streamlines the process for configuring an L3Out, which defines how the ACI fabric connects to external layer 3 networks. With the Create L3Out wizard, you make the necessary basic configurations for the L3Out components in the following pages:

- **Identity page**: This page is used to configure the basic settings for the L3Out, as well as the static routing and dynamic routing protocols settings.

- **Nodes and Interfaces page**: This page is used to configure the node profiles and interface profiles for the Layer 3 and Layer 2 interface types.

- **Protocols page**: This page is used to configure specific polices based on the protocols that you selected in the Identity page.
- **External EPG page**: This page is used to configure the contract and subnets for the external EPG.

MP-BGP Route Reflectors

Configuring an MP-BGP Route Reflector Using the GUI

Procedure

Step 1 On the menu bar, choose **System** > **System Settings**.

Step 2 In the **Navigation** pane, right-click **BGP Route Reflector**, and click **Create Route Reflector Node**.

Step 3 In the **Create Route Reflector Node** dialog box, from the **Spine Node** drop-down list, choose the appropriate spine node. Click **Submit**.

Note Repeat the above steps to add additional spine nodes as required.

The spine switch is marked as the route reflector node.

Step 4 In the **BGP Route Reflector** properties area, in the **Autonomous System Number** field, choose the appropriate number. Click **Submit**.

Note The autonomous system number must match the leaf connected router configuration if Border Gateway Protocol (BGP) is configured on the router. If you are using routes learned using static or Open Shortest Path First (OSPF), the autonomous system number value can be any valid value.

Step 5 On the menu bar, choose **Fabric** > **Fabric Policies** > **Pods** > **Policy Groups**.

Step 6 In the **Navigation** pane, expand and right-click **Policy Groups**, and click **Create Pod Policy Group**.

Step 7 In the **Create Pod Policy Group** dialog box, in the **Name** field, enter the name of a pod policy group.

Step 8 In the **BGP Route Reflector Policy** drop-down list, choose the appropriate policy (default). Click **Submit**.
The BGP route reflector policy is associated with the route reflector pod policy group, and the BGP process is enabled on the leaf switches.

Step 9 On the menu bar, choose **Fabric** > **Fabric Policies** > **Profiles** > **Pod Profile default** > **default**.

Step 10 In the **Work** pane, from the **Fabric Policy Group** drop-down list, choose the pod policy that was created earlier. Click **Submit**.
The pod policy group is now applied to the fabric policy group.

Verifying the MP-BGP Route Reflector Configuration

Procedure

Step 1 Verify the configuration by performing the following actions:

a) Use secure shell (SSH) to log in as an administrator to each leaf switch as required.
b) Enter the **show processes | grep bgp** command to verify the state is S.

If the state is NR (not running), the configuration was not successful.

Step 2 Verify that the autonomous system number is configured in the spine switches by performing the following actions:

a) Use the SSH to log in as an administrator to each spine switch as required.
b) Execute the following commands from the shell window

Example:

cd /mit/sys/bgp/inst

Example:

grep asn summary

The configured autonomous system number must be displayed. If the autonomous system number value displays as 0, the configuration was not successful.

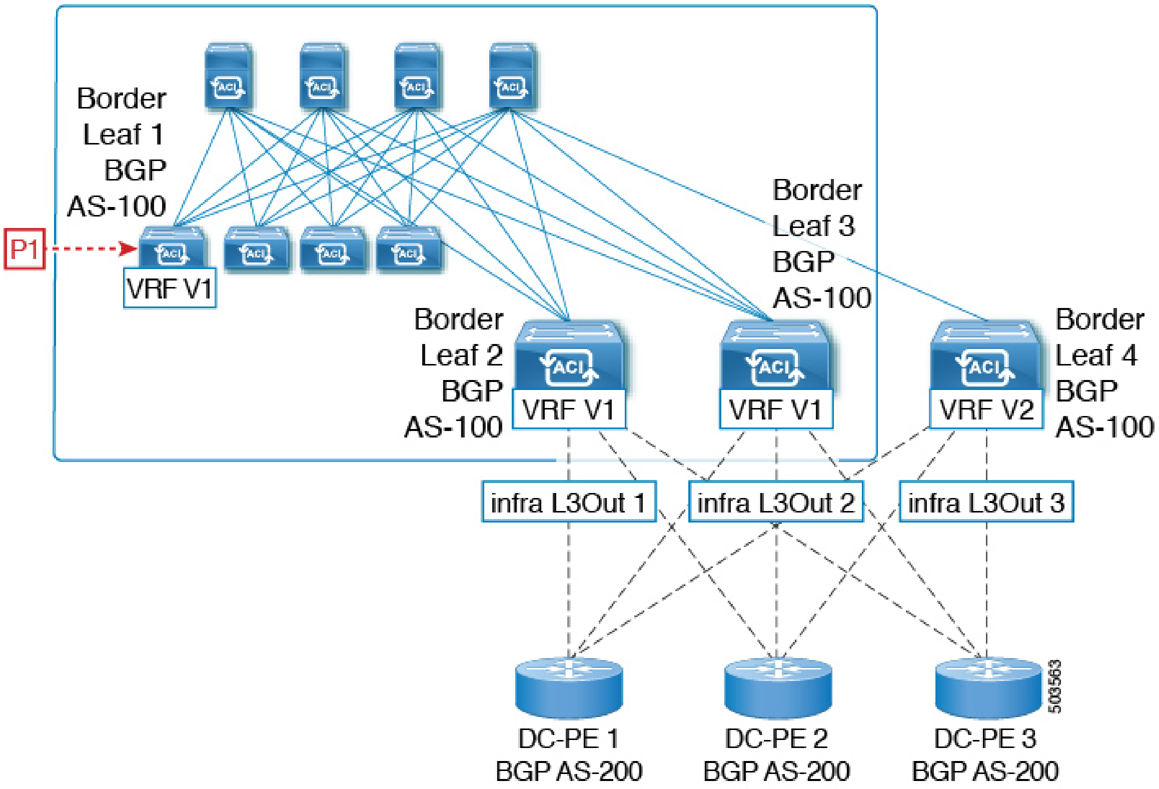
About the BGP Domain-Path Feature for Loop Prevention

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

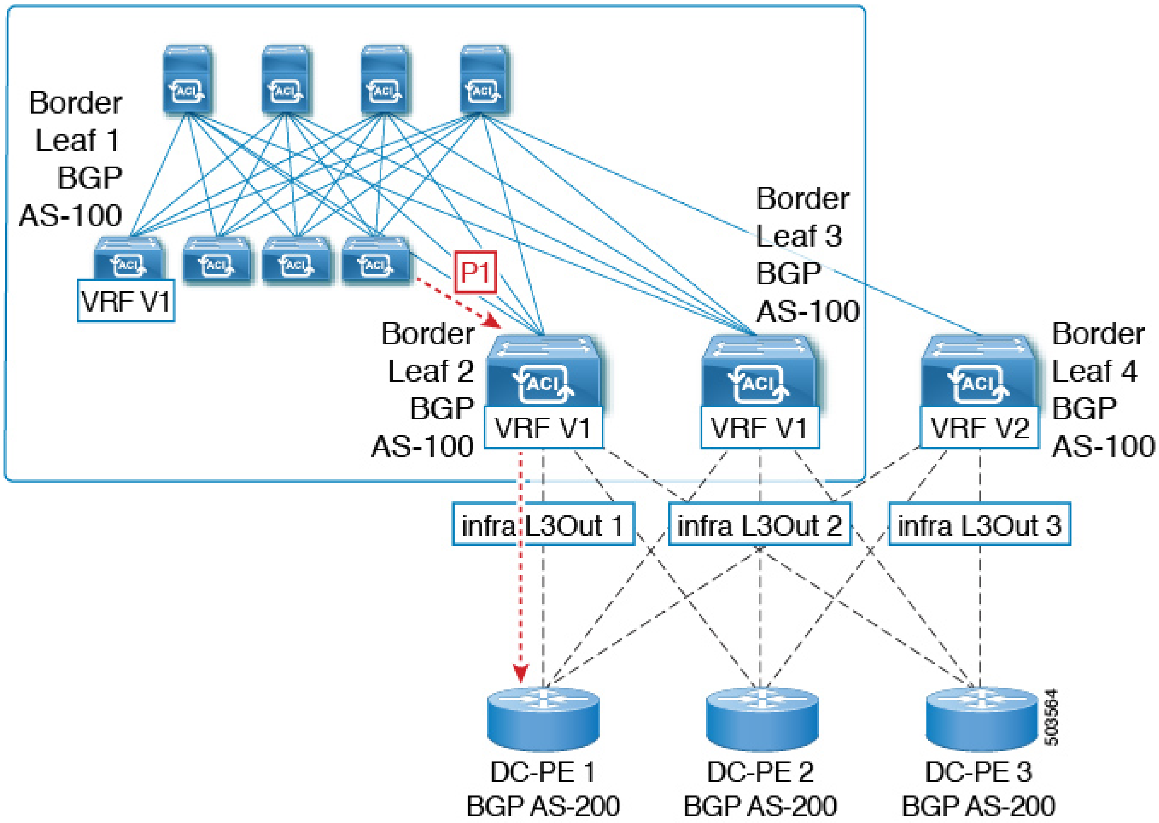
- Intentional disabling of existing BGP loop prevention mechanisms, such as AS Path checks
- Route leaks across different VRFs or VPNs

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

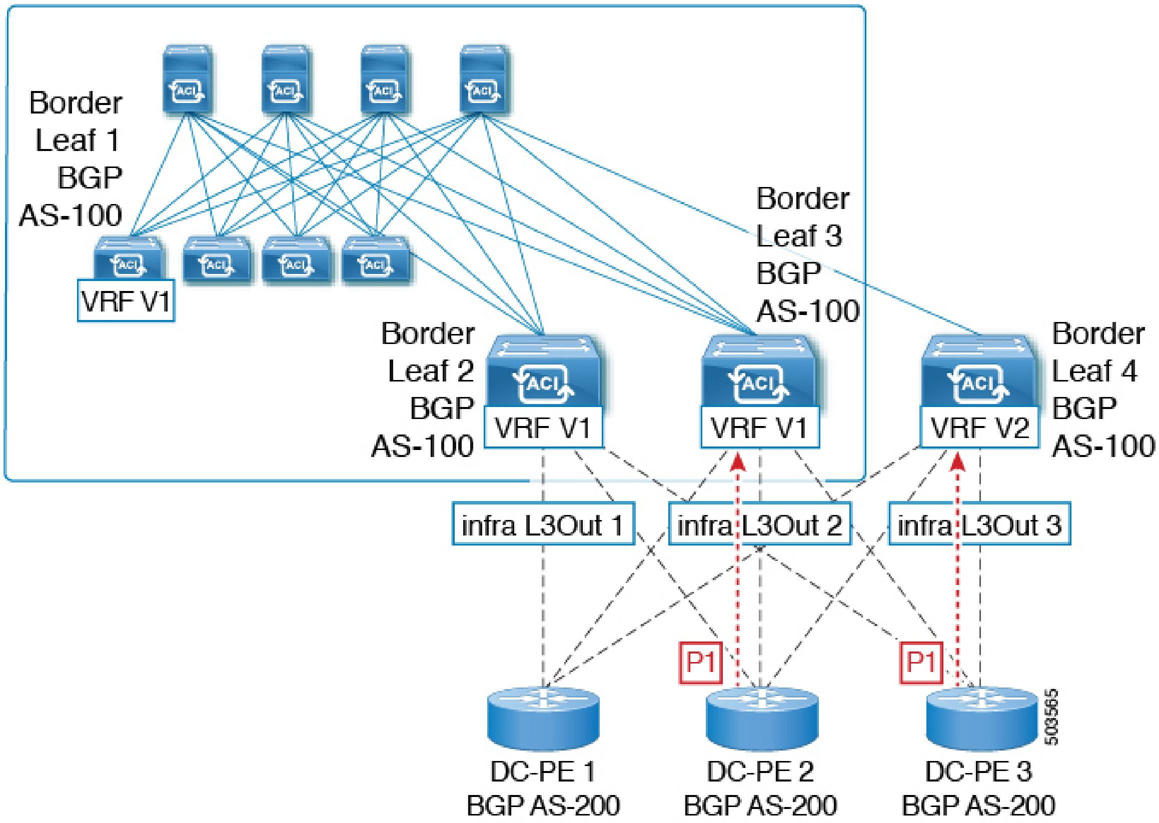
1. A prefix P1 received from a BGP IP L3Out peer is advertised in the ACI fabric using the Multiprotocol Border Gateway Protocol (MP-BGP).



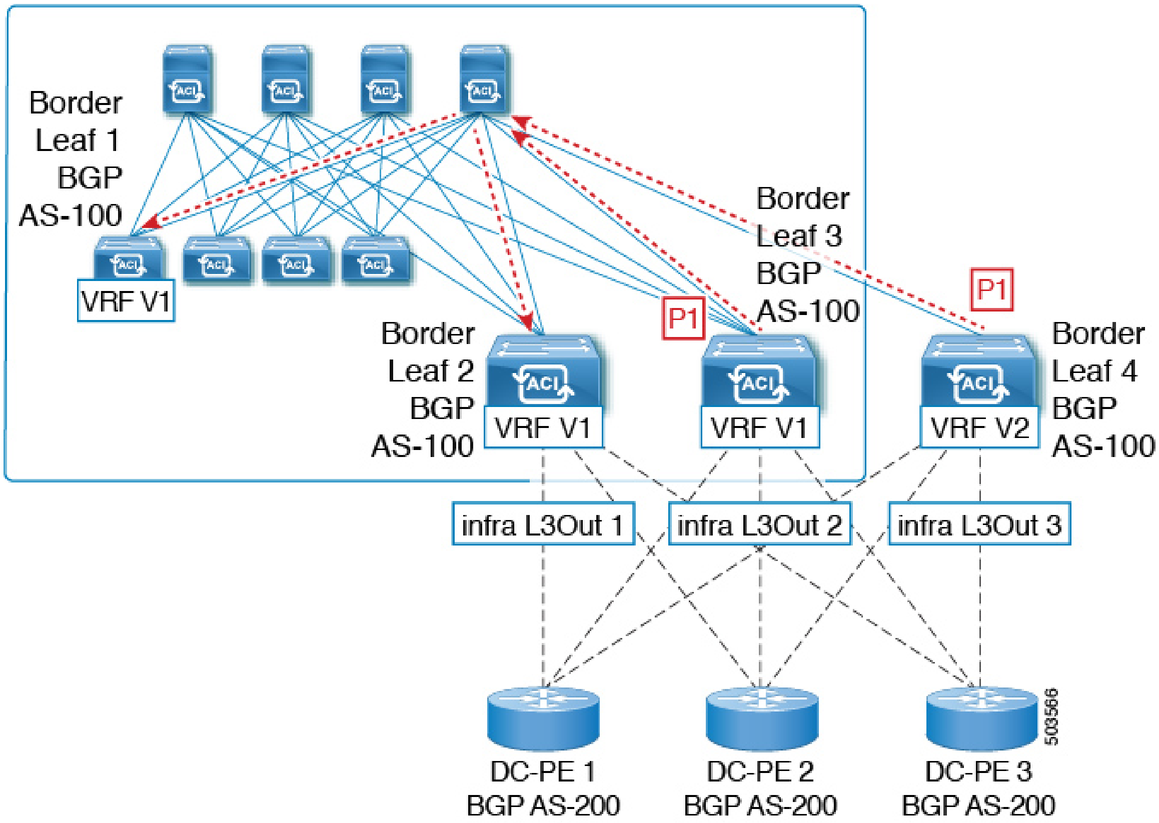
2. As a transit case, this prefix can be advertised out externally through an SR-MPLS infra L3Out.



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.



4. 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 3.

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 9 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 9, 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 9 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.

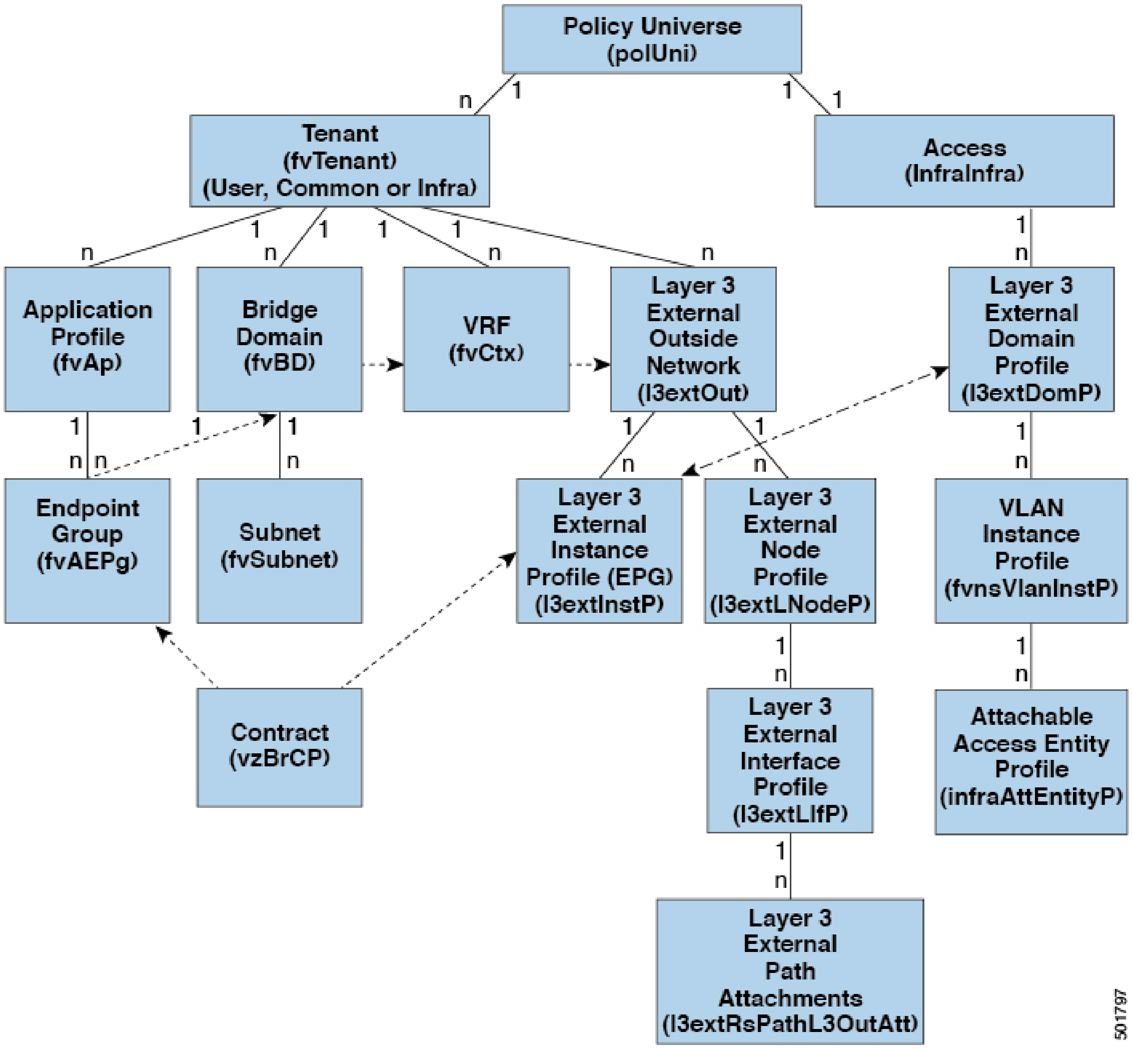
- To see the routing ID assigned to the SR-MPLS infra L3Out domain, navigate to:

  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 VRFL3Out.

Layer 3 Out for Routed Connectivity to External Networks

Routed connectivity to external networks is enabled by associating a fabric access (`infraInfra`) external routed domain (`l3extDomP`) with a tenant Layer 3 external instance profile (`l3extInstP` or external EPG) of a Layer 3 external outside network (`l3extOut`), in the hierarchy in the following diagram:

Figure 1: Policy Model for Layer 3 External Connections


Policy Universe (polUni)
Tenant (fvTenant) (User, Common or Infra)
Access (InfraInfra)
Application Profile (fvAp)
Bridge Domain (fvBD)
VRF (fvCtx)
Layer 3 External Outside Network (l3extOut)
Layer 3 External Domain Profile (l3extDomP)
Endpoint Group (fvAEPg)
Subnet (fvSubnet)
Layer 3 External Instance Profile (EPG) (l3extInstP)
Layer 3 External Node Profile (l3extLNodeP)
VLAN Instance Profile (fvnsVlanInstP)
Contract (vzBrCP)
Layer 3 External Interface Profile (l3extLIfP)
Attachable Access Entity Profile (infraAttEntityP)
Layer 3 External Path Attachments (l3extRsPathL3OutAtt)
1
n
501797

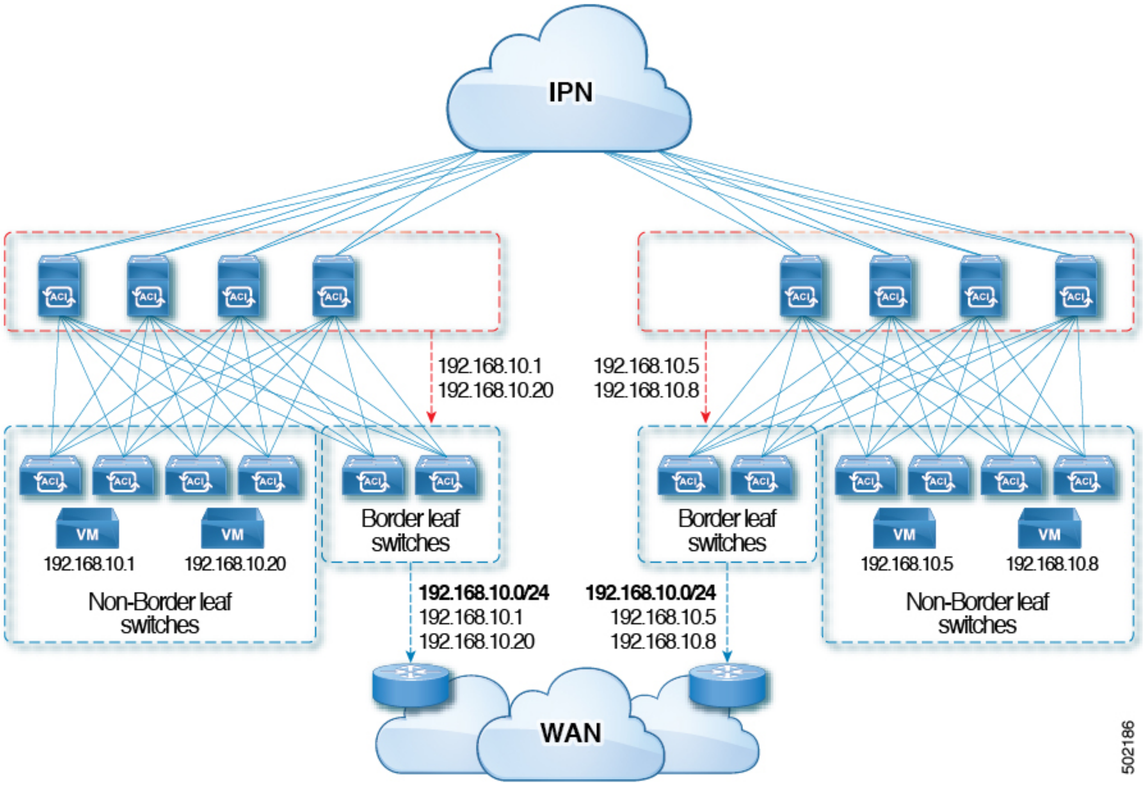
A Layer 3 external outside network (`l3extOut` object) includes the routing protocol options (BGP, OSPF, or EIGRP or supported combinations) and the switch-specific and interface-specific configurations. While the `l3extOut` contains the routing protocol (for example, OSPF with its related Virtual Routing and Forwarding (VRF) and area ID), the Layer 3 external interface profile contains the necessary OSPF interface details. Both are needed to enable OSPF.

The `l3extInstP` EPG exposes the external network to tenant EPGs through a contract. For example, a tenant EPG that contains a group of web servers could communicate through a contract with the `l3extInstP` EPG according to the network configuration contained in the `l3extOut`. The outside network configuration can easily be reused for multiple nodes by associating the nodes with the L3 external node profile. Multiple nodes that use the same profile can be configured for fail-over or load balancing. Also, a node can be added to multiple l3extOuts resulting in VRFs that are associated with the l3extOuts also being deployed on that node. For scalability information, refer to the current *Verified Scalability Guide for Cisco ACI*.

Advertise Host Routes

Enabling Advertise Host Routes on the BD, individual host-routes (/32 and /128 prefixes) are advertised from the Border-Leaf switches (BL). The BD must be associated to the L3out or an explicit prefix list matching the host routes. The host routes must be configured to advertise host routes out of the fabric.

Border-Leaf switches along with the subnet advertise the individual end-point(EP) prefixes. The route information is advertised only if the host is connected to the local POD. If the EP is moved away from the local POD or once the EP is removed from EP database (even if the EP is attached to a remote leaf), the route advertisement is then withdrawn.



Advertise Host Route configuration guidelines and limitations are:

- When host routes are advertised, the VRF Transit Route Tag is set in order to prevent them from being advertised back into the fabric and installed. In order for this loop protection to work properly, external routers must preserve this route-tag if advertising to another L3Out.
- If a bridge domain is tied to an EPG that has the same subnet configured for internal leaking, you must also enable the "Advertised Externally" flag on the EPG subnet.
- The Advertise Host Routes feature is supported on Generation 2 switches or later (Cisco Nexus N9K switches with "EX", "FX", or "FX2" on the end of the switch model name or later; for example, N9K-93108TC-EX).
- Host route advertisement supports both BD to L3out Association and the explicit route map configurations. We recommend using explicit route map configuration which allows you greater control in selecting individual or a range of host routes to configure.
- EPs/Host routes in SITE-1 will not be advertised out through Border Leafs in other SITEs.

- When EPs is aged out or removed from the database, Host routes are withdrawn from the Border Leaf.
- When EP is moved across SITEs or PODs, Host routes should be withdrawn from first SITE/POD and advertised in new POD/SITE.
- EPs learned on a specific BD, under any of the BD subnets are advertised from the L3out on the border leaf in the same POD.
- EPs are advertised out as Host Routes only in the local POD through the Border Leaf.
- Host routes are not advertised out from one POD to another POD.
- In the case of Remote Leaf, if EPs are locally learned in the Remote Leaf, they are then advertised only through a L3out deployed in Remote Leaf switches in same POD.
- EPs/Host routes in a Remote Leaf are not advertised out through Border Leaf switches in main POD or another POD.
- EPs/Host routes in the main POD are not advertised through L3out in Remote Leaf switches of same POD or another POD.
- The BD subnet must have the **Advertise Externally** option enabled.
- The BD must be associated to an L3out or the L3out must have explicit route-map configured matching BD subnets.
- There must be a contract between the EPG in the specified BD and the External EPG for the L3out.

Note If there is no contract between the BD/EPG and the External EPG the BD subnet and host routes will not be installed on the border leaf.

- Advertise Host Route is supported for shared services. For example: epg1/BD1 deployed is in VRF-1 and L3out in another VRF-2. By providing shared contract between EPG and L3out host routes are pulled from one VRF-1 to another VRF-2.
- When Advertise Host Route is enabled on BD custom tag cannot be set on BD Subnet using route-map.
- When Advertise Host Route is enabled on a BD and the BD is associated with an L3Out, BD subnet is marked public. If there's a rogue EP present under the BD, that EP is advertised out on L3Out.

Guidelines for Routed Connectivity to Outside Networks

Use the following guidelines when creating and maintaining Layer 3 outside connections.

Topic	Caution or Guideline
Floating SVIs	When running ESXi on UCS B Series blade switches behind a fabric interconnect, we recommend that you leave "Fabric Failover" disabled and allow the DVS running on ESXi itself to achieve redundancy in the event of a failure. If enabled, the LLDP/CDP packets that Cisco ACI uses for deployment will be seen on the active and standby virtual switch ports (vEths), which could cause constant flapping and deployment issues.

Topic	Caution or Guideline
Issue where a border leaf switch in a vPC pair forwards a BGP packet with an incorrect VNID to an on-peer learned endpoint	If the following conditions exist in your configuration: • Two leaf switches are part of a vPC pair • For the two leaf switches connected behind the L3Out, the destination endpoint is connected to the second (peer) border leaf switch, and the endpoint is on-peer learned on that leaf switch If the endpoint is on-peer learned on the ingress leaf switch that receives a BGP packet that is destined to the on-peer learned endpoint, an issue might arise where the transit BGP connection fails to establish between the first layer 3 switch behind the L3Out and the on-peer learned endpoint on the second leaf switch in the vPC pair. This might happen in this situation because the transit BGP packet with port 179 is forwarded incorrectly using the bridge domain VNID instead of the VRF VNID. To resolve this issue, move the endpoint to any other non-peer leaf switch in the fabric so that it is not learned on the leaf switch.
Border leaf switches and GIR (maintenance) mode	If a border leaf switch has a static route and is placed in Graceful Insertion and Removal (GIR) mode, or maintenance mode, the route from the border leaf switch might not be removed from the routing table of switches in the ACI fabric, which causes routing issues. To work around this issue, either: • Configure the same static route with the same administrative distance on the other border leaf switch, or • Use IP SLA or BFD for track reachability to the next hop of the static route
L3Out aggregate stats do not support egress drop counters	When accessing the **Select Stats** window through **Tenants** > ***tenant_name*** > **Networking** > **L3Outs** > ***L3Out_name*** > **Stats**, you will see that L3Out aggregate stats do not support egress drop counters. This is because there is currently no hardware table in the ASICs that record egress drops from the EPG VLAN, so stats do not populate these counters. There are only ingress drops for the EPG VLAN.
Updates through CLI	For Layer 3 external networks created through the API or GUI and updated through the CLI, protocols need to be enabled globally on the external network through the API or GUI, and the node profile for all the participating nodes needs to be added through the API or GUI before doing any further updates through the CLI.
Loopbacks for Layer 3 networks on same node	When configuring two Layer 3 external networks on the same node, the loopbacks need to be configured separately for both Layer 3 networks.

Topic	Caution or Guideline
Ingress-based policy enforcement	Starting with Cisco APIC release 1.2(1), ingress-based policy enforcement enables defining policy enforcement for Layer 3 Outside (L3Out) traffic for both egress and ingress directions. The default is ingress. During an upgrade to release 1.2(1) or higher, existing L3Out configurations are set to egress so that the behavior is consistent with the existing configuration. You do not need any special upgrade sequence. After the upgrade, you change the global property value to ingress. When it has been changed, the system reprograms the rules and prefix entries. Rules are removed from the egress leaf and installed on the ingress leaf, if not already present. If not already configured, an `Actrl` prefix entry is installed on the ingress leaf. Direct server return (DSR), and attribute EPGs require ingress based policy enforcement. vzAny and taboo contracts ignore ingress based policy enforcement. Transit rules are applied at ingress.
Bridge Domains with L3Outs	A bridge domain in a tenant can contain a public subnet that is advertised through an `l3extOut` provisioned in the common tenant.
Bridge domain route advertisement For OSPF and EIGRP	When both OSPF and EIGRP are enabled on the same VRF on a node and if the bridge domain subnets are advertised out of one of the L3Outs, it will also get advertised out of the protocol enabled on the other L3Out. For OSPF and EIGRP, the bridge domain route advertisement is per VRF and not per L3Out. The same behavior is expected when multiple OSPF L3Outs (for multiple areas) are enabled on the same VRF and node. In this case, the bridge domain route will be advertised out of all the areas, if it is enabled on one of them.
BGP Maximum Prefix Limit	Starting with Cisco APIC release 1.2(1x), tenant policies for BGP `l3extOut` connections can be configured with a maximum prefix limit, that enables monitoring and restricting the number of route prefixes received from a peer. Once the maximum prefix limit has been exceeded, a log entry is recorded, and further prefixes are rejected. The connection can be restarted if the count drops below the threshold in a fixed interval, or the connection is shut down. Only one option can be used at a time. The default setting is a limit of 20,000 prefixes, after which new prefixes are rejected. When the reject option is deployed, BGP accepts one more prefix beyond the configured limit, before the APIC raises a fault.

Topic	Caution or Guideline
MTU	• Cisco ACI does not support IP fragmentation. Therefore, when you configure Layer 3 Outside (L3Out) connections to external routers, or Multi-Pod connections through an Inter-Pod Network (IPN), it is recommended that the interface MTU is set appropriately on both ends of a link. On some platforms, such as Cisco ACI, Cisco NX-OS, and Cisco IOS, the configurable MTU value does not take into account the Ethernet headers (matching IP MTU, and excluding the 14-18 Ethernet header size), while other platforms, such as IOS-XR, include the Ethernet header in the configured MTU value. A configured value of 9000 results in a max IP packet size of 9000 bytes in Cisco ACI, Cisco NX-OS, and Cisco IOS, but results in a max IP packet size of 8986 bytes for an IOS-XR untagged interface. • The MTU settings for the Cisco ACI physical interfaces vary: • For sub-interfaces, the physical interface MTU is fixed and is set to 9216 for the front panel ports on the leaf switches. • For SVI, the physical interface MTU is set based on the fabric MTU policy. For example, if the fabric MTU policy is set to 9000, then the physical interface for the SVI is set to 9000.
QoS for L3Outs	To configure QoS policies for an L3Out and enable the policies to be enforced on the BL switch where the L3Out is located, use the following guidelines: • The VRF Policy Control Enforcement Direction must be set to**Egress**. • The VRF Policy Control Enforcement Preference must be set to **Enabled**. • When configuring the contract that controls communication between the EPGs using the L3Out, include the QoS class or Target DSCP in the contract or subject of the contract.
ICMP settings	ICMP redirect and ICMP unreachable are disabled by default in Cisco ACI to protect the switch CPU from generating these packets.

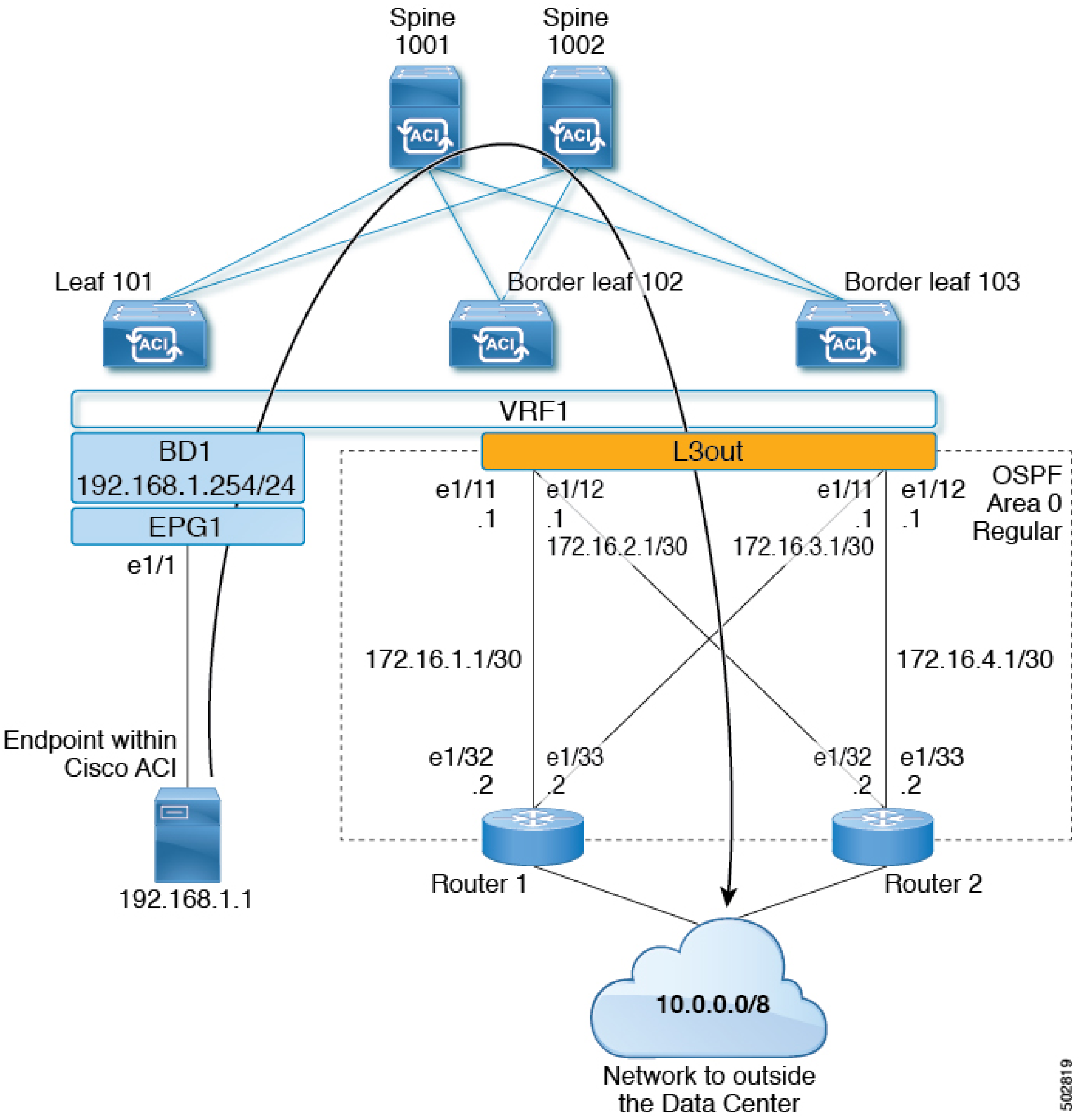
Example L3Out Configuration

There are many different options available to you when configuring an L3Out using the **Create L3Out** wizard. Following is one example L3Out configuration, where you will configure an OSPF L3Out with two external routers, that will help you to understand the general configuration process.

Note This example uses Cisco APIC release 4.2(x) and the associated GUI screens.

Example Topology

Figure 2: Example Topology for an OSPF L3Out with Two External Routers

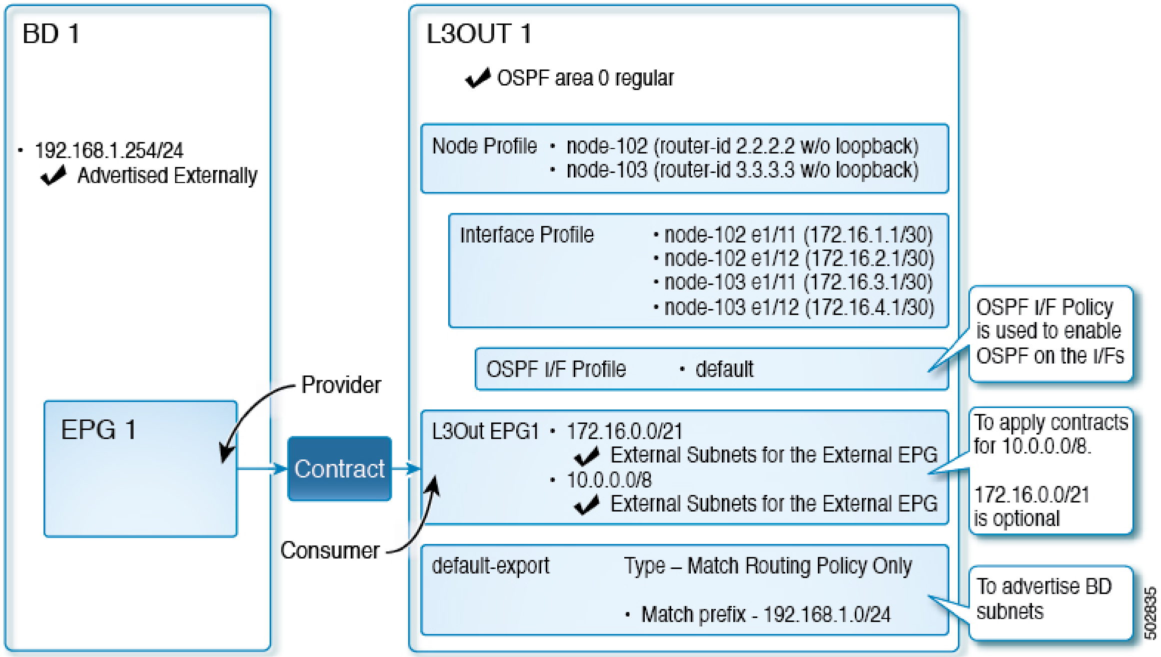


This basic L3Out example shows you how to:

- Configure an L3Out with the following specifications:
  - With Area 0 OSPF
  - With two external routers
  - With routed interfaces
  - On two border leaf switches

- Advertise a BD subnet using default route-map (default-export)
- Allow communication with a contract between EPG1 and external route (10.0.0.0/8)

Figure 3: OSPF Configuration Diagram

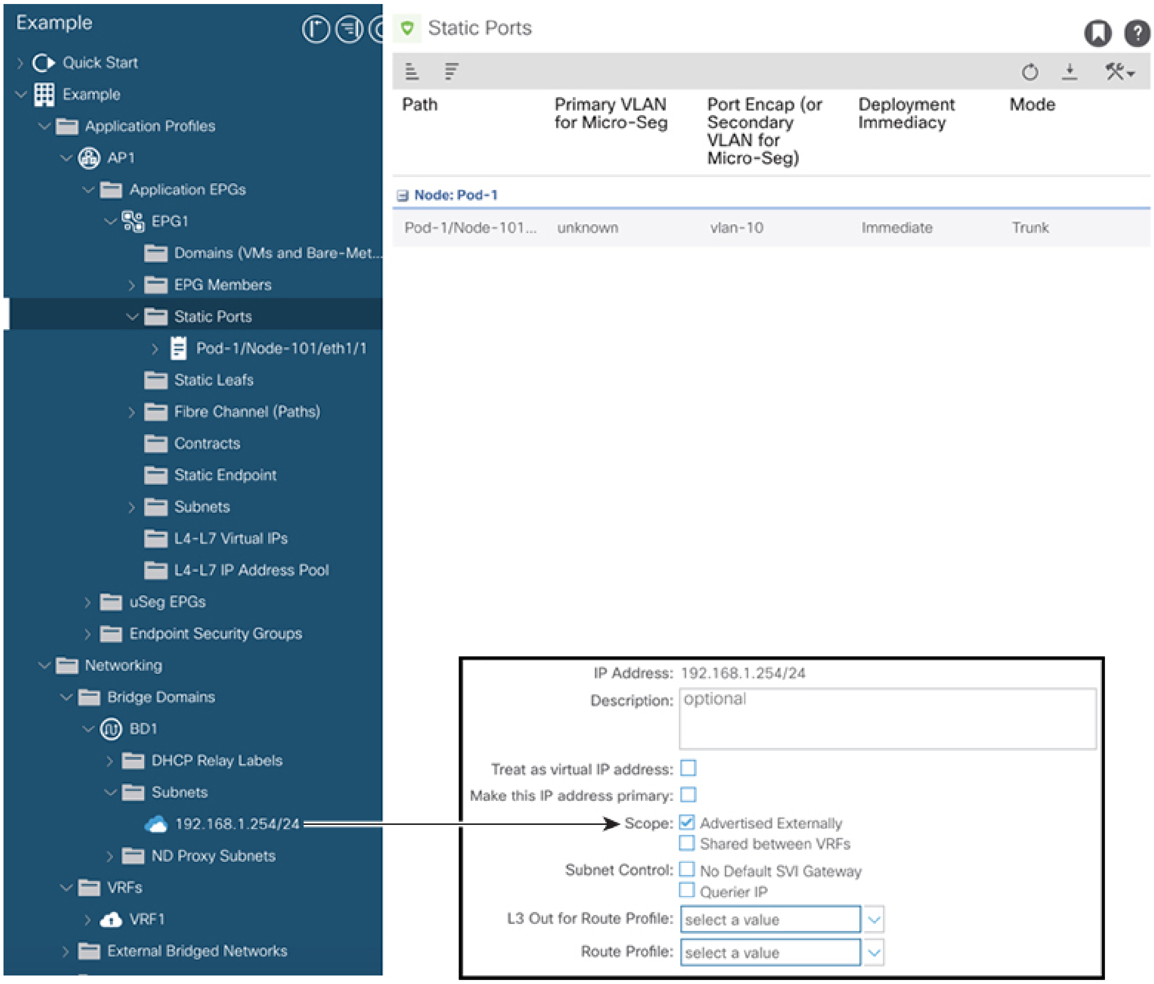

BD 1
• 192.168.1.254/24
✓ Advertised Externally
EPG 1
Provider
Contract
Consumer
L3OUT 1
✓ OSPF area 0 regular
Node Profile • node-102 (router-id 2.2.2.2 w/o loopback)
• node-103 (router-id 3.3.3.3 w/o loopback)
Interface Profile • node-102 e1/11 (172.16.1.1/30)
• node-102 e1/12 (172.16.2.1/30)
• node-103 e1/11 (172.16.3.1/30)
• node-103 e1/12 (172.16.4.1/30)
OSPF I/F Profile • default
L3Out EPG1 • 172.16.0.0/21
✓ External Subnets for the External EPG
• 10.0.0.0/8
✓ External Subnets for the External EPG
default-export Type – Match Routing Policy Only
• Match prefix - 192.168.1.0/24
OSPF I/F Policy is used to enable OSPF on the I/Fs
To apply contracts for 10.0.0.0/8.
172.16.0.0/21 is optional
To advertise BD subnets
502835

The preceding diagram illustrates the configuration for the example topology in Figure 2: Example Topology for an OSPF L3Out with Two External Routers, on page 18. The configuration flow for this example is as follows:

1. L3Out: This creates
   - L3Out itself (OSPF parameters)
   - Node, Interface, OSPF I/F Profiles
   - L3Out EPG with **External Subnets for the External EPG** scope
2. Advertise a BD subnet: This uses
   - **default-export** route-map
   - BD subnet with **Advertise Externally** scope
3. Allow EPG - L3Out communication: This uses a contract between EPG1 and L3Out EPG1

Prerequisites

Figure 4: Example Screen of Objects Created as Prerequisites



Path	Primary VLAN for Micro-Seg	Port Encap (or Secondary VLAN for Micro-Seg)	Deployment Immediacy	Mode
Node: Pod-1				
Pod-1/Node-101...	unknown	vlan-10	Immediate	Trunk

IP Address: 192.168.1.254/24

Description: optional

Treat as virtual IP address: □

Make this IP address primary: □

Scope: ☑ Advertised Externally
□ Shared between VRFs

Subnet Control: □ No Default SVI Gateway
□ Querier IP

L3 Out for Route Profile: select a value

Route Profile: select a value

- This configuration example focuses only on the L3Out configuration part. The other configurations such as for VRF, BD, EPG, Application Profiles, and Access Policies (Layer 3 Domain etc.) are not covered. The preceding screenshot displays the prerequisite tenant configurations that are as follows:
  - VRF1
  - BD1 with the subnet 192.168.1.254/24
  - EPG1 with a static port towards endpoints

Create Example L3Out Using the Create L3Out Wizard

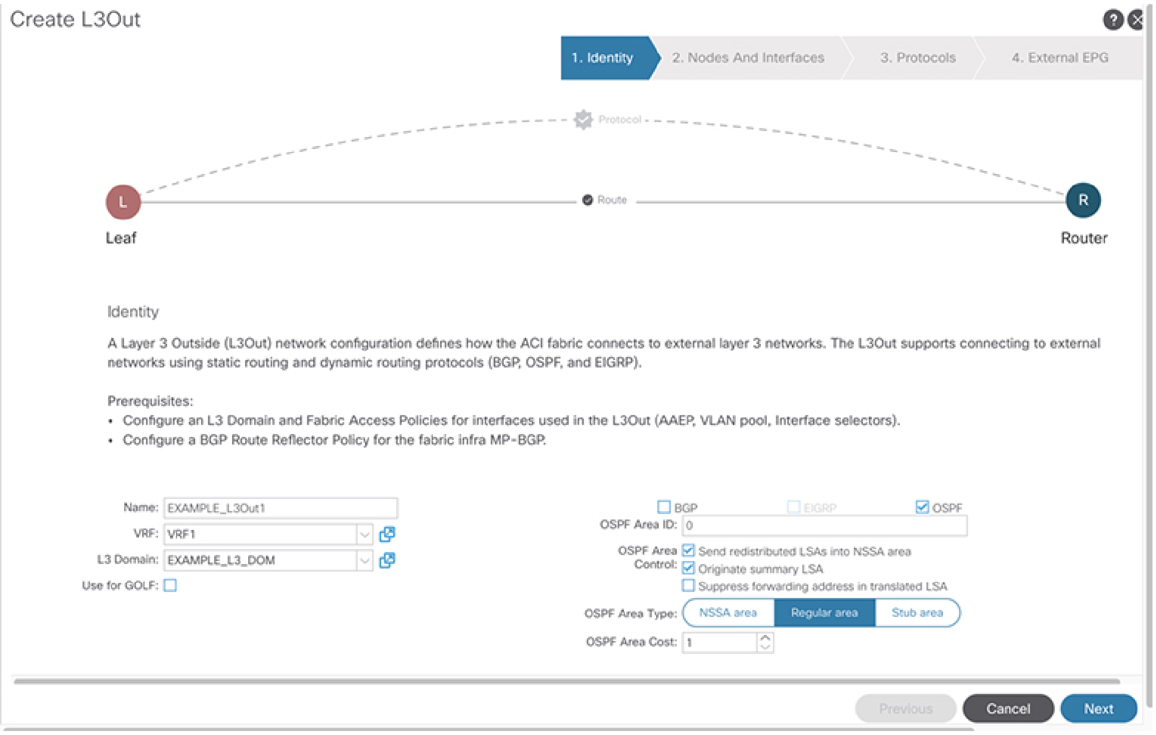
This task creates the OSPF L3Out described in Example Topology. Following this task, Cisco ACI will be configured with two border leaf switches and OSPF neighborship with two external routers as shown in Figure 2: Example Topology for an OSPF L3Out with Two External Routers, on page 18.

Procedure

Step 1 In the GUI **Navigation** pane, under the Tenant Example, navigate to **Networking** > **L3Outs**.

Step 2 Right-click and choose **Create L3Out**.

Step 3 In the **Create L3Out** screen, **Identity** tab, perform the following actions:



Create L3Out

Identity

A Layer 3 Outside (L3Out) network configuration defines how the ACI fabric connects to external layer 3 networks. The L3Out supports connecting to external networks using static routing and dynamic routing protocols (BGP, OSPF, and EIGRP).

Prerequisites:

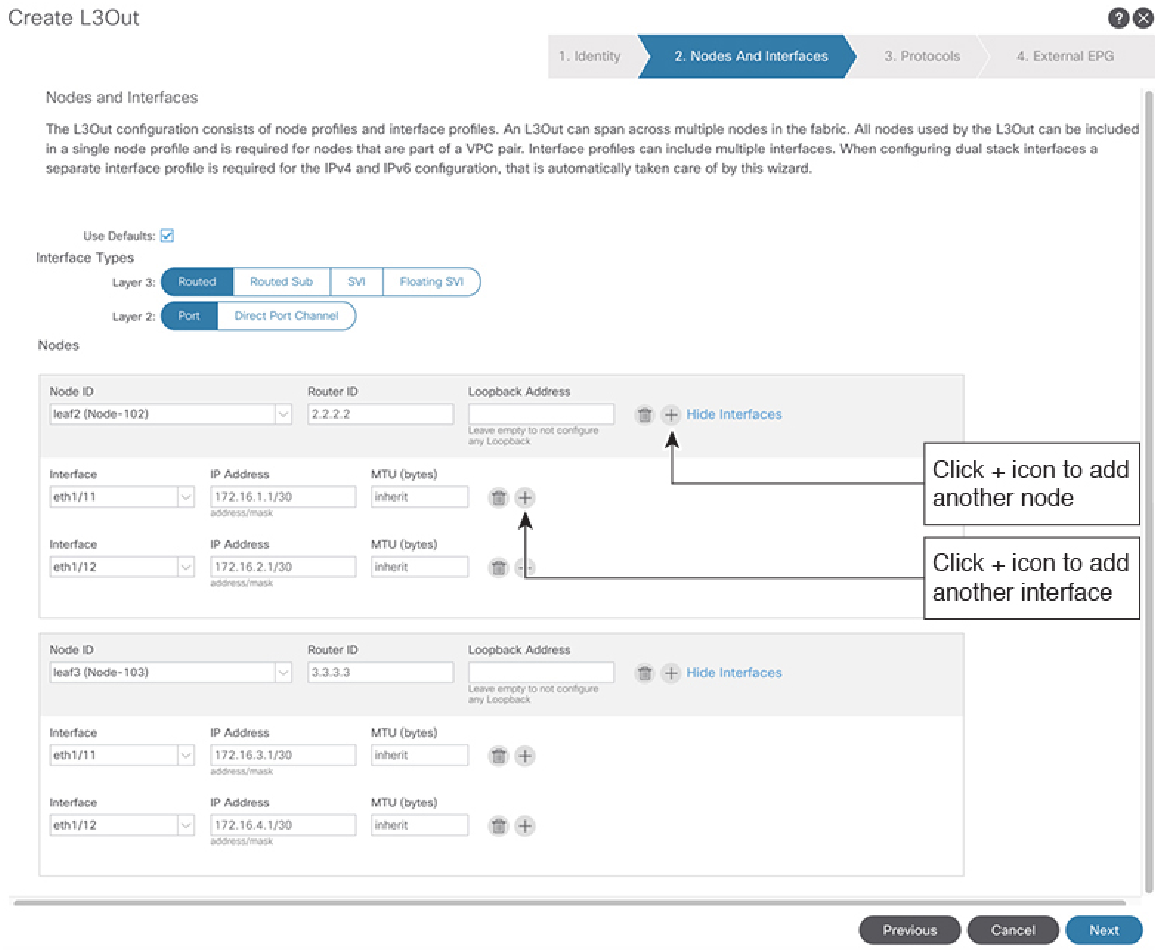
- Configure an L3 Domain and Fabric Access Policies for interfaces used in the L3Out (AAEP, VLAN pool, Interface selectors).
- Configure a BGP Route Reflector Policy for the fabric infra MP-BGP.

Name: EXAMPLE_L3Out1
VRF: VRF1
L3 Domain: EXAMPLE_L3_DOM
Use for GOLF: □

□ BGP □ EIGRP ☑ OSPF
OSPF Area ID: 0
OSPF Area Control: ☑ Send redistributed LSAs into NSSA area
☑ Originate summary LSA
□ Suppress forwarding address in translated LSA
OSPF Area Type: NSSA area [✓ Regular area] Stub area
OSPF Area Cost: 1

a) In the **Name** field, enter the name for an L3Out. (EXAMPLE_L3Out1)
b) In the **VRF** field and the **L3 Domain** field, choose the appropriate values. (VRF1, EXAMPLE_L3DOM)
c) In the **OSPF** field, check the checkbox.
d) In the **OSPF Area ID** field, choose the value **0** or the text **backbone**.
e) In the **OSPF Area Type** field, choose **Regular area**.
f) Keep the rest of the fields with their default values.

Step 4 Click **Next** to display the **Nodes and Interfaces** screen, and perform the following actions:



Create L3Out

Nodes and Interfaces

The L3Out configuration consists of node profiles and interface profiles. An L3Out can span across multiple nodes in the fabric. All nodes used by the L3Out can be included in a single node profile and is required for nodes that are part of a VPC pair. Interface profiles can include multiple interfaces. When configuring dual stack interfaces a separate interface profile is required for the IPv4 and IPv6 configuration, that is automatically taken care of by this wizard.

Use Defaults: ☑

Interface Types

Layer 3: [✓ Routed] Routed Sub SVI Floating SVI

Layer 2: [✓ Port] Direct Port Channel

Nodes

Node ID: leaf2 (Node-102)
Router ID: 2.2.2.2
Loopback Address: (Leave empty to not configure any Loopback)

Interface	IP Address	MTU (bytes)
eth1/11	172.16.1.1/30	inherit
eth1/12	172.16.2.1/30	inherit

Node ID: leaf3 (Node-103)
Router ID: 3.3.3.3
Loopback Address: (Leave empty to not configure any Loopback)

Interface	IP Address	MTU (bytes)
eth1/11	172.16.3.1/30	inherit
eth1/12	172.16.4.1/30	inherit

Click + icon to add another node

Click + icon to add another interface

a) In the **Interface Types** area, in the **Layer 3** field and in the **Layer 2** field, ensure that your selections match the choices in the preceding screenshot (Routed and Port).

b) In the **Nodes** area, in the **Node ID** field, from the drop-down list, choose the appropriate node ID. (leaf2 (Node 102))

c) In the **Router ID** field, enter the appropriate router ID. (2.2.2.2)

   The **Loopback Address** field auto populates based on the router ID value you enter. You do not require the loopback address, so delete the value and leave the field blank.

d) In the **Interface** field, choose the interface ID. (eth1/11)

e) In the **IP Address** field, enter the associated IP address. (172.16.1.1/30)

f) In the **MTU** field, keep the default value. (inherit)

g) Click the + icon next to the **MTU** field to add an additional interface for node leaf2. (Node-102)

h) In the **Interface** field, choose the interface ID. (eth1/12)

i) In the **IP Address** field, enter the associated IP address. (172.16.2.1/30)

j) In the **MTU** field, keep the default value. (inherit)

Step 5 To add another node, click the + icon next to the **Loopback Address** field, and perform the following actions:

Note When you click the + icon, the new **Nodes** area is displayed below the area that you had populated earlier.

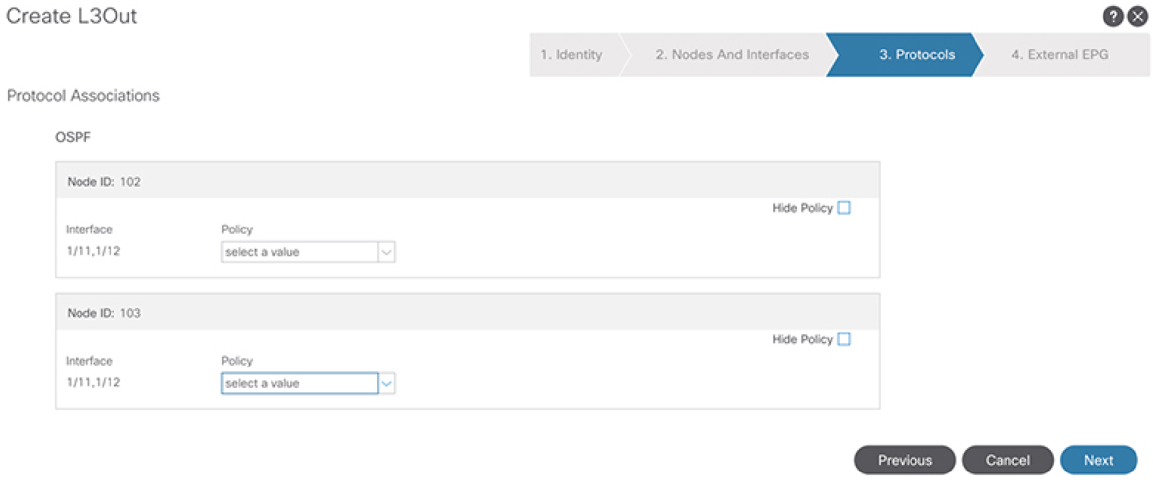
a) In the **Nodes** area, in the **Node ID** field, from the drop-down list, choose the node ID. (leaf3 (Node-103))
b) In the **Router ID** field, enter the router ID. (3.3.3.3)

The **Loopback Address** field auto populates based on the router ID value you enter. You do not require the loopback address, so delete the value and leave the field blank.

c) In the **Interface** field, choose the interface ID. (eth1/11)
d) In the **IP Address** field, enter the IP address. (172.16.3.1/30)
e) In the **MTU** field, keep the default value. (inherit)
f) Click the **+** icon next to the **MTU** field to add an additional interface for node leaf3. (Node-103)
g) In the **Interface** field, choose the interface ID. (eth1/12)
h) In the **IP Address** field, enter the associated IP address. (172.16.4.1/30)
i) In the **MTU** field, keep the default value. (inherit), and click **Next**.
We have specified the node, interface, and IP address for each interface.

Step 6 Click **Next** to view the **Protocols** screen.

This screen allows you to specify the OSPF interface level policy to configure hello-interval, network-type,



Protocol Associations

OSPF

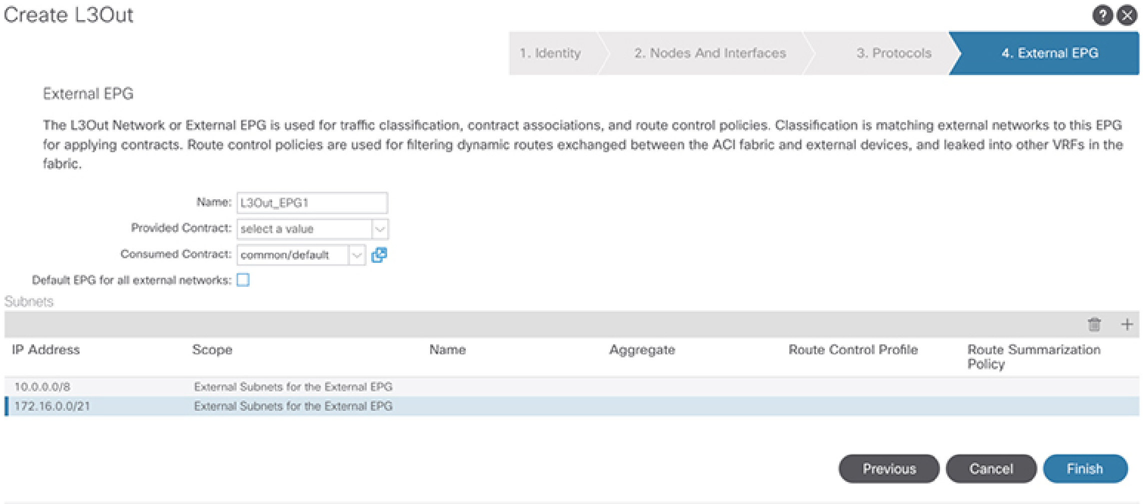
Node ID: 102
Hide Policy □
Interface: 1/11,1/12
Policy: select a value

Node ID: 103
Hide Policy □
Interface: 1/11,1/12
Policy: select a value

In this example, nothing is selected. Therefore, the default policy is used. The default OSPF interface profile uses **Unspecified** as network-type which defaults to broadcast network type. To optimize this with point-to-point network-type for sub-interface, see **Change the OSPF Interface Level Parameters (Optional)**.

Step 7 Click **Next**.
The **External EPG** screen is displayed with L3Out EPG details. This configuration is to classify the traffic into the EPG to apply to the contract.

Step 8 In the **External EPG** screen, perform the following actions:



Create L3Out

External EPG

The L3Out Network or External EPG is used for traffic classification, contract associations, and route control policies. Classification is matching external networks to this EPG for applying contracts. Route control policies are used for filtering dynamic routes exchanged between the ACI fabric and external devices, and leaked into other VRFs in the fabric.

Name: L3Out_EPG1

Provided Contract: select a value

Consumed Contract: common/default

Default EPG for all external networks: □

Subnets

IP Address	Scope	Name	Aggregate	Route Control Profile	Route Summarization Policy
10.0.0.0/8	External Subnets for the External EPG				
172.16.0.0/21	External Subnets for the External EPG				

a) In the **External EPG** area, **Name** field, enter a name for the external EPG. (L3Out_EPG1)

b) In the **Provided Contract** field, do not choose a value.

In this example, there is no provided contract for L3Out_EPG1 because a normal EPG (EPG1) is the provider.

c) In the **Consumed Contract** field, choose **default** from the drop-down list.

Step 9 In the **Default EPG for all external networks** field, uncheck the checkbox, and perform the following actions:

a) Click the **+** icon in the **Subnets** area, to display the **Create Subnet** dialog box.

b) In the **IP Address** field, enter the subnet. (10.0.0.0/8)

c) In the **External EPG Classification** field, check the checkbox for **External Subnets for the External EPG**. Click **OK**.

Step 10 Click the **+** icon in the **Subnets** area once more to display the **Create Subnet** dialog box, and perform the following actions:

Note Although this is an optional configuration, it is a best practice to specify the L3Out interface subnets in case endpoints have to communicate with those IPs.

a) In the **IP Address** field, enter the subnet. (172.16.0.0/21)

This subnet covers all the interfaces in the L3Out. This can be each individual subnet for each routed interface instead.

b) In the **External EPG Classification** field, check the checkbox for **External Subnets for the External EPG**. Click **OK**.

c) Click **Finish**.

The L3Out OSPF is now deployed.

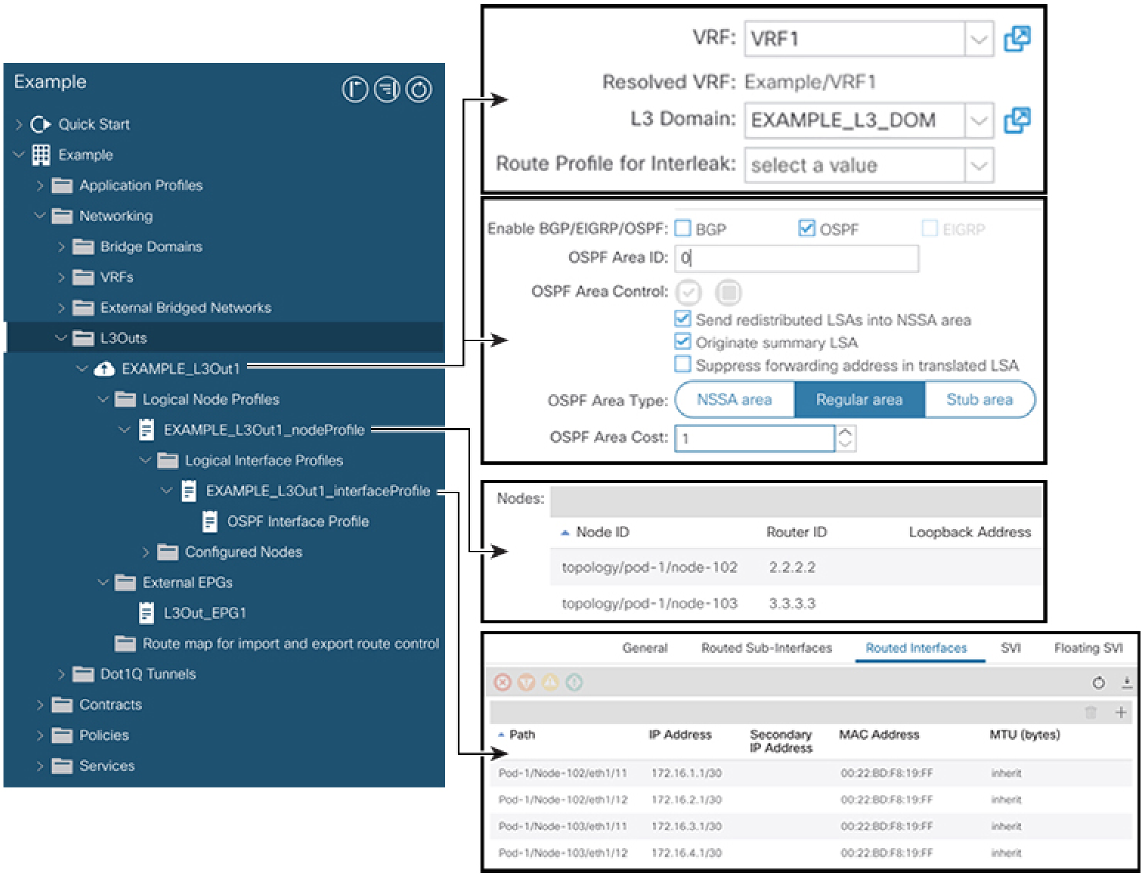
Review - Create Example L3Out Using the Create L3Out Wizard

Review how the configuration using the wizard is presented in the Cisco APIC GUI, and verify that the configurations are accurate.

Procedure

Step 1 Navigate to your **Tenant_name** > **Networking** > **L3Outs** > **EXAMPLE_L3Out1**, in the **Work** pane, scroll to view the details as follows:

At this location in the GUI, verify the main L3Out parameters such as VRF, domain, and OSPF parameters that are configured in the **Identity** screen in the **Create L3Out** wizard.



VRF: VRF1

Resolved VRF: Example/VRF1

L3 Domain: EXAMPLE_L3_DOM

Route Profile for Interleak: select a value

Enable BGP/EIGRP/OSPF: □ BGP ☑ OSPF □ EIGRP

OSPF Area ID: 0

OSPF Area Control:
☑ Send redistributed LSAs into NSSA area
☑ Originate summary LSA
□ Suppress forwarding address in translated LSA

OSPF Area Type: NSSA area [✓ Regular area] Stub area

OSPF Area Cost: 1

Nodes:

Node ID	Router ID	Loopback Address
topology/pod-1/node-102	2.2.2.2	
topology/pod-1/node-103	3.3.3.3	

Path	IP Address	Secondary IP Address	MAC Address	MTU (bytes)
Pod-1/Node-102/eth1/11	172.16.1.1/30		00:22:BD:F8:19:FF	inherit
Pod-1/Node-102/eth1/12	172.16.2.1/30		00:22:BD:F8:19:FF	inherit
Pod-1/Node-103/eth1/11	172.16.3.1/30		00:22:BD:F8:19:FF	inherit
Pod-1/Node-103/eth1/12	172.16.4.1/30		00:22:BD:F8:19:FF	inherit

Step 2 Verify that OSPF is enabled with the specified parameters such as Area ID and Area Type.

Step 3 Under **Logical Node Profiles**, *EXAMPLE_L3Out1_nodeProfile* is created to specify border leaf switches with their router IDs.

Step 4 Under **Logical Interface Profile**, *EXAMPLE_L3Out1_interfaceProfile* is created.

Verify the interface parameters such as interface ID, IP addresses, in this example, as routed interfaces. The default MAC addresses gets auto populated. OSPF interface profile is also created under this for OSPF interface level parameters.

The review is complete.

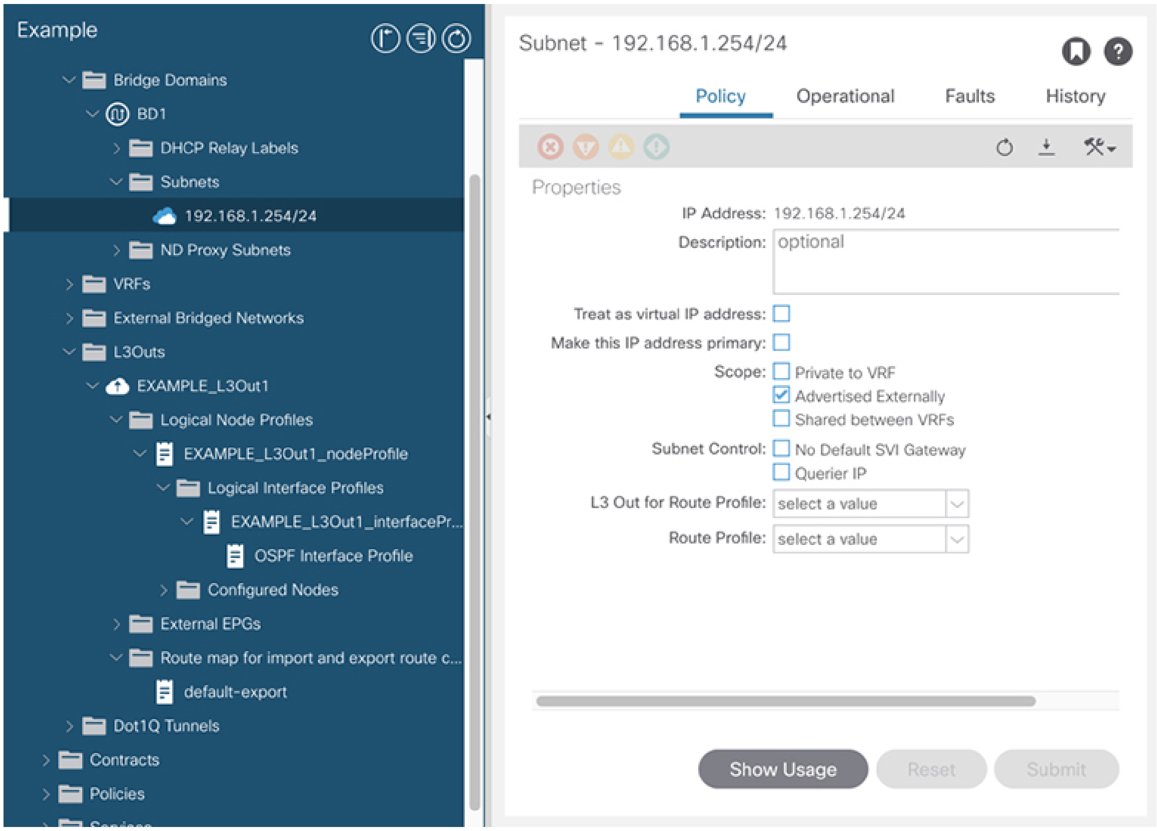
Configure Advertise the BD Subnet with a Route Map

In this example, a route map, **default-export**, is used with the IP prefix list to advertise the BD subnet.

Note This **default-export** route map will be applied to the L3Out (EXAMPLE_L3Out1) without being associated to anything specific.

Procedure

Step 1 To enable a BD subnet to be advertised, navigate to **Tenant** > **Networks** > **Bridge Domains** > ***BD1*** > **Subnets** > ***192.168.1.254/24***, and select **Advertised Externally** scope.



Subnet - 192.168.1.254/24

Properties

IP Address: 192.168.1.254/24

Description: optional

Treat as virtual IP address: □

Make this IP address primary: □

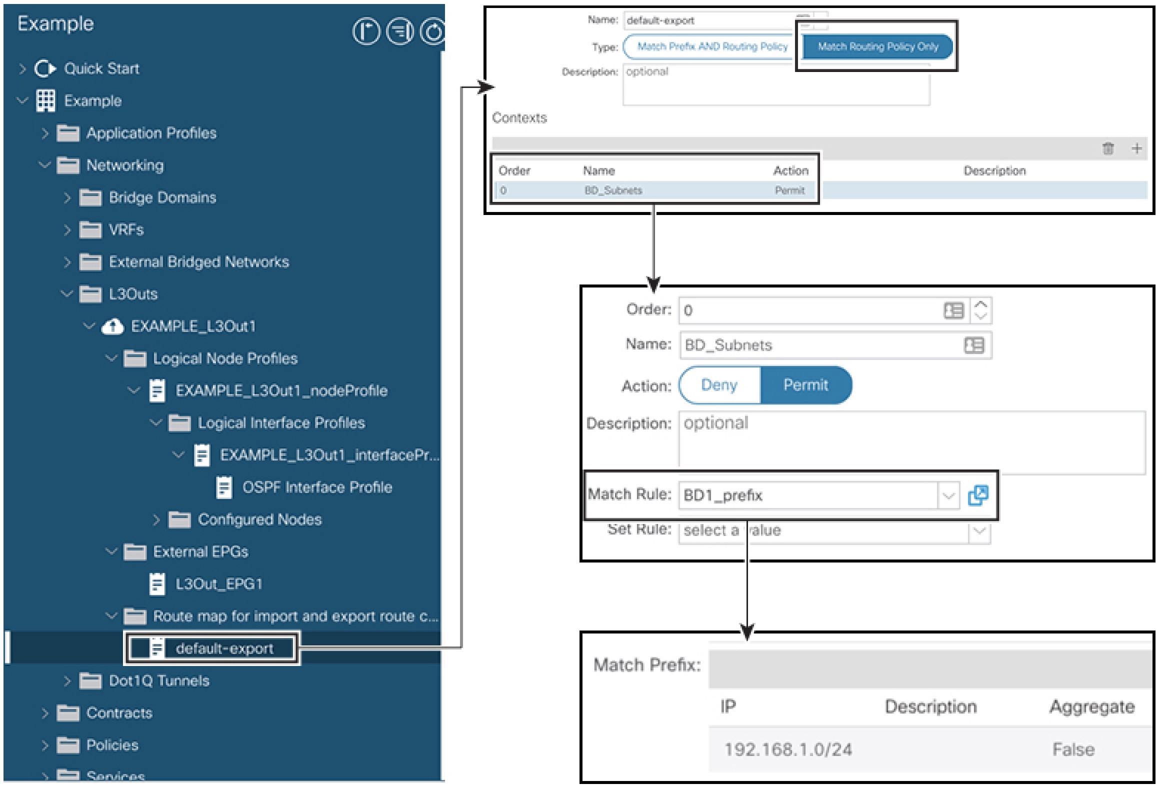
Scope: □ Private to VRF
☑ Advertised Externally
□ Shared between VRFs

Subnet Control: □ No Default SVI Gateway
□ Querier IP

L3 Out for Route Profile: select a value

Route Profile: select a value

Step 2 To create a route map under your L3Out (EXAMPLE_L3Out1), navigate to **Route map for import and export route control**.



Name: default-export

Type: Match Prefix AND Routing Policy | Match Routing Policy Only

Description: optional

Contexts

Order	Name	Action	Description
0	BD_Subnets	Permit	

Order: 0

Name: BD_Subnets

Action: Deny | Permit

Description: optional

Match Rule: BD1_prefix

Set Rule: select a value

Match Prefix:

IP	Description	Aggregate
192.168.1.0/24		False

Step 3 Right-click and choose **Create Route map for import and export route control**.

Step 4 In the **Create Route map for import and export route control** dialog box, in the **Name** field, choose **default-export**.

Step 5 In the **Type** field, choose **Matching Route Policy Only**.

Note **Match Routing Policy Only**: By choosing this **Type** with default-export route map, all route advertisement configuration is performed by this route map. BD associations and export route control subnets configured under the external EPG will not apply. You should configure all match rules within this route-map for all routes that will be advertised from this L3Out.

Match Prefix and Routing Policy: By choosing this **Type** with default-export route map, route advertisement is matched by any match rules configured in this route map **in addition to** any BD to L3Out associations and export route control subnets defined under the External EPG.

When using a route profile, it is recommended to use **Match Routing Policy Only** for a simpler configuration that is easier to maintain.

Step 6 In the **Contexts** area, click the **+** icon, to display the **Create Route Control Context** dialog box, and perform the following actions:

a) In the **Order** field, configure the order. (0)

In this example, we have only one order.

b) In the **Name** field, enter a name for the context. (BD_Subnets)

c) In the **Action** field, choose **Permit**.

This enables the route map to permit the prefix we will configure.

In this example, we require the match rule that requires the IP prefix list, **BD1_prefix**. This IP prefix list points to the BD subnet advertised.

Step 7 In the **Match Rule** field, create the IP prefix-list by performing the following actions;

a) Choose **Create Match Rule for a Route-Map**.
b) In the **Name** field, enter a name *BD1_prefix*.
c) In the **Match Prefix** area, click the + icon, and enter the BD subnet (192.168.1.0/24).

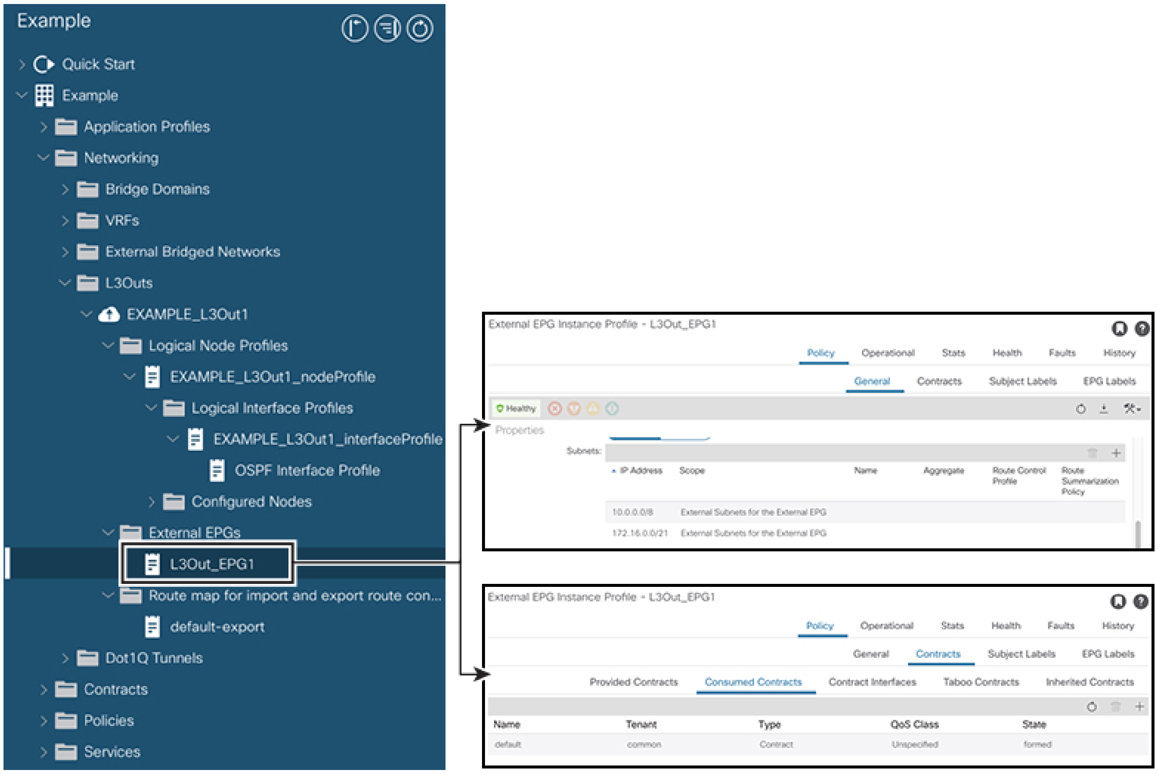
Verify the Contract

In this task, you verify the contract to enable communication between an endpoint (192.168.1.1) and external prefixes (10.0.0.0/8, and optionally 172.16.0.0/21). In this example, the EPG for the endpoint is EPG1 and the external EPG for external prefixes is L3Out_EPG1.

The required configuration should already be present from the **Create L3Out** wizard.

Procedure

Step 1 Under your L3Out, navigate to **External EPGs** > **L3Out_EPG1**.



IP Address	Scope	Name	Aggregate	Route Control Profile	Route Summarization Policy
10.0.0.0/8	External Subnets for the External EPG				
172.16.0.0/21	External Subnets for the External EPG				

Name	Tenant	Type	QoS Class	State
default	common	Contract	Unspecified	formed

Step 2 In the **Work** pane, in the **External EPG Instance Profile** area, under **Policy** > **General** sub-tab, look at the Properties and verify that the two subnets are displayed with **External Subnets for the External EPG**.

Step 3 Next, click the **Contracts** sub-tab and verify the contract you specified earlier is consumed correctly. In case you want to add more contracts, you can perform the actions from this location in GUI.

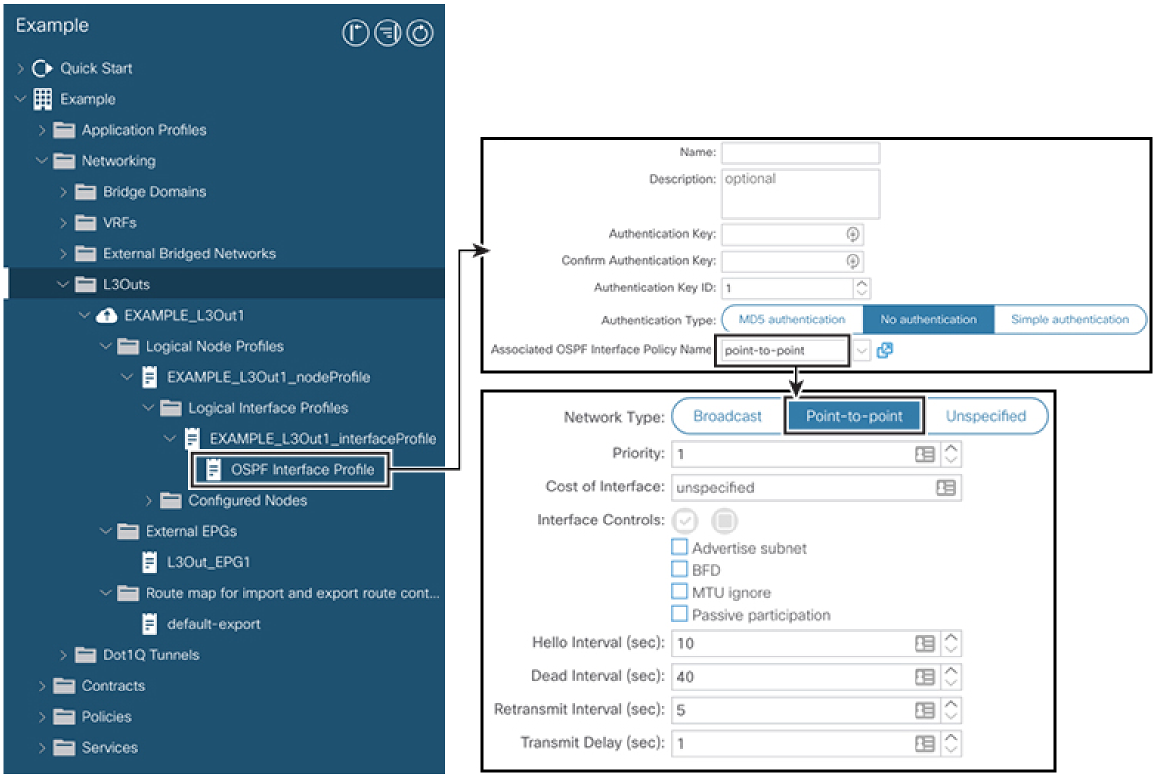
Step 4 Navigate to **Application Profile** > **Application EPGs** > ***EPG1*** > **Contracts**, and verify that EPG1 is providing the appropriate contract.

Change the OSPF Interface Level Parameters (Optional)

If you wish to change the OSPF interface-level parameters, such as **Hello Interval, OSPF network type**, then you can configure it in the OSPF Interface Profile. The node level OSPF parameters are already configured.

Procedure

Step 1 Under your L3Out, navigate to **Logical Interface Profile** > **EXAMPLE_L3Out1_interfaceProfile** > **OSPF Interface Profile**.



Name:
Description: optional
Authentication Key:
Confirm Authentication Key:
Authentication Key ID: 1
Authentication Type: MD5 authentication [✓ No authentication] Simple authentication
Associated OSPF Interface Policy Name: point-to-point

Network Type: Broadcast [✓ Point-to-point] Unspecified
Priority: 1
Cost of Interface: unspecified
Interface Controls:
□ Advertise subnet
□ BFD
□ MTU ignore
□ Passive participation
Hello Interval (sec): 10
Dead Interval (sec): 40
Retransmit Interval (sec): 5
Transmit Delay (sec): 1

Step 2 In the **Work** pane, in the **Properties** area, choose the OSPF Interface Policy you wish to use.

This modifies your OSPF interface level parameters.